Statistics for Biology and Health

Richard J. Cook · Jerald F. Lawless

The Statistical Analysis of Recurrent Events

Springer

Statistics for Biology and Health

Series Editors
M. Gail, K. Krickeberg, J. Sarmet, A. Tsiatis, W. Wong

Statistics for Biology and Health

Bacchieri/Cioppa: Fundamentals of Clinical Research
Borchers/Buckland/Zucchini: Estimating Animal Abundance: Closed Populations
Burzykowski/Molenberghs/Buyse: The Evaluation of Surrogate Endpoints
Cook/Lawless: The Statistical Analysis of Recurrent Events
Duchateav/Janssen: The Frailty Model
Everitt/Rabe-Hesketh: Analyzing Medical Data Using S-PLUS
Ewens/Grant: Statistical Methods in Bioinformatics: An Introduction, 2nd ed.
Gentleman/Carey/Huber/Irizarry/Dudoit: Bioinformatics and Computational Biology Solutions Using R and Bioconductor
Hougaard: Analysis of Multivariate Survival Data
Keyfitz/Caswell: Applied Mathematical Demography, 3rd ed.
Klein/Moeschberger: Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed.
Kleinbaum/Klein: Survival Analysis: A Self-Learning Text, 2nd ed.
Kleinbaum/Klein: Logistic Regression: A Self-Learning Text, 2nd ed.
Lange: Mathematical and Statistical Methods for Genetic Analysis, 2nd ed.
Manton/Singer/Suzman: Forecasting the Health of Elderly Populations
Martinussen/Scheike: Dynamic Regression Models for Survival Data
Moyé: Multiple Analyses in Clinical Trials: Fundamentals for Investigators
Nielsen: Statistical Methods in Molecular Evolution
O'Quigley: Proportional Hazards Regression
Parmigiani/Garrett/Irizarry/Zeger: The Analysis of Gene Expression Data: Methods and Software
Proschan/Lan Wittes: Statistical Monitoring of Clinical Trials: A Unified Approach
Siegmund/Yakir: The Statistics of Gene Mapping
Simon/Korn/McShane/Radmacher/Wright/Zhao: Design and Analysis of DNA Microarray Investigations
Sorensen/Gianola: Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics
Stallard/Manton/Cohen: Forecasting Product Liability Claims: Epidemiology and Modeling in the Manville Asbestos Case
Sun: The Statistical Analysis of Interval-censored Failure Time Data
Therneau/Grambsch: Modeling Survival Data: Extending the Cox Model
Ting: Dose Finding in Drug Development
Vittinghoff/Glidden/Shiboski/McCulloch: Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models
Wu/Ma/Casella: Statistical Genetics of Quantitative Traits: Linkage, Map and QTL
Zhang/Singer: Recursive Partitioning in the Health Sciences
Zuur/Ieno/Smith: Analyzing Ecological Data

Richard J. Cook
Jerald F. Lawless

The Statistical Analysis of Recurrent Events

Springer

Richard J. Cook
Dept. Statistics & Actuarial Science
University of Waterloo, Waterloo, Ontario
200 University Avenue W.
Waterloo N2L 3G1
Canada
rjcook@uwaterloo.ca

Jerald F. Lawless
Dept. Statistics & Actuarial Science
University of Waterloo, Waterloo, Ontario
200 University Avenue W.
Waterloo N2L 3G1
Canada
jlawless@uwaterloo.ca

Library of Congress Control Number: 2007929451

ISBN 978-0-387-69809-0 e-ISBN 978-0-387-69810-6

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

To Joan and John Cook

To Jill, Kim, and Sarah

Preface

Recurrent event data arise in fields such as medicine and public health, business and industry, reliability, the social sciences, and insurance. The literature on the statistical analysis of recurrent events has grown rapidly over the past twenty years and a variety of models and methods has been developed. This book provides a comprehensive treatment of the area. We describe important models, explain their underlying assumptions and properties, consider settings where they are appropriate, and discuss in detail how to fit and base inferences on these models. Parametric, nonparametric, and semiparametric methods are covered. Many illustrative examples are given, most of which are taken from health or industrial settings.

This book is intended as a resource for persons interested in the modeling and analysis of recurrent events and as a text for a graduate course in statistics or biostatistics. We discuss results and models from stochastic processes in some detail, and have attempted to present the material in an accessible way with discussion of model formulation, estimation and inference, and numerous applications. The importance of model assessment is emphasized. Chapters are concluded with Problems and Supplements sections which give exercises as well as extensions to material in the text. An important feature of this book is the coverage of practical issues such as observation and subject-selection schemes, the planning of randomized experiments, incomplete data, and the prediction of future events. Areas needing further methodological development are also discussed.

Likelihood methods are emphasized as a basis for inference whenever possible. Estimating function theory is also used, especially for inference about marginal features when models are not fully specified. Appendix A provides a summary of relevant material on likelihood and estimating function methodology, but familiarity with statistical inference is assumed. Martingale representations are used for certain estimating functions, but we do not discuss asymptotic theory used to rigorously justify large sample results. Our approach is to indicate clearly the statistical basis of methodology without dwelling on regularity conditions and detailed proofs of asymptotic results.

Some background in survival analysis is beneficial, inasmuch as many methods for recurrent events are related to survival analysis and can be implemented with software for that area. Kalbfleisch and Prentice (2002) and Lawless (2003a) are references with a similar style of presentation to this book. Books which discuss models for recurrent event data include Cox and Lewis (1966), Cox and Isham (1980), Daley and Vere-Jones (1988), and other books on point processes. Andersen et al. (1993) provide a rigorous discussion of models and methods for the analysis of data arising from counting processes, and emphasize Markov processes. Therneau and Grambsch (2000) present methods for the analysis of recurrent event data along with applications using S-PLUS®, R and SAS. Nelson (2003) gives graphical procedures and simple methods for the analysis of recurrent events based on rate or mean functions. Other recent books which include some discussion of the analysis of recurrent event data include Hougaard (2000), Kalbfleisch and Prentice (2002), Martinussen and Scheike (2006), and Sun (2006). The present book goes beyond these treatments in the breadth of models addressed and in the attention paid to practical issues of design and analysis.

The data in examples are analyzed using S-PLUS, although identical code can be used in R (see www.r-project.org). In most cases there exist analogous procedures in SAS software. Datasets that are available to the public are listed in Appendix D and are posted at www.stats.uwaterloo.ca/cook-lawless/book.shtml along with sample code for S-PLUS or R and SAS.

Our interests in statistical methods for recurrent events have developed from working with several colleagues in various areas of research. We would like to acknowledge Nancy Heddle (McMaster University), Pierre Major (McMaster University), and Jeff Robinson (General Motors) for stimulating collaborations which have led to methodological development in this area. We also wish to thank colleagues at GlaxoSmithKline Inc., Novartis Pharmaceuticals Inc., and Bayer Canada Inc. for permission to use the data from clinical trials in several examples.

We are grateful to the faculty, visiting fellows, graduate students, and staff at University of Waterloo who help create a stimulating environment for research. In particular we would like to acknowledge collaborations involving recurrent events with Jean-Marie Boher, Bingshu Chen, Charmaine Dean, Daniel Fong, Marc Fredette, Joan Hu, Jack Kalbfleisch, Claude Nadeau, Edmund Ng, Wei Wei, Grace Yi, and Min Zhan. Mary Lou Dufton and Joan Hatton provided secretarial assistance in the preparation of this book, for which we are grateful. We would especially like to thank Ker-Ai Lee, whose expert statistical programming helped in the preparation of the examples, and who provides important support to our research.

Much of the work here was developed while the first author held an Investigator Award from the Canadian Institutes of Health Research and a Canada Research Chair in Statistical Methods for Health Research, and while the second author held an Industrial Research Chair co-sponsored by General

Motors Canada and the Natural Sciences and Engineering Research Council of Canada. This support is gratefully acknowledged.

Finally we would like to thank our wives Alison (R.J.C.) and Liz (J.F.L.) for their patience and support during the preparation of this book.

University of Waterloo *Richard Cook*
December 2006 *Jerry Lawless*

Glossary

The following is a summary of the notation used throughout this book.

- $I(A)$ is the indicator function, equaling 1 if A is true and 0 otherwise
- $\Pr(A)$ is the probability of event A
- $E(\cdot)$ denotes expectation, $\text{var}(\cdot)$ denotes variance, $\text{cov}(\cdot)$ denotes covariance, $\text{corr}(\cdot)$ denotes correlation, $\text{asvar}(\cdot)$ and $\text{ascov}(\cdot)$ denote asymptotic variance and covariance, respectively
- $m_X(t) = E(\exp(Xt))$ is the moment generating function of X
- $\Gamma(a) = \int_0^\infty u^{a-1} \exp(-u) du$ is the gamma function, where $a > 0$
- $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ is the beta function, where $a > 0$ and $b > 0$
- $g(x) \sim o(x)$ means $g(x)/x \to 0$ as $x \to 0$
- The transpose of a matrix A is A'
- Vectors are written in column form so, for example, $\theta = (\theta_1, \ldots, \theta_r)'$
- If $g(\theta) = (g_1(\theta), \ldots, g_k(\theta))'$ is a vector of functions, then $\partial g(\theta)/\partial\theta'$ is the $k \times r$ matrix with (i, j) element $\partial g_i(\theta)/\partial\theta_j$
- $\prod_{[a,b]}\{1 + g(u)du\}$ is a product integral; see Section 2.1
- The integral $\int_a^b dG(u)$ is a Riemann–Stieltjes integral; see Section 2.1.
- $L(\theta)$, $\ell(\theta)$, $U(\theta)$, $I(\theta)$, and $\mathcal{I}(\theta)$ represent the likelihood, log-likelihood, score, observed information, and expected information functions, respectively; see Appendix A
- $\widehat{\theta}$ denotes an estimate of the parameter θ
- If $\theta = (\theta_1', \theta_2')'$, then $\tilde{\theta}_1(\theta_2)$ is the profile likelihood estimate of θ_1 for fixed θ_2

- T_k is the time of the kth event for an individual
- $W_k = T_k - T_{k-1}$ is the duration of time between the $(k-1)$st and kth events for an individual
- $B(t) = t - T_{N(t^-)}$ is the backwards recurrence time, or the time since the last event before t
- $Y_k(t) = I(T_k < t \leq T_{k+1})$ indicates whether an individual is at risk of a $(k+1)$st event at time t; if there are J types of events then $Y_{jk}(t)$ indicates whether an individual is at risk of a $(k+1)$st type j event at time t, $j = 1, \ldots, J$.
- $\{N(t), 0 \leq t\}$ is a right-continuous counting process giving the number of events in $[0, t]$
- $N(s,t) = N(t) - N(s)$ records the number of events over $(s,t]$
- $\Delta N(t) = N(t + \Delta t^-) - N(t^-)$
- $dN(t) = \lim_{\Delta t \downarrow 0} [N(t + \Delta t^-) - N(t^-)]$
- $[\tau_0, \tau]$ is an interval of observation measured on the scale of process age t and when $\tau_0 = 0$, τ denotes the right-censoring or end-of-followup time
- $Y(t)$ indicates whether an individual is under observation at time t; often $Y(t) = I(\tau_0 \leq t \leq \tau)$
- $\{\bar{N}(t), 0 \leq t\}$, where $\bar{N}(t) = \int_0^t Y(u) dN(u)$, is the observable counting process
- $\bar{Y}_k(t) = Y(t)Y_k(t)$ indicates whether an individual is observed and at risk of a $(k+1)$st event at time t ; $\bar{Y}_{jk}(t) = Y(t)Y_{jk}(t)$
- $x(t) = (x_1(t), x_2(t), \ldots, x_p(t))'$ is a $p \times 1$ column vector denoting external, possibly time-dependent, covariates, which are assumed left-continuous
- $z(t) = (z_1(t), z_2(t), \ldots, z_p(t))'$ is a $p \times 1$ column vector used to denote $x(t)$ and possibly derived covariates
- $x^{(t)} = \{x(s) : 0 \leq s \leq t\}$ is the history of the covariate process at t and $x^{(\infty)} = \{x(s) : 0 \leq s\}$ is the realization of the entire covariate process
- $H(t) = \{N(s) : 0 \leq s < t\}$ is the history of the counting process at time t, which we sometimes also denote $N^{(t)}$
- $H(t) = \{N(s) : 0 \leq s < t; x(s) : 0 \leq s \leq t\}$ is also used
- An intensity function for an event process is denoted $\lambda(t|H(t))$; see Section 1.2
- The intensity function for the observable event process is $\bar{\lambda}(t|\bar{H}(t))$, where $\bar{H}(t) = \{\bar{N}(s) : 0 \leq s < t; Y(s) : 0 \leq s \leq t\}$; We have $\bar{\lambda}(t|\bar{H}(t)) = Y(t)\lambda(t|H(t))$ if $\Delta N(t)$ and $Y(t)$ are conditionally independent

- $S(t) = \Pr(T \geq t)$ is the survivor function and $F(t) = \Pr(T \leq t)$ is the cumulative distribution function for a random variable T; $h(t) = -d \log S(t)/dt$ is the hazard function. Sometimes we define $S(t)$=$\Pr(T > t)$
- $\mu(t) = E\{N(t)\}$ and $\mu(t|x) = E\{N(t)|x\}$ are mean functions
- $\rho(t) = d\mu(t)/dt$ and $\rho(t|x) = d\mu(t|x)/dt$ are rate functions; we often write $E\{dN(t)\} = \rho(t)dt$
- β is a vector of regression coefficients
- $\{Z(t), 0 \leq t\}$ is a multistate stochastic process where $Z(t)$ is the state occupied at t
- $\mathcal{Q}(t)$ is a $K \times K$ matrix of transition intensities for a multistate process with K states
- $\mathcal{P}(s,t)$ is a $K \times K$ matrix of transition probabilities for a multistate process with K states, with (k,ℓ) entry $p_{k\ell}(s,t) = \Pr(Z(t) = \ell | Z(s) = k)$
- Under a discrete time scale $t = 0, 1, 2, \ldots$, $n(t) = N(t) - N(t-1)$
- If $\{N(t), 0 \leq t\}$ is a multivariate counting process then $N_j(t)$ counts the number of events of type j and $N(t) = (N_1(t), \ldots, N_J(t))'$

Contents

1
Introduction

1.1 The Scope of Recurrent Events

In science and technology, interest often lies in studying processes which generate events repeatedly over time. Such processes are referred to as *recurrent event processes* and the data they provide are called *recurrent event data*.

In some settings interest may lie in a relatively small number of processes generating a large number of events. This is the case when studying stoppages to assembly lines, when analyzing processes for software fault detection and removal, or when investigating the incidence of injuries in manufacturing plants. In other settings data may be available for a larger number of processes exhibiting a relatively small number of recurrent events. These types of processes arise frequently in medical studies, where information is often available on many individuals, each of whom may experience transient clinical events repeatedly over a period of observation. Examples include the occurrence of asthma attacks in respirology trials, epileptic seizures in neurology studies, and fractures in osteoporosis studies. In business, examples include the filing of warranty claims on automobiles, or insurance claims for policy holders.

In this book we focus primarily, but not exclusively, on situations in which data are available from a large number of individuals under study. Frequent objectives in analyzing recurrent event data include (i) understanding and describing individual event processes, (ii) identifying and characterizing variation across a population of processes, (iii) comparing groups of processes, and (iv) determining the relationship of fixed covariates, treatments, and time-varying factors to event occurrence.

In most applications the scale used to characterize the distributions of events is in units of time, but other scales may be used. For example, early work on the development of models for point processes was in the textile industry where the unit of measurement was fibre length and the event of interest was a flaw in the fibre. Other time-related scales include distance driven for automobile warranty claims and hours or cycles of operation in manufacturing processes.

Before providing some notation and describing modeling concepts for recurrent events, we give some examples of recurrent event data.

1.2 Some Preliminary Examples

We present here several examples of recurrent event data, and introduce some simple graphical and numerical summaries that display basic features of the data. The data presented here are subjected to more formal methods of analysis in subsequent chapters.

1.2.1 Mammary Tumors in a Carcinogenicity Study

Gail et al. (1980) presented data from a carcinogenicity experiment on the times to the development of mammary tumors for 48 female rats. Rats were exposed to a carcinogen and further conditioned for 60 days prior to randomization to receive either a treatment or control. A followup period of 122 days began after randomization, during which they were examined every few days for the development of new tumors. The data given in Gail et al. (1980) are displayed in Table 1.1 except we report the times from the beginning of the period of examination instead of from the time of exposure to the carcinogen. Following each identifier are times corresponding to days on which new tumors were detected.

Figure 1.1 displays the data in separate *event plots* for each rat where dots are placed on the days that events occurred. In cases where there was more than one event for a given animal on a given day, the events are separated slightly. Such plots, which are feasible when there are not too many individual processes, convey an impression of the frequency and patterns of events. They also show the total followup times for each process; in this case they are all the same at 122 days, but in many settings they vary considerably.

Plots like Figure 1.1 have limitations, however, because it is often not easy to determine visually whether a trend or other pattern exists. Another useful plot is a *cumulative sample mean function plot*, defined as follows. Suppose that m individual processes are observed, with each process being observed over the time interval $[0, \tau]$. Let $N_i(t)$ represent the number of events over the time interval $[0, t]$ for the ith process. Then the *cumulative sample mean function* is

$$\widehat{\mu}(t) = \frac{1}{m} \sum_{i=1}^{m} N_i(t) \,. \tag{1.1}$$

A *cumulative sample variance function*

$$\widehat{\mathrm{var}}\{N(t)\} = \sum_{i=1}^{m} \{N_i(t) - \widehat{\mu}(t)\}^2/(m-1)$$

Table 1.1. Times to tumor (in days) for laboratory rats (numbers in parentheses indicate number of tumors detected).

Treatment Group		Control Group	
ID	Days of Tumor Detection	ID	Days of Tumor Detection
1	122	1	3, 42, 59, $61^{(2)}$, 112, 119
2	-	2	28, 31, 35, 45, 52, $59^{(2)}$, 77, 85, 107, 112
3	3, 88	3	31, 38, 48, 52, 74, 77, $101^{(2)}$, 119
4	92	4	11, 114
5	70, 74, 85, 92	5	35, 45, $74^{(2)}$, 77, 80, 85, $90^{(2)}$
6	38, 92, 122	6	$8^{(2)}$, 70, 77
7	28, 35, 45, 70, 77, 107	7	17, 35, 52, 77, 101, 114
8	92	8	21, 24, 66, 74, $101^{(2)}$, 114
9	21	9	8, 17, 38, $42^{(3)}$
10	11, 24, 66, 74, 92	10	52
11	56, 70	11	$28^{(2)}$, 31, 38, 52, $74^{(2)}$, $77^{(2)}$, $80^{(2)}$, $92^{(2)}$
12	31	12	17, 119
13	3, 8, 24, 35, 92	13	52
14	45, 92	14	$11^{(2)}$, 14, 17, 52, $56^{(2)}$, $80^{(2)}$, 107
15	3, 42, 92	15	17, 35, 66, 90
16	3, 17, 52, 80	16	28, 66, $70^{(2)}$, 74
17	17, 59, 92, 101, 107	17	3, 14, $24^{(2)}$, 28, 31, 35, 48, 74, 77, 119
18	45, 52, 85, 101, 122	18	21, 28, 45, 56, 63, 80, 85, 92, $101^{(2)}$, 119
19	92	19	28, 35, 52, 59, $66^{(2)}$, 90, 97, 119
20	21, 35	20	$8^{(2)}$, 24, 42, 45, 59, $63^{(2)}$, 77, 101, 119, 122
21	24, 31, 42, 48, 70, 74	21	80
22	-	22	92, $122^{(2)}$
23	31	23	21
		24	3, 28, 74
		25	24, 74, 122

may similarly be defined. Figure 1.2 shows plots of $\widehat{\mu}(t)$ and of $\widehat{\text{var}}\{N(t)\}$ for each treatment group. The left panel shows the cumulative sample mean functions to be roughly linear, with the control group having a little over twice as many events (tumors) per animal as the treatment group at any given time. The variance function plot shows the variability in cumulative tumor occurrences per animal, at different times.

Cumulative mean and variance functions may correspondingly be defined for recurrent event process models; this is discussed in Chapter 2. An important question, addressed in Chapter 3, is how to estimate a process mean or variance function when individual processes are not all observed over the same time interval, as they are here.

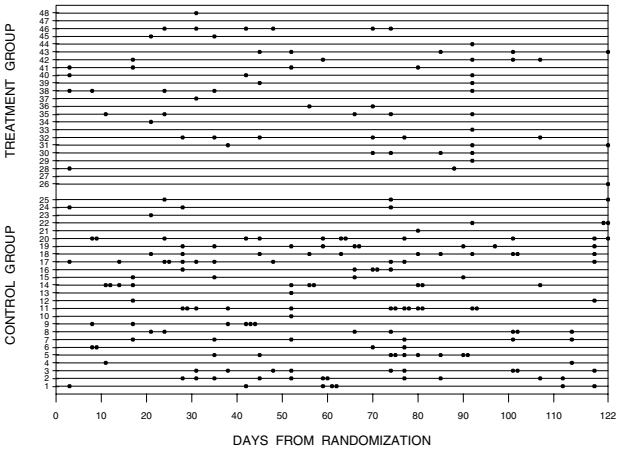


Fig. 1.1. Event plots for tumor occurrences in 48 rats.

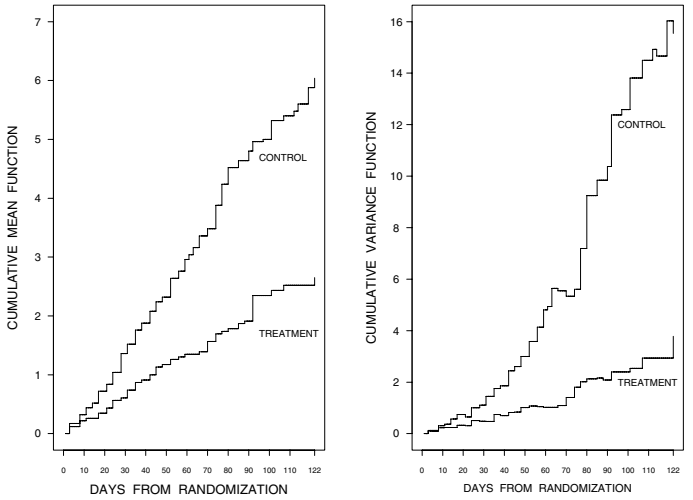


Fig. 1.2. Cumulative sample mean and variance functions for tumor occurrences in rats.

1.2.2 Testing and Debugging a Large Software System

Dalal and McIntosh (1994) describe a testing and debugging process for a large software system consisting of roughly seven million noncommentary source lines (NCSL) of code. During testing which was carried out by different individuals, the testers recorded the amount of time they spent each day on testing, and the number of faults found. Faults found were generally not fixed immediately, but instead were repaired at convenient times. When faults were repaired, and for other reasons, additional lines of code (NCSL) were added to the software system over time. Testing covered 160 calendar days, during which 870 faults were detected and over 342,000 new NCSL were added.

For purposes of analysis we treat this as a single large process, with fault detections being the events of interest. The most relevant time scale for assessing these events is the cumulative staff time spent testing the system. The data given by Dalal and McIntosh (1994), and reproduced in Table D.2 of Appendix D, show the cumulative staff days of testing (t) and the cumulative number of faults detected ($N(t)$) at the end of different test days. Because the introduction of new NCSL can affect the number of faults in the system, the cumulative number of source lines added during the testing period is also given.

Figure 1.3 is a plot of $N(t)$ versus t, which shows that the rate of fault detection has decreased over the testing period. A major issue for the software developers is when to stop testing. This decision problem is discussed in Section 3.8.3; an important factor is prediction of the number of new faults that would be detected if the testing period were extended.

1.2.3 Pulmonary Exacerbations in Cystic Fibrosis

Therneau and Hamilton (1997) discussed data that arose in a clinical trial involving persons with cystic fibrosis (Fuchs et al. 1994). These individuals are susceptible to an accumulation of mucus in the lungs, which leads to pulmonary exacerbations and deterioration of lung function. In a randomized clinical trial, a purified recombinant form of the human enzyme DNase I, called rhDNase, was administered daily to patients in an rhDNase treatment group and the remaining patients were administered a placebo; patients and their physicians did not know which treatment (rhDNase or placebo) they were receiving. Most subjects were followed for approximately 169 days, and the occurrences of exacerbations over the study period were recorded for each. Subjects had as many as five exacerbations, and Table 1.2 shows the total number experienced by the 324 and 321 subjects in the placebo and rhDNase groups, respectively.

A main objective of the study was to compare the two treatment groups in terms of exacerbation occurrence. Because the subjects were randomly assigned to treatment and the length of followup for almost all subjects was close to 169 days, a simple comparison could be based on the counts in Table

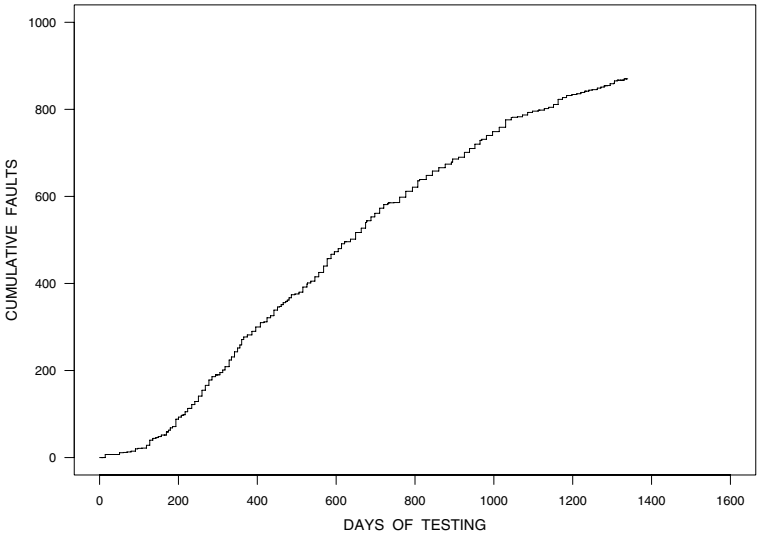


Fig. 1.3. Cumulative software faults detected versus staff days of testing.

1.2. The patterns of exacerbations are also of interest, however, and there is in addition a baseline measurement of forced expiratory volume taken for each subject at the start of the study; this is a measure of lung function and is expected to be related to the frequency of exacerbations. This suggests that some type of regression modeling would be useful.

Table 1.2. Distribution of the numbers of exacerbations by treatment group for subjects in the rhDNase study.

Number of	Number of Patients	
Exacerbations	Placebo Group	rhDNase Group
0	185	217
1	97	65
2	24	30
3	13	6
4	4	3
5	1	0

A final noteworthy point is that when an exacerbation occurs, the subject receives antibiotics and is not considered at risk for a new exacerbation until the end of antibiotic therapy. The majority of periods of antibiotic therapy are in the 10–15-day range, but there is considerable variation in their lengths. With the event defined as the onset of an exacerbation, this creates a complication with the construction of cumulative plots such as Figures 1.2 and 1.3, and with model-based analysis. The complication is easily addressed by careful consideration of when individuals are at risk of an exacerbation. Periods "at risk" are discussed in Section 1.4.2.

1.2.4 Automobile Warranty Claims

Manufacturers whose products are covered by warranties collect and track information on warranty claims. This is done for financial reasons, because claims incur costs to the manufacturer, which must be understood and predicted, but also because claims data reflect certain dimensions of product quality and reliability, and may suggest areas for improvement. Lawless and Nadeau (1995) examined data on cars of one model year and type, which were sold over a period of 60 weeks. A slightly updated database with 38,401 vehicles generating 5760 claims is considered in this book. The warranty data are for one system on the vehicle, and there were one-year and 12,000-mile limits on coverage. At the time the claims database was finalized, not all vehicles had been in service for 365 days, and for simplicity we consider in this section only those cars which had; there were 15,775 such cars. All vehicles are considered in later analyses in Chapter 3.

There were 2620 claims among the 15,775 cars, and Table 1.3 shows the frequency distribution of claims per vehicle. It is generally of interest to examine warranty claims according to the date of manufacture of the product, because quality problems occasionally arise over certain limited periods. Figure 1.4 shows a plot of the 2620 claims across the 15,775 cars. The cars in the plot are numbered from 1 to 15,775 according to their date of manufacture; the manufacturing period covered 209 days. The event (claim) "times" are the age of the car at the time of the claim, where age is the number of days since the vehicle was sold. It is noted that some claims in the figure are at age zero; this is because claims were made by the dealer before the car was sold. This figure is analogous to the the plot used in Figure 1.1 for a much smaller number of individual processes.

As with Figure 1.1, certain patterns are suggested by Figure 1.4, but summaries of the data provide a clearer picture. We could, for example, show a cumulative sample mean function plot, as in Figure 1.2. As another illustration, we show in Figure 1.5 a histogram of the 2620 claim times, with times (ages) grouped into 20-day intervals. Smoothed histograms, or data density plots, could similarly be given. It is noted that the claims frequency is lower for the higher ages. The main reason for this is that many cars are no longer

Table 1.3. Frequency distribution of car warranty claims.

Number of Claims	Number of Cars (frequency)
0	13,987
1	1,243
2	379
3	103
4	34
5+	29
Total	15,775

covered by the warranty at ages close to 365 days because they passed the 12,000-mile limit some time earlier.

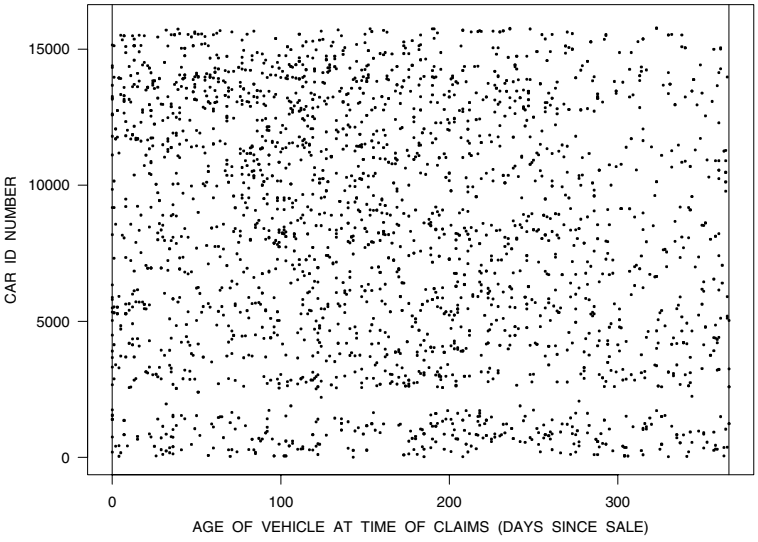


Fig. 1.4. Warranty claim occurrences for 15,775 cars.

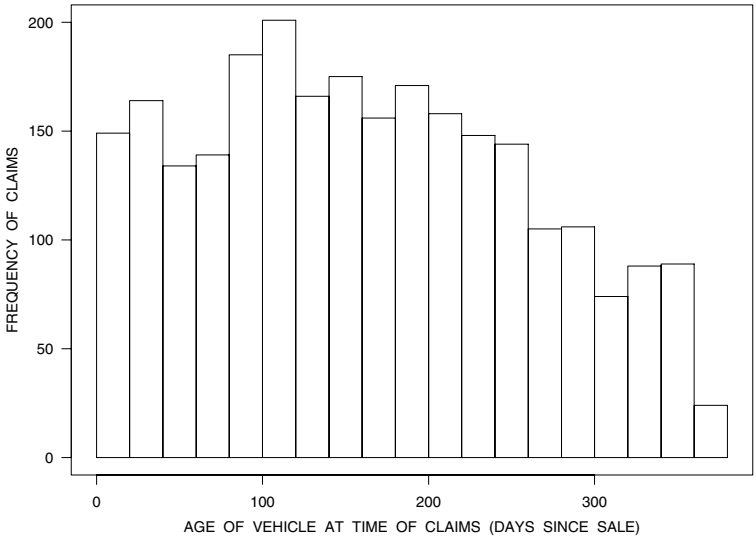


Fig. 1.5. Histogram of warranty claim occurrence times (ages).

1.3 Notation and Frameworks

Modeling of recurrent events can be approached in a number of ways, which are described in books on stochastic processes and, more specifically, point processes. For purposes of both modeling and statistical analysis, the concepts of intensity functions and *counting processes* are especially useful. In the following discussion and throughout the book we consider events that occur in continuous time, but discrete time models are also considered in specific places.

For a single recurrent event process starting for simplicity at $t = 0$, let $0 \leq T_1 < T_2 < \cdots$ denote the event times, where T_k is the time of the kth event. The associated *counting process* $\{N(t), 0 \leq t\}$ records the cumulative number of events generated by the process; specifically, $N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$ is the number of events occurring over the time interval $[0, t]$. More generally, $N(s, t) = N(t) - N(s)$ represents the number of events occurring over the interval $(s, t]$.

In this framework, note that we use square and round brackets to indicate whether the endpoint of an interval is in or not in the interval, respectively. In addition, we use t^- and t^+ to denote times that are infinitesimally smaller or larger than t, respectively. As defined here, counting processes are right-continuous; that is, $N(t) = N(t^+)$. Figure 1.6 portrays a realization of an event process in terms of its counting process.

Models for recurrent events can be specified very generally by considering the probability distribution for the number of events in short intervals $[t, t + \Delta t)$, given the history of event occurrence before time t. To set up some essential notation, we let $\Delta N(t) = N(t + \Delta t^-) - N(t^-)$ denote the number of events in the interval $[t, t + \Delta t)$, and let $H(t) = \{N(s) : 0 \leq s < t\}$ denote the *history* of the process at time t. For events occurring in continuous time we make the mathematically convenient assumption that two events cannot occur simultaneously. Then, the event *intensity function* gives the instantaneous probability of an event occuring at t, conditional on the process history, and defines the process mathematically. The intensity is defined formally as

$$\lambda(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N(t) = 1|H(t)\}}{\Delta t} . \tag{1.2}$$

Throughout the book we use intensity functions to model event processes. Section 2.1 develops mathematical background and relates intensity functions to other characteristics of a recurrent event process. The assumption that two events cannot occur at exactly the same time is plausible in most settings and is retained throughout the book. Processes that allow simultaneous occurrences can be handled using approaches described in Section 8.1.

When a heterogeneous group of individuals or processes is considered, the assumption of a common event intensity may be implausible. Greater generality can be obtained by broadening the definition of the process history to include information on fixed or time-varying covariates and by letting the event intensity function depend on such covariates. Covariates are discussed in Section 1.3.4 and throughout the book.

The definition (1.2) is very general and accommodates any possible dependence of the intensity on the process history $H(t)$. Models which make explicit assumptions about the dependence of $\lambda(t|H(t))$ on $H(t)$ are used throughout this book to facilitate analysis. Two fundamental ways of describing and modeling event occurrences are through *event counts* and through *gaps* or *waiting times* between successive events. These are discussed in the following subsections. Sometimes the most natural framework is clear for a particular problem but it may be driven by features of the underlying process, the objectives of analyses, or the results of model checking. In many contexts, analyses based on both counts and waiting times may be relevant.

Two process features that are often of interest are time trends and event clustering. Broadly, a *time trend* in a process refers to a tendency for the rate of event occurrence to change over time in some systematic way. Monotone trends are common, but nonmonotonic trends can also occur. For example, seasonal fluctuations in the occurrence of bronchial infections have been noted in respiratory disease. Figure 1.3 in Section 1.2 shows a decreasing trend in the rate of detection of software faults as testing time increases and Figure 1.5 indicates a decreasing trend in the population rate of car warranty claims as the ages of vehicles approach one year. Figure 1.2, on the other hand, shows

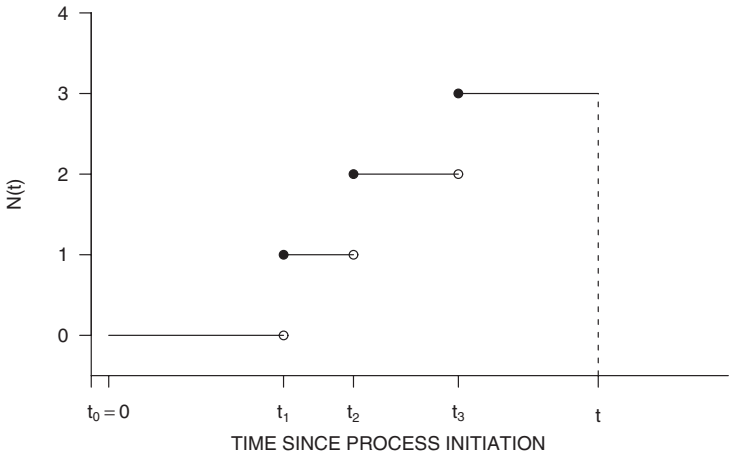


Fig. 1.6. Counting process representation of data on recurrent events.

an absence of trend in the average rate of occurrence of tumors in the animal carcinogenicity study described in Section 1.2.1.

Clustering refers to the tendency for events to cluster together in time. To some extent clustering is similar to nonmonotonic fluctuations in the event occurrence rate; the broad distinction is that clusters of events (i.e. several events close together in time) are considered to occur randomly in time whereas trends in the event rate are more time related. Time trends are rather easily built into models discussed below, but dealing with clustering is often more complicated.

1.3.1 Methods Based on Event Counts

Models and methods based on counts are often useful when individuals frequently experience the events of interest, and the events are "incidental" in the sense that their occurrence does not materially alter the process itself, either directly or through resulting interventions. Examples of incidental events include mild epileptic seizures or asthmatic attacks in humans and most warranty claims for cars. Events which may recur but which are not incidental include myocardial infarction and stroke in cardiovascular studies, or the development of new sites of metastatic disease in cancer trials. In these latter settings the events may substantially alter the condition of the individual, thus affecting the event process in the future; physicians may also alter treatment in response to these events.

The canonical framework for the analysis of event counts is the *Poisson process*. Poisson models typically use calendar time or the age of the process as the time scale. The independent increments property of Poisson processes states that $N(s_1, s_2)$ is independent of $N(s_3, s_4)$ provided $s_2 < s_3$. This implies that for Poisson processes the process history at time t does not affect the instantaneous probability of events at time t, and in the absence of covariates the only factor determining the intensity is t. Poisson processes are therefore *Markov*, with intensity function of the form

$$\lambda(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N(t) = 1\}}{\Delta t} = \rho(t)\,. \tag{1.3}$$

In addition to being the intensity function, $\rho(t)$ is the *rate function* giving the marginal (i.e. unconditional) instantaneous probability of an event at time t. Specifically, $\rho(t)\Delta t \doteq E\{\Delta N(t)\}$, and if $\mu(t)$ denotes the expected cumulative number of events at t, then

$$\mu(t) = E\{N(t)\} = \int_0^t \rho(s)ds \tag{1.4}$$

and $\rho(t) = \mu'(t) = d\mu(t)/dt$. Sections 1.2.1 to 1.2.4 all described settings where event counts are relevant.

Extensions to accommodate between-subject variability in event rates through fixed or time-varying covariates, or random effects are straightforward. Poisson processes and extensions are discussed extensively in Section 2.2 and in Chapter 3.

1.3.2 Methods Based on Waiting or Gap Times

We define $W_j = T_j - T_{j-1}$ as the *waiting time* or *gap time* between the $(j-1)$st and jth event. Analyses based on waiting times are often useful when events are relatively infrequent, when some type of individual renewal occurs after an event, or when prediction of the time to the next event is of interest. Analyses based on waiting times are natural in studies of system failures, where repairs made at each failure return the system to a working state. In some studies involving bladder cancer, patients undergo transurethral resection to remove all recently detected tumors, and prophylatic treatment to retard the development of new tumors; this intervention makes waiting time analyses reasonable. Other settings include studies of cyclical phenomena where characterization of cycle length is of interest. These include studies of recurrent phenomena such as infections, where an individual returns to a similar state after the infection has been cleared, and recurrent episodes of hospitalization or disability.

Renewal processes are the canonical models for waiting times and are defined as processes for which

$$\lambda(t|H(t)) = h(t - T_{N(t^-)}) \,. \tag{1.5}$$

That is, $h(\cdot)$ is the hazard function for the gap times between events, which are independent and identically distributed. Generalizations of renewal processes that accommodate within-subject association or trends in gap times are often useful. Renewal processes and extensions are discussed in Section 2.3 and in Chapter 4.

Table 1.4. Lengths of successive bowel motility cycles.

Individual	Complete Observed Periods									Censored
1	112	145	39	52	21	34	33	51		54
2	206	147								30
3	284	59	186							4
4	94	98	84							87
5	67									131
6	124	34	87	75	43	38	58	142	75	23
7	116	71	83	68	125					111
8	111	59	47	95						110
9	98	161	154	55						44
10	166	56								122
11	63	90	63	103	51					85
12	47	86	68	144						72
13	120	106	176							6
14	112	25	57	166						85
15	132	267	89							86
16	120	47	165	64	113					12
17	162	141	107	69						39
18	106	56	158	41	41	168				13
19	147	134	78	66	100					4

Example: Study of Motility of the Small Bowel

Aalen and Husebye (1991) discuss the analysis of data from a study of small bowel motility involving 19 healthy individuals. Catheters positioned in the proximal small bowel were used to monitor intraluminal pressure. Individuals were examined continuously from 5:45 p.m. to 7:25 a.m. on the next day, giving a total of 13 hours and 40 minutes of observation. A standardized mixed meal of 405 kcal was given to each individual at 6:00 p.m. to induce contraction of the small bowel, a key feature of the digestive process. After a variable length of time in which irregular bowel contractions occur, a "fasting state" begins, with a cyclical bowel motility (activity) pattern. The time between two consecutive fasting cycles is called the migrating motor complex (MMC)

period. In this study, the start of the bowel motility cycles represents the recurrent events, and it is of interest to analyze the durations of successive cycles for an individual. These duration times are shown in Table 1.4. Note that the end of the monitoring period leads to a censored, or incomplete, observation of the last MMC period. Analysis of these data is considered in Section 2.3.2 and in Chapter 4.

1.3.3 More General Models

It is sometimes convenient to consider models for recurrent events in terms of multistate processes such as the one given in Figure 1.7. The states in Figure 1.7 correspond to the cumulative number of events experienced by a subject since the onset of the process. The event intensity function can then be viewed as a "transition" intensity function. Generalizations of the Poisson and renewal models discussed earlier are naturally formulated. For example, one might adopt a Markov model $\lambda(t|H(t)) = \alpha_k(t)$ where $N(t^-) = k$, for which the Markov transition intensity depends on the state occupied and the time since the start of the process. This model generalizes the Poisson process for which $\alpha_k(t) = \rho(t)$, $k = 0, 1, \ldots$. Alternatively, one could adopt a semi-Markov model of the form $\lambda(t|H(t)) = h_k(B(t))$ where again $N(t^-) = k$ and $B(t) = t - T_{N(t^-)}$. As in a renewal process, the time scale is based on the time since entry to the current state, but here the distributions of the sojourns in the states are not identical. This represents a generalization of a simple renewal process for which $h_k(s) = h(s)$, $k = 0, 1, 2, \ldots$. More general intensity-based models where $\lambda(t|H(t))$ is affected by previous event history are likewise conveniently interpreted within the framework of Figure 1.7.

Fig. 1.7. Multistate representation of a recurrent event process.

1.3.4 Covariates

In many applications an objective is to relate event occurrence to fixed or time-varying covariates, or to adjust an analysis for the presence of covariates. In the car warranty claims data of Section 1.2.4, for example, we may want to relate the occurrence intensity or rate for claims to the time period in which the car was manufactured. In Sections 1.2.1 and 1.2.3, studies were described in which the comparison of treatment and control groups of subjects was of interest. In such cases it is customary to use an indicator covariate to

represent the group to which an individual belongs. In the study in Section 1.2.3, the frequency of pulmonary exacerbations for subjects is also related to the person's forced expiratory volume measurement at the time of randomization, so it could be included as a covariate in the analysis of the data. Time-varying covariates are also common. For example, in studies on the frequency of visits to hospital emergency clinics because of breathing problems, air pollution measures, temperature, and humidity may be important covariates. In the software debugging example of Section 1.2.2, the number of lines of code changed in response to previous faults that were detected generally affects the fault intensity function, so it is in effect a time-varying covariate.

We typically use x or z to denote fixed covariates, and $x(t)$ or $z(t)$ for time-varying covariates. An important distinction with a time-varying covariate is whether it is external or internal. An *external covariate* is one whose values are determined independently of the recurrent event process (fixed covariates are therefore external). A covariate that is not external is called *internal*. Thus, air pollution is an external covariate in a study on hospital visits due to breathing problems. However, the number of lines of code changed in a software debugging process is an internal covariate, because it depends on prior events (i.e. faults detected) in the process.

Covariates in models for recurrent events are discussed in Chapter 2.

1.3.5 Factors Influencing Model Choice

Recurrent event analyses may have a range of objectives that are determined by the setting, and this guides the formulation of models and methods for analysis. In trials where patients are randomized to treatments upon study entry and prospectively followed, for example, simple comparisons between two or more treatment groups are often of interest based on marginal features. Analyses based on expected event counts are appealing in such settings, because they provide a basis for simple treatment comparisons which exploit the randomization and facilitate causal inferences regarding treatment effects.

In contrast, in prospective observational studies, interest may lie in developing intensity-based models in order to gain a better understanding of factors that drive the event process. Fully specified models are also required when models are to be developed for simulation purposes, or when probabilistic predictions of future events are desired. Fixed and time-varying covariates are often relevant in such settings.

One modeling distinction is between full probability models for the recurrent event processes and partial specifications that model only certain aspects such as the mean function $E\{N(t)\}$. A second distinction is between parametric and nonparametric specifications. As noted above, full models are needed for certain purposes and this often involves fully parametric specifications. Conversely, simple comparisons of treatments or of groups of individuals is often best done with easily interpreted, robust analysis of marginal features that does not make many strong assumptions. The completeness and type of

data are also factors; data with missing components often require more assumptions for their analysis. We present a full range of methods for analysis including parametric, semiparametric, and nonparametric methods, as well as methods based on full and partial model specifications.

1.4 Selection of Individuals and Observation Schemes

In planning an event history study, one must decide on two key aspects: (i) how individuals will be selected for the study, and (ii) what information will be collected about the event histories for the individuals in the study.

An initial important distinction is between *prospective* and *retrospective* observation of individuals. In a *prospective study* we take one or more individual processes and follow them longitudinally over time; events and related outcomes that are recorded during this "followup" period are the responses of interest. Prospective studies can be *experimental* or *observational*. In the former, individuals selected for the study are often a random sample from some population, but the defining characteristic is that control is exercised over the assignment of treatments or other experimental factors. In observational studies no such control is exercised. In a *retrospective study*, the time period $[\tau_0, \tau]$ over which events are recorded for an individual is physically prior to the time the individual is selected, so they are inherently observational. Some studies involve both retrospective and prospective observation.

Sections 1.2.1 and 1.2.3 give examples of prospective experimental studies in which individuals are randomly assigned a treatment and then followed for a specified period of time. Sections 1.2.2 and 1.2.4 describe observational studies; the former involves a single process involving the detection of faults during software development and testing and the latter involves warranty claims on a large number of cars. These two studies could technically be either prospective or retrospective, depending on when the data were assembled. However, both are essentially prospective here because the processes were identified a priori and data were collected as they occurred. An example of a retrospective study is where women are selected randomly from some population, and their history of pregnancies and live births determined. With such studies a key issue is whether sufficiently good records exist to allow the history of past events and covariates to be determined. Selection bias may also arise, as discussed in Section 1.4.3.

In Section 2.6 we discuss the specification of likelihood functions or sampling distributions, and how they are affected by the two study aspects above. In the following subsections we introduce some related issues.

1.4.1 The Choice of Time Scale

In recording, modeling, and analyzing the occurrence of events it is necessary to have an appropriate time scale. The time variable t is often chronological or

calendar time, especially with processes that apply to humans or animals. In technological areas, measures of usage or exposure are often used, for example distance accumulated for motor vehicles or (discrete) usage cycles for printers or copiers. In the software testing illustration in Section 1.2.2, cumulative staff days of testing were deemed most appropriate.

The time scale also involves a choice of origin, and this requires some care when multiple individuals are under study. Intensity-based analyses can adapt to the choice of a time origin through specification of the intensity, but it is nevertheless desirable to use an origin that is consistent across individuals and facilitates interpretation and analysis. In many contexts this may be clear. For example, in studies of car warranty claims the time origin for a car would correspond to its purchase date, and analyses could be based on calendar time or usage since purchase. In health research, the choice of time origin is frequently less clear. Consider a study directed at modeling the development of newly damaged joints among patients in an arthritis clinic. Possible time origins include the time of birth of the patient (with age as the time scale), the time of disease onset, or the time of entry to the clinic. In many senses disease onset is a natural choice, although this is often difficult to determine precisely; incorporating patient age into analyses as a covariate would then be sensible. In randomized clinical trials, it is customary to consider the time of randomization or start of treatment as the time origin. This is usually reasonable because interest typically lies in making treatment comparisons, but when interest lies in features of the disease process other time origins may be preferred. It should also be noted that once an underlying time scale is chosen, it is necessary to decide whether it is most suitable to develop models based on the cumulative time or gap times between events. Although this could be viewed as a model specification decision, it affects the analysis and interpretation of results.

In studies, data are collected over some calendar time period, and different individuals are not necessarily observed over the same intervals for their event processes. That is, for individual i, we typically observe $\{N_i(t) : \tau_{i0} \le t \le \tau_i\}$, for an interval $[\tau_{i0}, \tau_i]$ that is determined by the calendar time period for the study and the time origin for that individual's event process. The examples in Section 1.3 all involved situations where $\tau_{i0} = 0$ for each individual. However, for settings such as the arthritis study mentioned above, we may have $\tau_{i0} > 0$, depending on the choice of time origin.

Figure 1.8 shows the relationship between the process age (t) and calendar time for two individuals with processes initiating at C_{10} and C_{20} in calendar time and observed in a study over calendar time period $[C_S, C_E]$. Unless stated otherwise, in the subsequent development the term "time" refers to the time scale t for individual event processes.

Sometimes it is useful to define more than one time scale. In studies of car warranty claims, for example, one might wish to develop models with time scales for both usage and age of the vehicle. Intensity models incorporating both scales can provide insight into which is more appropriate for a given

problem. Similarly, intensity functions that involve both cumulative time and gap times are often useful.

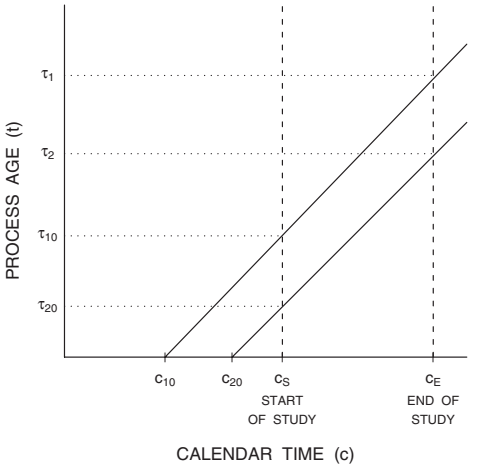


Fig. 1.8. Lexis diagram relating calendar time to process time t.

1.4.2 Defining Periods "At Risk"

Let us denote the time period over which an individual's events are recorded as $[\tau_0, \tau]$ as in the preceding section. For now we suppose events occur and are recorded in continuous time. Let $I(\cdot)$ be an indicator function such that $I(A) = 1$ if A is true and $I(A) = 0$ otherwise and define $Y(t) = I(\tau_0 \leq t \leq \tau)$. Thus, $Y(t) = 1$ if an individual is under observation and therefore at risk of events being observed at time t, and $Y(t) = 0$ otherwise. The function $Y(t)$ is called an "at risk" indicator and is useful for denoting which individuals provide information about event occurrence at a given time. The time τ is sometimes referred to as a censoring or end-of-followup time for the observed event process.

More general observational or censoring patterns can arise if subjects temporarily cease to be under observation. This happens, for example, if individuals are asked to record events on daily diary cards and stop doing so for a period of time. It is also possible for an individual to cease to be at risk temporarily because of the nature of the process. For example, in the clinical trial involving pulmonary exacerbations described in Section 1.2.3, an individual

who has an exacerbation is not at risk for another until a period of antibiotic treatment has ended. Such observational schemes and processes are discussed further in Section 2.6 and in Chapter 6.

1.4.3 Initial Conditions and Selecting Individuals for Study

Suppose the individual event processes of interest are considered to start at time $t = 0$. The simplest type of study is where individuals are followed prospectively starting from time $\tau_0 = 0$. In some settings, however, an individual is sampled and observation begins from a time τ_0 after the event process has begun, with events observed over the time interval $[\tau_0, \tau]$; see Figure 1.8. Moreover, the selection of the individual may depend on events prior to τ_0. If information on the history of the process over $[0, \tau_0)$ is available, then analyses based on intensity functions for the period $\tau_0 \leq t \leq \tau$ are easy to carry out as discussed in Section 2.6. Difficulties can arise, however, if all the necessary history over $[0, \tau_0)$ is not available. The information for the interval $[0, \tau_0)$, including the values of any covariates, is sometimes referred to as the *initial conditions* for the process $\{N(t), \tau_0 \leq t\}$.

It is always important to be aware of any distinction between the *target population*, which is the desired reference population for a study, and the actual *study population* from which the individual processes in the study are drawn. This can be problematic in observational studies.

In some settings the mechanism by which individuals are sampled or chosen for a study is not fully specified. For example, consider a registry database of osteoporosis patients attending one of a number of tertiary care clinics. Information on the incidence of fractures, medication uses, and health care resource utilization may be collected. Simple analyses of fracture rates may be of interest, but for inferences about the general population such analyses should incorporate the mechanism for referral to the tertiary care clinics and the process which determines when individuals are sent to them.

If selection of individuals is completely independent of their event processes then analyses involving full or partially specified models are both relatively straightforward. If selection of an individual at τ_0 depends on $H(\tau_0)$, then this must often be taken into account in the analysis by, for example, modeling the intensity function for $t > \tau_0$. In retrospective studies, where an individual is chosen for study after time τ_0 and perhaps even after time τ, care must be taken to reflect any conditions for selection in the analysis. In particular, the inclusion of an individual in a study may depend on her event history over $[\tau_0, \tau]$ or on her satisfying a certain condition (e.g. being alive) at time τ. These issues are discussed in subsequent chapters and, in particular, in Section 7.3.

1.4.4 Intermittent Observation and Interval Censoring

Frequently it is difficult to observe the precise times of events and all that is known is how many events occurred between successive examination times. In studies of osteoporosis, for example, interest may lie in the occurrence of asymptomatic fractures which are only detectable upon radiographic examination at specific followup times. If examinations occur at common times for each patient, analyses can be easily based on the interval event counts (i.e. numbers of fractures), although considering gap times will be difficult. However, if examination times vary between patients then the times between assessments must be taken into account; we refer to this as *interval-censored data*. Such data often arise in medical contexts, such as studies of metastatic cancer where new metastases are detectable upon magnetic resonance imaging, in radiographic studies of joint damage in arthritic patients, and in studies of the development of tumors in superficial bladder cancer. Similar phenomena occur in studies on the degradation of systems or materials, for example, in the detection of cracks in metal surfaces during intermittent inspections. Problems involving *intermittent inspection* are discussed in Chapter 7.

1.5 Multitype Event Data

1.5.1 Multivariate Event Processes

Multiple types of recurrent events arise frequently. In some cases the events represent different severities of the same phenomenon (e.g. mild, moderate, or severe epileptic seizures) or subtypes thought to be caused by different underlying physiological processes (e.g. asthma exacerbations may be classified according to the concentration of different cell counts in sputum samples), or they may correspond to different types of problems (e.g. automobile insurance claims may be due to theft, breakdown, collisions, etc.).

In some cases interest may lie in differentiating these events because they have different implications; for example, severe epileptic seizures have more serious consequences than milder ones. Interest may also lie in comparing or understanding the relation between events; one may, for example, investigate whether individual asthma patients tend to have predominantly one type of exacerbation, whether one type of exacerbation increases the risk of the other, or whether the same environmental factors are associated with both types of exacerbations.

Multiple event types are discussed at length in Chapter 6. The following subsections introduce some important special cases.

1.5.2 Recurrent Events with Termination

A special case of multiple types of events arises when a recurrent event process is terminated by another event. An example is in organ transplant studies

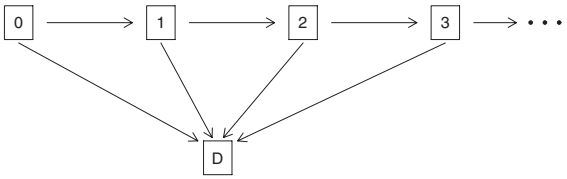


Fig. 1.9. Multistate diagram for recurrent events with a terminal event.

where transient graft rejection episodes are terminated by total graft rejection or patient death. In studies of cancer patients with bone metastases, the development of new metastases is terminated by patient death. Intensity-based models handle this type of recurrent event data relatively easily. If marginal process characteristics such as the expected number of recurrent events are of interest, however, analyses are less straightforward if the terminating event is related to the recurrent events. This is discussed in Section 6.6.

Example: Skeletal Complications from Bone Metastases

Breast cancer patients frequently develop bone metastases over the course of the disease. These bone lesions reduce the integrity of the bone thereby increasing risk of fractures, spinal cord compression, and bone pain, as well as the need for clinical interventions including radiation therapy and orthopedic surgery. An international multicenter randomized trial of 380 breast cancer patients with bone metastases was designed to evaluate a bisphosphonate, pamidronate, for the prevention of skeletal-related events which included fractures, spinal cord compression, need for radiation, and need for surgery (Hortobagyi et al., 1998). Patients were randomized in a balanced fashion to receive pamidronate or placebo medication via monthly infusions. Data on the incidence of skeletal complications and survival were observed prospectively and the exact dates of all these events were recorded. Figure 1.9 is a multistate diagram reflecting the event process in which death is the absorbing state D, precluding the occurrence of future skeletal events. Analyses which distinguish between different types of skeletal events (e.g. fractures, the need for radiotherapy for bone pain) may also be conducted; these are considered in Section 6.7.

1.5.3 Recurrent Episodes

Another special case involving multiple event types is when the events correspond to the onset and termination of a relapsing and remitting condition. This setting may be more appropriately characterized as recurrent episodes

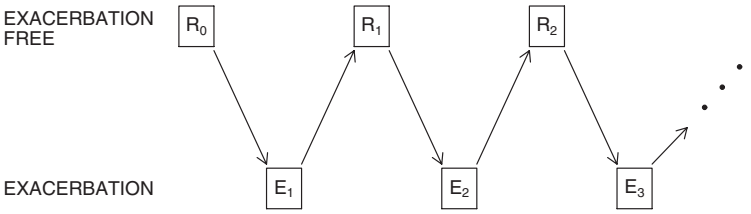


Fig. 1.10. Multistate diagram for recurrent episodic conditions.

or alternating states rather than recurrent events, but models and methods similar to those for ordinary recurrent events can be used for analysis. For example, Section 1.2.3 discussed a study on patients with cystic fibrosis who are at risk of recurrent pulmonary exacerbations. When an individual experiences an exacerbation, therapeutic interventions are provided to bring about the resolution of symptoms, which then typically clear up after a relatively short time. Figure 1.10 represents this type of process as a sequence of alternating states where R_0 is the initial state and E_k and R_k represent the states of k episodes and recoveries, respectively, $k = 1, 2, \ldots$. Similar phenomena arise in studies of outbreaks of herpes simplex virus, exacerbations in bronchitis, recurrent hospitalizations among psychiatric patients, and failures in systems which must be shut down in order for repair to take place. For such data, the history $H(t)$ is the collection of all event times and types over $[0, t)$ and the intensity function governing transitions between states can then be defined. Sometimes interest may lie only in factors related to the intensity for the onset of exacerbations and with a suitably defined "at risk" indicator, analyses can be relatively straightforward. This topic is discussed in Section 6.5.

Example: Exacerbations in Chronic Bronchitis

Here we consider a multicenter randomized trial designed to examine the effect of a quinolone, Ciprofloxacin, versus standard care on the occurrence and resolution of acute exacerbations of chronic bronchitis (Grossman et al., 1998). In this study clinic visits for subjects were scheduled at three-month intervals from the time of randomization, as part of a regular one-year followup assessment program. In order to enter the study and be randomized, patients were required to be experiencing an exacerbation. In addition, patients were required to visit a participating clinic when they perceived that a new exacerbation had begun or when an exacerbation had been resolved. As a result it was possible to determine the times of the onset and resolution of the exacerbations prospectively, with the exception of the first one; the time

of onset for this exacerbation was determined retrospectively at the time of study entry.

One hundred and fifteen patients were randomized to take Ciprofloxacin and 107 were randomized to receive standard care upon the development of symptoms. The average duration of followup was 357 and 350 days for the Ciprofloxacin and standard care groups, respectively. Figure 1.11 displays the profiles of eight patients from the standard care arm in which the black regions represent exacerbations during which treatment was taken, and the white regions exacerbation-free periods. From this plot it is evident that there is considerable variability in the frequency and duration of exacerbations between patients. Additional issues arising in this study include the fact that patients entered during an exacerbation and this must therefore be handled differently than the subsequent exacerbations. For example, the duration of the first exacerbation is subject to "length-biased" sampling; individuals with longer exacerbation episodes are more likely to be selected for the study. Moreover, the reliability of the reported onset times for this exacerbation may be poor compared to the times for the subsequent exacerbations. Furthermore, patients randomized to receive Ciprofloxacin received it after this first exacerbation had been present for some time, rather than at the start of the exacerbation. We consider the analysis of data from this study in Section 6.7.

1.6 Some Other Aspects of Analysis and Design

As discussed in the preceding sections, choices involving time scales and frameworks for analysis must be made. These are driven by the objectives of the study and to some extent by constraints on resources for the selection and observation of the processes of interest. In subsequent chapters we cover a broad range of models and methods of analysis, and consider applications from various fields.

Many of the methods of analysis described in this book can be implemented using widely available software for survival or lifetime data analysis. Applications involving S-PLUS or R software are discussed within chapters and some sample code is provided in Appendix C; SAS and other packages are mentioned in Appendix B. Models that cannot be handled by survival software are readily fitted using general-purpose optimization software that is available in S-PLUS, R, SAS, and many other statistical or mathematical computation packages. Illustrations are provided within chapters, with general discussion in Appendix B.

Most of the book is about methods of analysis, but the planning of studies involving recurrent events is discussed in Chapter 8. Key issues are sample size, and the type of followup of individual processes that resources permit. Increased attention is being given to longitudinal surveys (e.g. Korn and Graubard 1999; Lawless 2003b) in which individuals are selected according to a complex survey design and then interviewed at rather widely spaced times

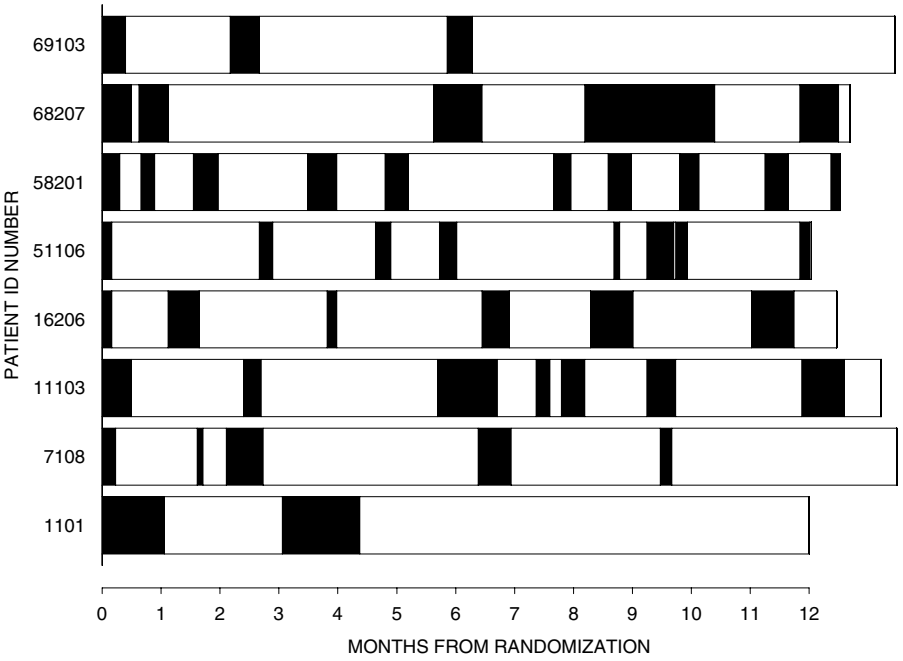


Fig. 1.11. Profiles of eight sample patients from a bronchitis study.

(e.g. annually) for the collection of information on events and other features. These and other studies lead to interesting methodological questions regarding design, clustering, missing data, and measurement error associated with covariates or the timing or type of events. We provide some limited discussion of these topics in Chapter 8.

1.7 Bibliographic Notes

Recurrent events are of interest in many fields of study. Much of the early statistical history of recurrent events deals with single processes or populations, for example, the emission of particles from a radioactive source, the occurrence of earthquakes, or the occurrence of accidents or cases of disease in a human population. Models for such settings were developed under the heading of point processes, and books on this topic provide extensive probabilistic developments and many examples of applications; see, for example, Cox and Lewis (1966), Cox and Isham (1980), Daley and Vere-Jones (1988, 2003), Lewis (1972), and Snyder and Miller (1991). Certain models, notably Poisson and renewal processes, were also studied in connection with the reliability of repairable systems (Ascher and Feingold, 1984; Rigdon and Basu,

2000) and of software (Singpurwalla and Wilson, 1999). Jiang et al. (2006) review and reference several types of applications to repairable systems.

More recently, the modeling and analysis of recurrent events for multiple individuals or systems has undergone extensive development (Andersen et al., 1993; Lawless, 1995; Nelson, 2003) and methods that incorporate inter-individual variability through covariates or random effects have become important. Application areas motivating these developments include medicine (Gail et al., 1980; Prentice et al., 1981; Byar, 1980), social science (Allison, 1984; Blossfeld and Rohwer, 1995), and product or equipment reliability (Lawless and Kalbfleisch, 1992; Nelson, 1988). Much of this later development has taken place within the general framework of counting processes and event history analysis, which encompasses events of multiple types and transitions among defined states. Andersen et al. (1993) provide an authoritative account of methodology and the underlying mathematical theory. There is relatively little comprehensive material focused on recurrent events. Shorter treatments providing coverage of specific topics include Blossfeld and Rohwer (1995), Cai and Schaubel (2004a), Cook and Lawless (2002), Kalbfleisch and Prentice (2002, Ch.7), Karr (1991), Nelson (2003), Rigdon and Basu (2000), and Therneau and Grambsch (2000).

Issues concerning the selection and observation of individuals in a study of recurrent events are discussed further in Section 2.6 and the Bibliographic Notes for Chapter 2. Software for the analysis of recurrent event data is surveyed in Appendix B.

2

Models and Frameworks for Analysis of Recurrent Events

2.1 Mathematical Background

From both a theoretical and a practical perspective, the counting process notation introduced in Section 1.3 provides a convenient framework for the treatment of recurrent events. For now we dispense with subscripts that denote individuals or units and let $N(s,t)$ denote the number of occurrences of some type of event over the time interval $(s,t]$, for a specific individual. Unless stated otherwise, for convenience we assume that the process starts at $t=0$ with $N(0)=0$ and define $N(t)=N(0,t)$ for $t>0$. The process $\{N(t), 0 \le t\}$ is then the *counting process* for the event occurrences. For this and the following three chapters it is assumed that one type of event is of interest; multiple event types are considered in Chapter 6. In this section we derive probability distributions for observed event occurrence patterns and for gap times. The results, which are used in developing statistical methods throughout the book, are contained in Theorems 2.1 and 2.2. Readers who wish to focus on applications can safely skip over the derivations of these results.

For most of this chapter it is assumed that events occur in continuous time; special consideration of discrete time is provided in Section 2.5. In the continuous time setting, models for recurrent events can be specified very generally by considering the probability distributions for the number of events in short intervals $[t, t+\Delta t)$, given the history of events before time t. As described in Section 1.3, we define the history $H(t)$ of the event process at time t as

$$H(t) = \{N(s) : 0 \le s < t\} \qquad t > 0\,,$$

and let $\Delta N(t) = N(t+\Delta t^-) - N(t^-)$ denote the number of events in $[t, t+\Delta t)$. The value $N(0)$ is included in $H(t)$; this is typically equal to 0, but there are situations where it may take on positive values as well. In such cases $N(t)$ may be defined as either the number of events in $(0,t]$ or $[0,t]$, according to what is most useful.

It is assumed in the continuous time case that two events cannot occur simultaneously, and the *intensity function* (sometimes called the *complete intensity*) for the event process as defined in (1.2) is

$$\lambda(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N(t) = 1|H(t)\}}{\Delta t}\,. \tag{2.1}$$

Mathematically, it is assumed that an intensity is bounded and continuous except possibly at a finite number of points over any finite time interval. The intensity function defines an event process, and all process characteristics can be determined from it.

If observable fixed or time-varying covariates $x(t)$ are related to event occurrence they may be incorporated in the model by redefining the process history to include covariate information. The covariates are all assumed to be external (exogenous) in the development that follows. As discussed in Section 1.3.4, covariates that are internal (nonexternal) are more difficult to deal with in terms of both modeling and interpretation; they are discussed in Section 2.5. We let $x^{(t)} = \{x(s) : 0 \le s \le t\}$ denote the history of the external covariates over $[0, t]$, and $x^{(\infty)}$ denote the complete covariate path. Unless stated otherwise, we assume that probabilities are conditional on the covariate path, and include $x^{(\infty)}$ in the initial information $H(0)$ for convenience. It is assumed, though, that $\lambda(t|H(t))$ depends only on $x^{(t)}$.

To facilitate the development that follows, we introduce the *product integral* of a continuous integrable function $g(u)$ over the interval $[a, b]$. Let $a = u_0 < u_1 < \cdots < u_R = b$ partition $[a, b]$ and define $\Delta u_r = u_{r+1} - u_r$, $r = 0, 1, \ldots, R$, where $u_{R+1} = u_R^+$. The product integral of $g(u)$ over $[a, b]$ is defined as

$$\prod_{[a,b]} \{1 + g(u)du\} = \lim_{R\to\infty} \prod_{r=0}^{R} \{1 + g(u_r)\Delta u_r\}\,, \tag{2.2}$$

where as $R \to \infty, \max(\Delta u_r)$ approaches 0. By noting that $\log\{1+g(u)\Delta u\} = g(u)\Delta u + o(\Delta u)$ we see that the log of (2.2) approaches the Riemann integral $\int_a^b g(u)du$ in the limit, and so

$$\prod_{[a,b]} \{1 + g(u)du\} = \exp\left\{\int_a^b g(u)du\right\}\,. \tag{2.3}$$

It is also easily seen that

$$\prod_{[a,b]} \{1 + g(u)du + o(du)\} = \exp\left\{\int_a^b g(u)du\right\}\,, \tag{2.4}$$

a result which is useful below. This development also applies if g has a finite number of discontinuities over $[a, b]$.

Another useful concept is the *Riemann–Stieltjes integral*, which allows extensions to deal with discrete and continuous cumulative functions $G(t)$. Let $G(t)$ be a nondecreasing, right-continuous function with left-hand limits and a finite number of discontinuities (jumps) in any finite interval. Assume that $g(t) = G'(t)$ exists except at points of discontinuity of $G(\cdot)$ and that at points of discontinuity t_j we have $G(t_j) - G(t_j^-) = g_j$. The Riemann–Stieltjes integral of $dG(\cdot)$ over the interval $[a, b]$ is then defined as

$$\int_a^b dG(u) = \int_a^b g(u)du + \sum_{j:a\le t_j\le b} g_j \,,$$

where the first integral on the right-hand side is a Riemann integral. The Riemann–Stieltjes integral is therefore a Riemann integral when $G(t)$ is continuous, and reduces to a sum when $G(t)$ is a step function with jumps g_j at a countable set of points $\{t_j\}$. More generally, it handles functions with discrete and continuous components. The product integral can be similarly extended, replacing $g(u)du$ by $dG(u)$ in (2.2) and $g(u_r)\Delta u_r$ with $G(u_{r+1}) - G(u_r)$.

We now derive the probability density function for an event process that is observed over the fixed time interval $[\tau_0, \tau]$, conditional on $H(\tau_0)$. The probability density of the outcome "n events occur, at times $t_1 < t_2 < \cdots < t_n \le \tau$," where $n \ge 0$, may be obtained by considering partitions $\tau_0 = u_0 < u_1 < \cdots < u_R = \tau$ of $[\tau_0, \tau]$, and then taking a limit. The probability distribution of $N(u_1), \ldots, N(u_R)$, given $H(u_0)$, is

$$\prod_{r=0}^{R} \Pr\{N(u_r)|H(u_r)\} = \prod_{r=0}^{R} \Pr\{\Delta N(u_r)|H(u_r)\} \,, \tag{2.5}$$

where $\Delta N(u_r)$ is the number of events in $[u_r, u_{r+1})$. It follows from the definition (2.1) of the intensity function and the property that events cannot occur simultaneously that

$$\begin{aligned}
\Pr\{\Delta N(u_r) = 0|H(u_r)\} &= 1 - \lambda(u_r|H(u_r))\Delta u_r + o(\Delta u_r) \,, \\
\Pr\{\Delta N(u_r) = 1|H(u_r)\} &= \lambda(u_r|H(u_r))\Delta u_r + o(\Delta u_r) \,,
\end{aligned}$$

and

$$\Pr\{\Delta N(u_r) \ge 2|H(u_r)\} = o(\Delta u_r) \,.$$

Thus (2.5) equals

$$\begin{aligned}
\prod_{r=0}^{R} &\{\lambda(u_r|H(u_r))\Delta u_r + o(\Delta u_r)\}^{\Delta N(u_r)} \\
&\times \{1 - \lambda(u_r|H(u_r))\Delta u_r + o(\Delta u_r)\}^{1-\Delta N(u_r)} \,.
\end{aligned} \tag{2.6}$$

As R increases and the size of the Δu_r terms approach zero, the n intervals that contain the event times $t_1, \ldots, t_n$ have $\Delta N(u_r) = 1$; for all others

$\Delta N(u_r) = 0$. By dividing (2.6) by $\prod_{r=0}^{R} (\Delta u_r)^{\Delta N(u_r)}$ and letting $R \to \infty$, we obtain the following result.

Theorem 2.1. *Conditional on $H(\tau_0)$, the probability density of the outcome "n events, at times $t_1 < \cdots < t_n$," where $n \geq 0$, for a process with intensity (2.1), over the specified interval $[\tau_0, \tau]$, is*

$$\prod_{j=1}^{n} \lambda(t_j|H(t_j)) \cdot \exp\left\{ -\int_{\tau_0}^{\tau} \lambda(u|H(u))du \right\} . \tag{2.7}$$

The exponential term in (2.7) is obtained from the product integration result (2.4) with $g(u) = -\lambda(u|H(u))$, noting that the limit is unchanged by the deletion of the vanishingly small intervals around the n event times.

A second result that is often useful is the following; it provides conditional probabilities for interevent times and other waiting times.

Theorem 2.2. *For an event process with integrable intensity (2.1),*

$$\Pr\{N(s,t) = 0|H(s^+)\} = \exp\left\{ -\int_{s}^{t} \lambda(u|H(u))du \right\} . \tag{2.8}$$

Proof. Partition the interval $(s,t]$ as $s = u_0 < u_1 < \cdots < u_R = t$ and note as in the preceding development that

$$\begin{aligned} \Pr\{N(s,t) = 0|H(s^+)\} &= \lim_{R\to\infty} \prod_{r=1}^{R} \Pr\{\Delta N(u_r) = 0|H(u_r)\} \\ &= \lim_{R\to\infty} \prod_{r=1}^{R} \{1 - \lambda(u_r|H(u_r))\Delta u_r + o(\Delta u_r)\} , \end{aligned}$$

where all Δu_r approach zero as $R \to \infty$. The expression (2.8) follows immediately from the product integration result (2.4). Note that it is implicitly assumed under conditions stated following (2.1) that given $H(s^+)$, the intensity $\lambda(u|H(u))$ in (2.8) is deterministic and integrable over $(s,t]$, and that $\int_s^t \lambda(u|H(u))du$ is continuous.

Corollary. Let $W_j = T_j - T_{j-1}$ be the waiting time between the $(j-1)$st and jth events, where $T_0 = 0$ and $j = 1, 2, \ldots$. Then

$$\Pr\{W_j > w|T_{j-1} = t_{j-1}, H(t_{j-1})\} = \exp\left\{ -\int_{t_{j-1}}^{t_{j-1}+w} \lambda(u|H(u))du \right\} . \tag{2.9}$$

Proof. The left side of (2.9) equals $\Pr\{N(t_{j-1}, t_{j-1} + w) = 0|H(t_{j-1}^+)\}$ and so (2.9) follows directly from (2.8).

Knowledge of the intensity function allows us to write down the probability of a specified event history and conditional probabilities for interevent times,

as made explicit in Theorems 2.1 and 2.2. Other characteristics of an event process are less readily obtained from the intensity. Chief among these are the distribution of the event count $N(s,t)$ in time interval $(s,t]$, and the joint distribution of counts $N(s_j,t_j)$ in nonoverlapping intervals $(s_j,t_j]$, $j = 1,\ldots,m$. Even the mean function (1.4) and variance function for the counting process $\{N(t), 0 \le t\}$, denoted

$$\mu(t) = E\{N(t)\} \quad \text{and} \quad V(t) = \text{var}\{N(t)\}\,, \tag{2.10}$$

are difficult to determine from general intensity functions. In settings where there is a strong interest in distributions of event counts it may be preferable to specify the process in ways other than through its intensity; this is considered in Section 2.2.

The following sections describe some important families of recurrent event processes, which serve as a basis for modeling and data analysis in subsequent chapters. We start with processes for which properties of counts are easily obtained.

2.2 Poisson Processes and Models for Event Counts

Two types of processes for recurrent events might be considered canonical. One is the Poisson process, which describes situations where events occur randomly in such a way that the numbers of events in nonoverlapping time intervals are statistically independent. The other is the renewal process, in which the waiting (gap) times between successive events are statistically independent; that is, an individual is "renewed" after each event occurrence. Poisson processes tend to be appropriate in settings where events for an individual or system are triggered or influenced by random external factors, whereas renewal processes tend to describe settings in which events flow from physical cycles that are internal to an individual or system. Another aspect is whether events are incidental. For incidental events, Poisson processes or other models based on counts are often useful. In general the occurrence of events may of course be driven by a variety of internal and external factors, and the degree to which they are incidental may vary. As we show here and in Section 2.3, the applicability of either type of process can be extended greatly through the inclusion of covariates or random effects.

2.2.1 Poisson Processes

Poisson processes can be defined in various mathematically equivalent ways. One way mentioned earlier is through the independent counts property for nonoverlapping time intervals. Another way, which was described in Section 1.3.1, is via the intensity function: a Poisson process is one for which the intensity is of the form

$$\lambda(t|H(t)) = \rho(t) \qquad t > 0\,, \tag{2.11}$$

where $\rho(t)$ is a nonnegative integrable function. It is also assumed that the cumulative intensity

$$\mu(t) = \int_0^t \rho(u)du \qquad t > 0\,, \tag{2.12}$$

is continuous and finite for all $t > 0$. In the important special case where $\rho(t) = \rho$ is a constant, the process is called *homogeneous*; otherwise it is *nonhomogeneous*. We defer discussion of covariates to Section 2.2.2.

The Poisson process is seen from the definition (2.11) to be a Markov process; the probability of an event in $(t, t + \Delta t)$ may depend on t but is independent of $H(t)$. The following properties ensue from the definition.

(i) *$N(s,t)$ has a Poisson distribution with mean $\mu(s,t) = \mu(t) - \mu(s)$, for $0 \le s < t$.*
(ii) *If $(s_1, t_1]$ and $(s_2, t_2]$ are nonoverlapping intervals then $N(s_1, t_1)$ and $N(s_2, t_2)$ are independent random variables.*

To prove (i), note that by Theorem 2.1 we have from (2.7) that the probability density for the outcome "n events occur, at times $t_1 < \cdots < t_n$ in $(s,t]$" is

$$\left\{\prod_{j=1}^{n} \rho(t_j)\right\} \exp\{-\mu(s,t)\}\,, \tag{2.13}$$

where $n \ge 0$. The marginal probability of n events is then

$$\Pr(n \text{ events in } (s,t]) = \left\{\int \cdots \int \left[\prod_{j=1}^{n} \rho(t_j)\right] dt_1 \ldots dt_n\right\} \exp\{-\mu(s,t)\}\,, \tag{2.14}$$

where the multiple integral is over the region $s < t_1 < \cdots < t_n \le t$. Because the integrand $\prod \rho(t_j)$ is symmetric in $t_1, \ldots, t_n$ it follows that the integral in (2.14) is

$$\begin{aligned}(n!)^{-1} \int_s^t \cdots \int_s^t \prod_{j=1}^{n} \rho(t_j) dt_1 \cdots dt_n &= (n!)^{-1} \prod_{j=1}^{n} \left\{\int_s^t \rho(t_j) dt_j\right\} \\ &= (n!)^{-1} \mu(s,t)^n\,.\end{aligned}$$

Thus by (2.14),

$$\Pr(n \text{ events in } (s,t]) = \frac{\mu(s,t)^n}{n!} \exp\{-\mu(s,t)\} \qquad n = 0, 1, \ldots \tag{2.15}$$

which is the Poisson probability mass function as stated in (i).

Property (ii) above easily follows by noting that $t_1 < s_2$ and using the fact that the random variable $N(s_2, t_2)$ is independent of the history $H(s_2)$ of events prior to s_2. It is therefore independent of $N(s_1, t_1)$.

By (2.15) or (1.4) in Chapter 1, the counting process $\{N(t), 0 \leq t\}$ has *mean function* $\mu(t)$:

$$E\{N(t)\} = \mu(t)\,.$$

The *rate function* (also called the *rate of occurrence function*) for a process is defined as $\rho(t) = \mu'(t)$, where $\mu'(t) = d\mu(t)/dt$. It follows that

$$E\{\Delta N(t)\} = \rho(t)\Delta t + o(\Delta t)\,, \tag{2.16}$$

where $\Delta N(t)$ represents the number of events in the short interval $[t, t + \Delta t)$. Thus for a Poisson process, the rate function equals the intensity function. This property, which does not hold for other processes, reflects the fact that $\Delta N(t)$ is independent of $H(t)$.

The conditional distributions of gap times $W_j = T_j - T_{j-1}$ are, from (2.9), given as

$$\Pr(W_j > w | T_{j-1} = t_{j-1}) = \exp\{-\mu(t_{j-1}, t_{j-1} + w)\} \qquad j = 1, 2, \ldots \tag{2.17}$$

and so the gap times are not in general statistically independent. However, in the important special case of the *homogeneous Poisson process*, where $\rho(t) = \rho$, they are independent. In fact, because $\mu(t) = \rho t$ in this case, it follows from (2.17) that

(iii) *For the homogeneous Poisson process with intensity ρ, the gap times W_j $(j = 1, 2, \ldots)$ between events are independent and identically distributed (i.i.d.) exponential random variables with mean ρ^{-1}, and survivor function*

$$\Pr(W_j > w) = \exp(-\rho w) \qquad w > 0\,. \tag{2.18}$$

A final useful result is

(iv) *Let $\{N(t), 0 \leq t\}$ be a Poisson process with mean function $\mu(t)$. Define a new time scale (sometimes referred to as operational time) by $s = \mu(t)$ and define the process $\{N^*(s), 0 \leq s\}$ by*

$$N^*(s) = N(\mu^{-1}(s)) \qquad 0 < s\,.$$

Then $\{N^*(s), 0 \leq s\}$ is a homogeneous Poisson process with rate function $\rho^*(s) = 1$.

This result is discussed in Problem 2.2.

The concept of *trend* is important in many applications, as described in Chapter 1. For a Poisson process, $\rho(t)$ determines whether there is a trend in the rate of events; if $\rho(t)$ is monotone increasing or decreasing then a monotone trend is said to exist, but nonmonotone trends are also common.

Poisson process models may be parametric or nonparametric. For parametric models $\rho(t)$ is specified as a function of a finite-dimensional parameter. Common models include the exponential and power law models, in which $\rho(t;\alpha,\beta) = \exp(\alpha+\beta t)$ and $\rho(t;\alpha,\beta) = \alpha\beta t^{\beta-1}$, respectively.

2.2.2 Covariates in Poisson Processes

External covariates $x(t)$, which include fixed covariates, can be incorporated in a Poisson process by specifying the intensity as a function of t and the covariate history $x^{(t)} = \{x(u) : 0 \le u \le t\}$. This is usually done by defining covariate vectors $z(t)$ that are based on $x^{(t)}$ and then considering intensities of the form

$$\rho(t|x^{(\infty)}) = \rho(t|x^{(t)}) = \rho_0(t)\exp(z'(t)\beta)\,, \tag{2.19}$$

where β is a vector of regression parameters of the same length as $z(t)$. As is the case throughout this book, vectors are written in column form. The positive-valued function $\rho_0(t)$ is sometimes called the *baseline rate* or intensity, and corresponds to an individual for whom $z(t) = 0$ for all $t > 0$. For this function to have a practical interpretation it is common to center $z(t)$ in some way. Note that in (2.19) we make the reasonable assumption that given $x^{(t)}$, the intensity is independent of covariate values after t.

The *multiplicative model* (2.19) is sometimes referred to as a *log-linear* model and represents a flexible and convenient way to ensure positive-valued multiplicative effects of $z(t)$ for any β. The exponential term can be replaced by some other function $g(z(t);\beta)$ if desired. For example, $g(z(t);\beta) = 1 + z'(t)\beta$ is occasionally useful. In this case, the parameter space for β must be constrained to guarantee that $1 + z'(t)\beta > 0$.

If the baseline function $\rho_0(t)$ is specified parametrically, the model is fully parametric, otherwise it is *semiparametric*. The semiparametric model (2.19), with $\rho_0(t)$ an arbitrary positive-valued function, is sometimes called the Andersen–Gill (1982) model, and statistical methods for it are discussed in Chapter 3.

When all covariates are fixed, their effects have a simple interpretation, because conditional on the covariate vector z, the process $\{N(t), 0 \le t\}$ is Poisson with rate function $\rho_0(t)\exp(z'\beta)$ and mean function

$$E\{N(t)|z\} = \mu_0(t)\exp(z'\beta)\,, \tag{2.20}$$

where $\mu_0(t) = \int_0^t \rho_0(u)du$ is the baseline mean function. This is a log-linear model, in which both the mean and rate functions for any two individuals are proportional; the ratio of the functions for individuals with covariate vectors z_1 and z_2 is $\exp\{(z_1 - z_2)'\beta\}$.

When covariates are time-varying but external, the recurrent event process is still Poisson, conditional on the associated covariate history $x^{(\infty)}$. However, although the effect of covariates on the rate function (2.19) is easy to interpret, the effect on the mean function may be complex:

$$E\{N_i(t)|x_i^{(\infty)}\} = \int_0^t \rho_0(u)\exp\{z_i'(u)\beta\}du \tag{2.21}$$

does not in general have a simple form.

In models (2.20) with fixed covariates, the mean functions for individuals differ only in level (are proportional). Models for which both the shape and level of the rate function depend on x can be formulated through derived time-varying covariates $z(t)$ that are functions of x and t. For example, if $z(t) = x \cdot t$ then by (2.19) the rate functions $\rho(t|x) = \rho_0(t)\exp(z'(t)\beta)$ are not proportional for different x, nor are the mean functions. In many applications it is of interest to determine whether the shape or level of the rate or mean functions for $N(t)$ vary with covariate values.

The multiplicative model (2.19) can also be extended by allowing $z(t)$ to include components based on prior event history, such as the time since the most recent event or the number of previous events. In that case $z(t)$ has internal covariate components and the process is no longer Poisson, but it is often referred to as a *modulated Poisson process*. For such processes, the event intensity may depend on prior event history.

Nonmultiplicative regression models can also be formulated. Two prominent types are *additive models* for which

$$\rho(t|x^{(t)}) = \rho_0(t) + g(z(t);\beta) \tag{2.22}$$

and *time transform models* for which

$$\rho(t|x^{(t)}) = \rho_0\left(\int_0^t \exp(z'(u)\beta)du\right)\exp(z'(t)\beta)\,, \tag{2.23}$$

where, in both cases, $\rho_0(\cdot)$ is a baseline rate function. Time transform models are analogous to accelerated failure time models for lifetime data. In particular

$$s = g(t) = \int_0^t \exp(z'(u)\beta)du \tag{2.24}$$

can be considered as a transformed time scale defined by the covariate process, such that the process $\{N^*(s), 0 \le s\}$, where $N^*(s) = N(g^{-1}(s))$, is a Poisson process with intensity $\rho_0(s)$. Moreover, it follows from (2.23) that

$$E\{N^*(g(t))|x^{(t)}\} = E\{N(t)|x^{(t)}\} = \mu_0(g(t))\,, \tag{2.25}$$

where $\mu_0(t) = \int_0^t \rho_0(u)du$.

2.2.3 Random Effects in Poisson Processes

Sometimes, even after conditioning on covariates, there is more interindividual variation in event occurrence than is accounted for by a Poisson process. One sign of this is when $\mathrm{var}\{N_i(t)\}$ appears to be substantially larger than

$E\{N_i(t)\}$; the two are identical under a Poisson model. If counts are of interest and Poisson processes are still thought to be reasonable models for individuals, we can consider the incorporation of unobservable random effects u_i for individuals $i = 1, \ldots, m$, such that, given u_i and fixed covariates z_i, the process $\{N_i(t), 0 \le t\}$ is Poisson with rate function

$$\rho(t|z_i, u_i) = u_i \rho_0(t) \exp(z_i'\beta)\,. \tag{2.26}$$

The terms $u_1, \ldots, u_m$ are taken here to be i.i.d. with finite mean and distribution function $G(u)$. By including an intercept term in $z_i'\beta$, or absorbing it in $\rho_0(t)$, we may assume without loss of generality that $E(u_i) = 1$.

Random effects can also be incorporated in other ways. For example, if both the shape and level of intensity functions vary across individuals in a way that cannot be explained by observable covariates, we might consider bivariate random effects (u_i, v_i) and conditional rate functions of the form

$$\rho(t|z_i, u_i, v_i) = u_i \rho_0(t; v_i) \exp(z_i'\beta)\,. \tag{2.27}$$

Another option would be to consider fixed or time-varying covariates for which the regression coefficients are random.

Models (2.26) where u_i has a gamma distribution with mean 1, variance ϕ, and density function

$$g(u;\phi) = \frac{u^{\phi^{-1}-1} \exp(-u/\phi)}{\phi^{\phi^{-1}} \Gamma(\phi^{-1})} \qquad u > 0\,, \tag{2.28}$$

are especially convenient, because various process characteristics have closed-form expressions. In particular, if we write

$$\mu_i(s,t) = \int_s^t \rho_0(v) \exp(z_i'\beta) dv = \mu_0(s,t) \exp(z_i'\beta)$$

then given z_i and u_i, the distribution of $N_i(s,t)$ is Poisson with mean $u_i \mu_i(s,t)$. Given only z_i the probability function is then

$$\begin{aligned} \Pr(N_i(s,t) = n|z_i) &= \int_0^\infty \frac{[u\mu_i(s,t)]^n}{n!} \exp\{-u\mu_i(s,t)\} g(u;\phi) du \qquad (2.29) \\ &= \frac{\Gamma(n+\phi^{-1})}{\Gamma(\phi^{-1})} \quad \frac{[\phi\mu_i(s,t)]^n}{[1+\phi\mu_i(s,t)]^{n+\phi^{-1}}} \qquad n = 0, 1, 2, \ldots \end{aligned}$$

which is of negative binomial form. Note that the limit as $\phi \to 0$ gives the Poisson distribution (2.15).

The unconditional mean and variance of $N_i(s,t)$ can be obtained from (2.29) or by noting that

$$\begin{aligned} E\{N_i(s,t)\} &= E\{E[N_i(s,t)|u_i]\} \\ &= E\{u_i \mu_i(s,t)\} = \mu_i(s,t)\,, \end{aligned} \tag{2.30}$$

and

$$\begin{aligned}\text{var}\{N_i(s,t)\} &= E\{\text{var}[N_i(s,t)|u_i]\} + \text{var}\{E[N_i(s,t)|u_i]\} \\ &= E\{u_i\mu_i(s,t)\} + \text{var}\{u_i\mu_i(s,t)\} \\ &= \mu_i(s,t) + \phi\mu_i(s,t)^2\,. \end{aligned} \tag{2.31}$$

The covariance for the event count in nonoverlapping intervals may similarly be obtained as

$$\text{cov}\{N_i(s_1,t_1), N_i(s_2,t_2)\} = \phi\mu_i(s_1,t_1)\mu_i(s_2,t_2)\,. \tag{2.32}$$

It should be noted that the relationships (2.30)–(2.32) hold for any model (2.26) in which $E(u_i) = 1, \text{var}(u_i) = \phi$.

The mean function $\mu_i(t) = \mu_i(0,t)$ and rate function $\rho_i(t) = \mu_i'(t)$ for the "negative binomial" process $\{N_i(t), 0 \le t\}$ above are independent of ϕ and, in particular, are the same as when $\phi = 0$, that is, when the process is Poisson. However, when $\phi > 0$ the process is not Poisson, and the intensity function at time t depends both on ϕ and on the process history prior to t. It is given by (see Problem 2.6)

$$\lambda(t|H_i(t)) = \left\{\frac{1+\phi N_i(t^-)}{1+\phi\mu_i(t)}\right\}\rho_i(t) \qquad t > 0\,, \tag{2.33}$$

and because it depends on the process history only through $N_i(t^-)$ it is still Markov.

Under the model (2.26), the marginal rate and mean functions are

$$\rho_i(t) = \rho_0(t)\exp(z_i'\beta) \quad \text{and} \quad \mu_i(t) = \mu_0(t)\exp(z_i'\beta)\,, \tag{2.34}$$

so we see from (2.33) that although the covariate z_i has a simple multiplicative effect on the process mean and rate function, it has a more complicated effect on the process intensity. Note that the intensity at t increases with the value of $N_i(t^-)$. This makes sense because large values of $N_i(t^-)$ are associated with larger values of u_i, which in turn are associated with larger event counts beyond t.

2.2.4 Example: Mammary Tumors in Rats

Section 1.2.1 described a situation discussed by Gail et al. (1980), who presented data on the times to development of mammary tumors for 48 female rats in a carcinogenicity experiment. The animals were randomly assigned to two groups: treatment (23 animals) and control (25 animals). The data in Table 1.1 give the days on which new tumors were discovered for each animal; animals were inspected every few days for a period of 122 days. The event (appearance of a new tumor) times are not known exactly because of the intermittent inspections.

The main objective of analysis is a comparison of the treatment and control animals with respect to the frequency of tumor occurrence. Figure 1.2 indicates that the average rate of tumor occurrence across animals is roughly constant for both groups, so a very simple comparison can be based on the expected total tumor counts over the 122-day observation period. Let $\mu_T(t)$ and $\mu_C(t)$ represent the mean functions for tumors over $(0,t]$ for animals in the treatment and control groups, respectively, and $V_T(t)$ and $V_C(t)$ denote the respective variance functions. If we let $\mu_T = \mu_T(122)$, $\mu_C = \mu_C(122)$, $V_T = V_T(122)$, and $V_C = V_C(122)$, then a good measure of treatment effect is the ratio $\psi = \mu_T/\mu_C$, which can easily be estimated from the observed data.

The sample means $\widehat{\mu}_T = 2.65$ and $\widehat{\mu}_C = 6.04$ for the treatment and control groups give the estimate $\widehat{\psi} = 0.44$. These are the maximum likelihood estimates under the assumption that individual tumor counts $N_i(122)$ follow identical Poisson distributions with means $\mu_T(122)$ and $\mu_C(122)$ for the treatment and control animals, respectively.

Confidence intervals for ψ can be obtained under the Poisson model, but this is not necessarily appropriate. In fact, the sample variances for the total tumor counts in the two groups at 122 days are found to be $\widehat{V}_T = 3.7826$ and $\widehat{V}_C = 15.5400$. Under a Poisson model the mean and variance of $N_i(122)$ are the same, and the values of $\widehat{\mu}_C$ and $\widehat{V}_C$ indicate that this may not be true for the control group. More generally, Figure 1.2 suggests that $V_C(t)$ is substantially larger than $\mu_C(t)$ for this group. Formal methods of testing the Poisson assumption are given in Chapter 3 and we give a more detailed analysis there; here we just compare the results from the Poisson model with those of a simple robust analysis.

Exact small sample confidence interval procedures are available for Poisson models (e.g. Cox and Lewis, 1966) but given the large total tumor counts in the two groups, a simple large sample approach is satisfactory. Because $\widehat{\mu}_T$ is approximately normally distributed as $N(\mu_T, \mu_T/23)$, it follows that $\log \widehat{\mu}_T$ is approximately $N(\log \mu_T, (23\mu_T)^{-1})$; the latter approximation is typically better when μ_T is small. Similarly $\log \widehat{\mu}_C$ is approximately $N(\log \mu_C, (25\mu_C)^{-1})$. Thus the random variable $W = \log(\widehat{\mu}_T/\widehat{\mu}_C) = \log \widehat{\psi}$ is approximately normal:

$$W \simeq N\left(\log \psi,\ \frac{1}{23\mu_T} + \frac{1}{25\mu_C}\right). \tag{2.35}$$

By treating W as exactly normal we can obtain approximate confidence intervals for $\log \psi$ and thus for ψ. For example, an approximate 95% confidence interval for $\log \psi$ is $\log \widehat{\psi} \pm 1.96\{(23\widehat{\mu}_T)^{-1} + (25\widehat{\mu}_C)^{-1}\}^{1/2}$.

The above procedure is not robust to departures from the Poisson model. A robust approach is to note that $\widehat{\mu}_T$ and $\widehat{\mu}_C$, or $\log \widehat{\mu}_T$ and $\log \widehat{\mu}_C$, are approximately normally distributed even when the total counts $(N_i(122))$ are not Poisson random variables. In particular, $\widehat{\mu}_T \simeq N(\mu_T, V_T/23)$ and $\widehat{\mu}_C \simeq N(\mu_C, V_C/25)$, where V_T and V_C represent $\mathrm{var}\{N_i(122)\}$ for treatment and control animals. This leads to

$$W \simeq N\left(\log\psi, \frac{V_T}{23\mu_T^2} + \frac{V_C}{25\mu_C^2}\right) \tag{2.36}$$

and to the approximate 95% confidence interval

$$\log\widehat{\psi} \pm 1.96\left(\frac{\widehat{V}_T}{23\widehat{\mu}_T^2} + \frac{\widehat{V}_C}{25\widehat{\mu}_C^2}\right)^{1/2}$$

for $\log\psi$, where $\widehat{V}_T$ and $\widehat{V}_C$ are sample variance estimates given above.

The 95% confidence intervals obtained for $\log\psi$ and for ψ from the two approaches are as follows.

Poisson method: $-1.12 \le \log\psi \le -0.526$ and $0.326 \le \psi \le 0.591$.

Robust method: $-1.22 \le \log\psi \le -0.429$ and $0.296 \le \psi \le 0.651$.

The robust confidence intervals are wider, reflecting the fact that the variance of W in (2.36) is larger than that in (2.35) when V_T/μ_T and V_C/μ_C are greater than one. This phenomenon is often referred to as *overdispersion* relative to a Poisson model, or it is said that there is *extra-Poisson variation*. One way to model this formally is to introduce random effects, as described in Section 2.2.3, and an alternative analysis would be to adopt a negative binomial model (2.29) for each of the treatment and control groups. Large sample maximum likelihood methods (Appendix A) could then be used to obtain confidence intervals for ψ. This approach is illustrated in Section 3.4, where inference for negative binomial and other mixed Poisson processes is considered. This gives confidence intervals for ψ that are close to those for the robust method above.

The simple analysis here indicates strongly that tumor frequency over the 122 day study is substantially lower under the treatment than under the control. These results are most meaningful and easily interpreted when the rate of occurrence functions $\rho(t)$ for the two groups are constant; in that case $\rho_T(t) = \alpha_T$, $\rho_C(t) = \alpha_C$, and $\psi = \alpha_T/\alpha_C$ is the ratio of the rates as well as the ratio of $\mu_T(t) = \alpha_T \times t$ and $\mu_C(t) = \alpha_C \times t$. Conversely, if $\rho_T(t)$ and $\rho_C(t)$ were markedly nonlinear and nonproportional, the ratio $\mu_T(t)/\mu_C(t)$ would vary over time and so the treatment effect $\psi = \mu_T(122)/\mu_C(122)$ would be dependent on the duration of the study and would not represent the effect over a shorter or longer study. As noted above, Figure 1.2 in Section 1.2.1 shows plots of $\widehat{\mu}_T(t)$ and $\widehat{\mu}_C(t)$ and they are both close to linear, indicating that the two rate functions are roughly constant.

2.3 Renewal Processes and Models for Gap Times

2.3.1 Models for Gap Times Between Events

Renewal processes are ones in which the gaps $W_j = T_j - T_{j-1}$ $(j = 1, 2, \ldots)$ between successive events are independent and identically distributed. This is

equivalent to the condition that the process intensity is of the form (1.5),

$$\lambda(t|H(t)) = h(B(t)) \qquad t > 0\,, \tag{2.37}$$

where $B(t) = t - T_{N(t^-)}$ is the time since the most recent event before t, or backwards recurrence time, and $h(w)$ is the *hazard function* for the variables W_j. That is, if the W_j have common density function $f(w)$ and survivor function $S(w) = P(W \geq w)$, then

$$h(w) = \frac{f(w)}{S(w)} = \lim_{\Delta w \downarrow 0} \frac{\Pr(W < w + \Delta w | W \geq w)}{\Delta w}\,.$$

It is assumed here that the time origin $t = 0$ corresponds to an event time. Sometimes this is relaxed and W_1 is allowed to have a different distribution from $W_2, W_3, \ldots$, with the gap times still being mutually independent.

The concept of no trend in a process of recurrent events can, as suggested in Chapter 1, be interpreted in various ways. A pure renewal process may be said to exhibit no trend inasmuch as the gap times are i.i.d. The rate function $\rho(t)$ is not in general linear but under quite general conditions $\rho(t)$ approaches the constant value $E\{W\}^{-1}$ as t becomes large.

The distribution for counts $N(s,t)$ in renewal processes is in general mathematically intractable. An exception is for the renewal process in which the W_j are exponential random variables; the process is then a homogeneous Poisson process. Another exception is for the distribution of $N(t)$, which can be obtained from the relationship

$$\Pr(N(t) \geq n) = \Pr(T_n \leq t) \tag{2.38}$$

and the fact that $T_n = W_1 + \cdots + W_n$ is a sum of i.i.d. random variables. It also follows from (2.38) that $\Pr(N(t) = n) = \Pr(T_n \leq t) - \Pr(T_{n+1} \leq t)$, and

$$\mu(t) = E\{N(t)\} = \sum_{n=1}^{\infty} F_n(t)\,, \tag{2.39}$$

where $F_n(t)$ is the distribution function for T_n. Calculation of (2.38) or (2.39) can be approached in various ways; Problem 2.7 provides some direction. The marginal distribution of $N(s,t)$ for $s > 0$ is much less tractable because the time of the last event prior to s is unspecified. When such count distributions are wanted, it is simplest to determine them by simulation; see Problem 2.8.

Covariates may be incorporated into renewal processes in straightforward ways. If fixed covariates z are associated with independent renewal processes, we can allow the common distribution of the gap times W_j for a given process to depend on z. Because the W_j are positive-valued, regression models used in connection with lifetime data (e.g. Lawless, 2003a) may be used. The two most important families of such models are the *proportional hazards model* where the hazard function of W_j given z is of the form

$$h(w|z) = h_0(w)\exp(z'\beta)\,, \tag{2.40}$$

and the *accelerated failure time model*, where the hazard function is of the form

$$h(w|z) = h_0(we^{z'\beta})\exp(z'\beta)\,. \tag{2.41}$$

In each of (2.40) and (2.41), $h_0(w)$ is a positive-valued function referred to as the "baseline" hazard function.

If there are external time-varying covariates $z(t)$, then renewal models in which the process intensity is of the form

$$\lambda(t|H(t)) = h(B(t)|z(t)) \tag{2.42}$$

can be considered. This is equivalent to incorporating the time-varying covariate $z(t)$ into the hazard function for the W_j. The multiplicative model with

$$h(w|z(t)) = h_0(w)\exp(z'(t)\beta)\,, \tag{2.43}$$

where $t = w + t_{N(t^-)}$, is very useful. In a model like (2.43) the W_j for a given process are independent (given the full covariate history) but are not identically distributed.

In many applications the assumption of independent gap times is not tenable, even after conditioning on covariates. Models based on gap times can then be approached for the case of fixed covariates through specification of the distribution of W_j given $W_1, \ldots, W_{j-1}$ and z ($j = 1, 2, \ldots$). Gaussian (normal) models for $Y_j = \log W_j$ are often convenient; in this case the W_j are said to have *log-normal distributions*, conditional on prior gap times and covariates. Very flexible modeling is also possible through the multiplicative model (2.43), by allowing $z(t)$ to include components of prior event history such as gap times or number of events. Such models are often called *modulated renewal processes*: the elapsed time $w = B(t)$ since the most recent event is taken as the baseline time variable for each new event, but the gap time may depend on previous event or covariate history.

Random effects can be introduced into renewal models in various ways. The simplest and most easily interpreted is where there are independent random effects u_i associated with individual processes, so that given u_i and any relevant covariate values, the gap times W_{ij} ($j = 1, 2, \ldots$) for process i are independent. Two useful such models when all covariates z_i are fixed, are

(i) The conditional multiplicative model in which W_{ij} has a hazard function, given u_i and z_i, of the form

$$h(w|u_i, z_i) = u_i h_0(w)\exp(z_i'\beta)\,, \tag{2.44}$$

where the u_i are i.i.d. random variables, and

(ii) The conditional Gaussian model in which $Y_{ij} = \log W_{ij}$ has the conditional distribution

$$Y_{ij}|u_i, z_i \sim N(u_i + z_i'\beta, \sigma^2)\,, \tag{2.45}$$

where $u_i \sim N(0, \sigma_u^2)$.

In such models, the gap times for an individual process are not independent once the conditioning on the unobservable u_i is removed. The joint distribution of gap times is, however, exchangeable.

2.3.2 Example: Bowel Motility Cycles

Section 1.3.3 described a study of motility (muscular activity) patterns in the small bowels of 19 human subjects. Table 1.4 shows the lengths of successive motility (or "MMC") cycles, which followed a "fed state" period that each subject experienced after consumption of a standard meal. The motility cycle patterns are believed to be independent of the duration of the fed state. A main objective is to characterize their variability both within and between subjects.

In this context we can associate events with the start of a motility cycle. Let w_{ij} denote the length of the jth cycle for subject i, where $i = 1, \ldots, 19$ and $j = 1, \ldots, n_i$; the length of the last cycle is right-censored for each subject because the total followup time for the study was fixed. A basic question is whether the w_{ij} for a given subject indicate either a trend or autocorrelation. The amount of variability in the cycles for an individual subject is also of interest. Figure 2.1 shows plots of points $(w_{i,j-1}, w_{i,j})$ for $j = 2, 3$. One point in the left panel and two in the right panel have censored gap times (for W_2 and W_3, respectively) and are denoted by the + symbols, but this has a minor effect and there is no indication of trend or autocorrelation. Figure 2.2 shows Kaplan–Meier estimates of the survivor functions $S_1(t) = \Pr(W_{i1} \geq t), S_2(t) = \Pr(W_{i2} \geq t)$, and $S_3(t) = \Pr(W_{ij} \geq t)$ where in the last case $j \geq 3$ and it is assumed that gap times W_{ij} ($i = 1, \ldots, 19$; $j = 3, 4, \ldots$) are independent and identically distributed. There is an indication from Figure 2.2 that first cycles tend to be slightly longer than second or subsequent cycles, which appear to have similar length distributions. A tentative conclusion is therefore that cycle lengths are approximately independent within subjects, that the lengths of second and subsequent motility cycles are close to identically distributed, and that first cycles tend to be slightly longer.

It is important to note a potential problem with this informal analysis, however. In the present study, and many others, the duration of the study is fixed for a given subject. If the gap times between events (which here are the starts of cycles, with the cycle length playing the role of gap time) are not independent then censoring times for second and subsequent gap times are not independent of the gap times. For example, if τ_i is the duration of followup for subject i, then the censoring time for W_{i2} is $C_{i2} = \tau_i - w_{i1}$. If W_{i1} and W_{i2} are dependent, then W_{i2} and C_{i2} are dependent and so one of the key conditions

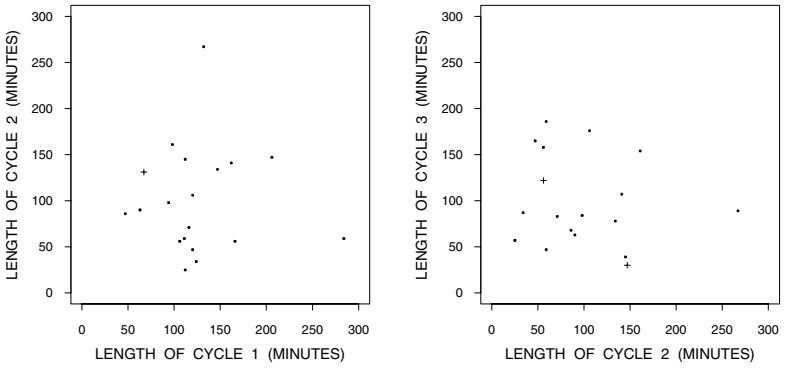


Fig. 2.1. Scatterplots of points (w_{i1}, w_{i2}) and (w_{i2}, w_{i3}) in the left and right panels, respectively, for bowel motility cycles.

for standard methods of survival analysis is violated. If W_{i1} and W_{i2} are positively correlated, it can be shown, for example, that the Kaplan–Meier estimate for $S_2(t) = \Pr(W_{i2} \geq t)$ computed from the available observations w_{i2} is biased downward; this is discussed in Section 4.4. If $\Pr(W_{i1}+W_{i2} > \tau_i)$ is not small, and (W_{i1}, W_{i2}) are strongly correlated, the bias will be substantial. The power of plots like Figure 2.1 to detect dependence or trend in gap times can also be compromised in studies with fixed followup times. For example, if W_1 and W_2 have positive association the fact that larger W_1 are more likely to have W_2 censored can mask the association in a plot of W_1 versus W_2.

In the case of gap time analysis, model-based procedures are essential. These are developed in Chapter 4, where we revisit these bowel motility data and demonstrate that there is indeed no significant evidence that the cycle lengths for an individual subject are dependent. It is also shown that individual Kaplan–Meier estimates for first, second, or subsequent gaps are in this case consistent, thus confirming the results of the informal analysis given here.

2.4 General Intensity-Based Models

Poisson and renewal processes have simple, easily interpreted properties in terms of counts and gap times, respectively. Their range of application is

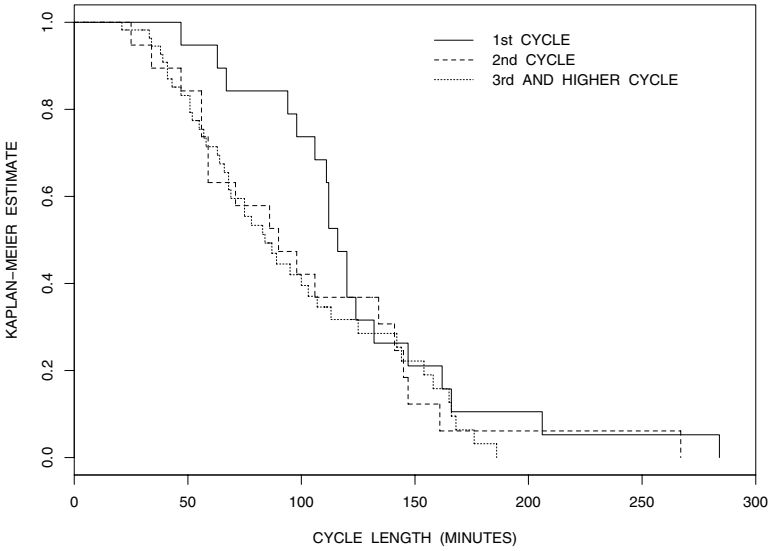


Fig. 2.2. Kaplan–Meier estimates of the survivor functions for first, second, and later bowel motility cycles.

limited, however, and in most settings we have to consider either extensions of these processes, or alternative models formulated via intensity functions. The multistate framework in Section 1.3.3 is often a convenient way to think about general models. Another option in some cases is to forgo modeling of the complete event process, and to consider only certain features of interest. For example, useful analysis can often be based on rate and mean functions for counts, perhaps supplemented with variance functions. This approach is developed in Section 3.6.

Intensity-based models, including modulated Poisson and renewal processes represented by (2.19) and (2.43), are relatively easy to fit and check when exact event times can be observed. This is discussed in Section 2.6 and demonstrated over the next several chapters, where methods of estimation, testing, and model checking are developed for the processes described in this chapter.

The result (2.9) gives the distribution of a gap time conditional on prior history. We use this result to describe how a process can be simulated, given the intensity function. Specifically, (2.9) indicates that if all covariates are external, then given $H(t_{j-1})$ and the $(j-1)$st event time t_{j-1}, the random variable

$$E_j = \int_{t_{j-1}}^{t_{j-1}+W_j} \lambda(t|H(t))dt \tag{2.46}$$

has a standard exponential distribution with survivor function $\exp(-u), u > 0$; see also Problem 2.2. Thus, given t_{j-1}, $H(t_{j-1})$, and the values of any external covariates beyond t_{j-1}, we can generate W_j by generating E_j and then solving (2.46) for W_j. The latter step has to be carried out numerically in most cases, but this is often easy to do. By repeating this for $j = 1, 2, \ldots$, we can generate successive event times $t_j = t_{j-1} + w_j$. Problem 2.2 also gives another simulation procedure which is useful when (2.46) is difficult to solve.

Example 2.4.1

Suppose that the process intensity is

$$\lambda(t|H(t)) = \exp\{\alpha_0 + g_1(t) + g_2(N(t^-)) + I(N(t^-) > 0)g_3(B(t))\}\,,$$

where $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ are specified functions. Then (2.46) becomes, for $j \geq 1$ and with a change of the variable of integration to $w = t - t_{j-1}$,

$$E_j = e^{\alpha_0 + g_2(j-1)} \int_0^{W_j} \exp\{g_1(t_{j-1} + w) + I(N(t^-) > 0)g_3(w)\}dw\,.$$

2.5 Discrete-Time Models and Time-Varying Covariates

Models for recurrent events may also be formulated in discrete time. For general discussion, assume that allowable t-values are $0, 1, 2, \ldots$, and let $n(t) = N(t) - N(t-1)$ denote the number of events at time t. In some applications $n(t)$ may equal only 0 or 1, but $n(t) \geq 2$ is allowed in general. Conditional on $n(0)$ and $x(0)$, the process is completely specified by the distributions of $n(t)$ given the event and covariate history $H(t) = \{n(0), \ldots, n(t-1), x(0), \ldots, x(t)\}$ for each $t \geq 1$. If $n(t)$ is binary, these distributions are defined by the discrete intensities (conditional event probabilities)

$$\Pr\{n(t) = 1|H(t)\} = \lambda(t|H(t)) \qquad t = 1, 2, \ldots \tag{2.47}$$

which are models for longitudinal binary responses (e.g. Diggle et al. 2002, Ch. 10). When $n(t) \geq 2$ is possible, models based on modulated Poisson processes are important; here $n(t)$ given $H(t)$ is taken to have a Poisson distribution with some specified mean $\rho(t|H(t))$. Multiplicative models where

$$E\{n(t)|H(t)\} = \rho(t|H(t)) = \rho_0(t)\exp\{x'(t)\beta\} \qquad t = 1, 2, \ldots \tag{2.48}$$

are very useful.

If the covariates $x(t)$ in a Poisson model (2.48) are random but external, then conditional on the full covariate history $x^{(\infty)}$, the $n(t)$ terms are independent Poisson random variables. Let us verify this statement, and at the

same time formalize the concept of an external covariate. Formally, we say a covariate process $\{x(t), 0 \leq t\}$ is external if it is not influenced by the event process. This means for $t = 1, 2, \ldots,$ that

$$\Pr\{x(t)|H_1(t)\} = \Pr\{x(t)|x^{(t-1)}\}\,, \tag{2.49}$$

where $H_1(t) = \{n(s), x(s); s = 0, 1, \ldots, t-1\}$. In that case, if we consider the joint distribution of the $n(t)$ and $x(t)$ up to time τ, we have

$$\prod_{t=1}^{\tau} \Pr\{n(t), x(t)|H_1(t)\} = \prod_{t=1}^{\tau} \Pr\{n(t)|H(t)\} \Pr\{x(t)|H_1(t)\} \tag{2.50}$$

$$= \prod_{t=1}^{\tau} \Pr\{n(t)|H(t)\} \Pr\{x(t)|x^{(t-1)}\}\,, \tag{2.51}$$

where $H(t) = (H_1(t), x(t))$. It then follows from (2.49) and (2.51) that

$$\Pr\{n(1), \ldots, n(\tau)|n(0), x^{(\tau)}\} = \prod_{t=1}^{\tau} \Pr\{n(t)|H(t)\} \tag{2.52}$$

and so under (2.48) and conditional on $x^{(\tau)}$, the $n(t)$ are indeed independent Poisson random variables. This argument also shows that the $n(t)$ are conditionally independent when the distribution of $n(t)$ given $H(t)$ depends only on $x(t)$, of which (2.48) with $n(t)$ Poisson is a special case.

From (2.50) we can also see where problems arise if covariates are internal (not external). In that case $\Pr\{x(t)|H_1(t)\}$ depends on $(n(0), \ldots, n(t-1))$ and (2.51) and (2.52) are not obtained. In fact, the right side of (2.52) is not the probability of any outcome and so although (2.48) has a probabilistic interpretation, the right side of (2.52) does not. In order to obtain the marginal probability of an outcome such as the sequence of counts $n(1), \ldots, n(\tau)$ it is necessary to average over the time-varying covariates $x(t)$ in (2.50). This makes it necessary to model the covariate and event occurrence processes simultaneously, whereas with external covariates we need only condition on the observed covariate values because of (2.52). Essentially the same conditions apply in the case of continuous-time processes.

Discrete time renewal or modulated renewal processes can be formulated when $n(t)$ is binary, by expressing (2.47) in terms of the time since the last event prior to t. Random effects can also be introduced into models in either the binary or general case; Problem 2.11 provides an example.

Strictly speaking, event times are always recorded on a discrete time scale. Nevertheless, it is customary to use continuous-time models when events occur in continuous time. In some applications events occur in continuous time but the process is observed intermittently. For example, one might record only the numbers of events $n(t)$ on successive days $t = 1, 2, \ldots$. In this case we can either use a discrete time model or deal with the interval counts $n(t) = N(t) - N(t-1)$ in a continuous time model. As discussed in previous sections,

the distributions of counts are intractable for many continuous time processes so we would try to avoid such processes in our models. In Chapter 7 we consider these issues in some detail.

2.6 Likelihood for Selection and Observation Schemes

Event history processes are studied through the collection and analysis of data. As discussed in Section 1.4, two questions about study design and the collection of data have important ramifications for analysis and interpretation of results: How are individuals selected for inclusion in the study? What information is collected about the event histories of individuals in the study, and how is this done?

Prospective studies, introduced in Section 1.4, typically take a group or "cohort" of individuals and then follow it longitudinally over time, recording events and covariates of interest. The group can often, but not always, be viewed as a random sample selected from a population of individuals or processes. The conditions for including an individual in the study can depend on their covariates or event history prior to the study, but these factors must then be considered in the analysis. Analyses based on a full specification of the event process can normally accommodate this easily, but difficulties may arise for methods based on partially specified processes such as mean function models.

Assume that event occurrences and covariate values $x(t)$ are recorded over the time interval $[\tau_0, \tau]$ for a specific individual. The time τ_0 may or may not correspond to the natural or "physical" start of the event process for the individual; the process time origin is dealt with in the specification of the intensity function and other process characteristics. In clinical trials t usually corresponds to time on study for the individual and $\tau_0 = 0$, although in a process involving, say, repeated infections, a person's prior history of infections before τ_0 may determine whether he is selected for the study. In an observational study on the occurrence of hospitalization episodes for psychiatric disorders, we may prefer to treat age as the time variable t for modeling and analysis. In that case τ_0 would be the individual's age at the start of followup.

The times τ_0 or τ can be random, as opposed to prespecified by the study. The latter is referred to as the *termination time* or *end-of-followup time* for the observed process. Sometimes it is also called a *censoring time*, by analogy with survival analysis. Two situations can be distinguished.

(i) Termination of observation is due to a study ending or a person becoming lost to followup; the event process may continue beyond τ but we are unable to observe it.

(ii) Termination is due to another type of event that ends the main event process. For example, if an individual dies at time τ then his process of recurrent bronchial infections is terminated at that time.

Dealing with case (ii) is a little more difficult. The terminating event may be related to the recurrent events, and process variables such as interval counts or event times are subject to constraints (e.g. $T_j \leq \tau$), so joint modeling of the recurrent and terminating events may be necessary.

Wherever possible we consider likelihood-based methods of statistical inference. It was shown in Section 2.1 that for a process observed over the fixed time interval $[\tau_0, \tau]$, the expression (2.7) gives the conditional probability density for the outcome "n events occur at times $t_1 < \cdots < t_n$." Then, the contribution to the likelihood function for an individual can be taken as

$$L = \prod_{j=1}^{n} \lambda(t_j|H(t_j)) \cdot \exp\left\{-\int_{\tau_0}^{\tau} \lambda(u|H(u))du\right\} \tag{2.53}$$

and the likelihood from a group of m independent individual processes is a product of such terms.

If τ_0 or τ_1 is random then under certain conditions (2.53) is still valid for inference purposes. This is obviously the case if τ_0 and τ_1 are determined independently of the event process, in which case (2.53) is the event history density, conditional on τ_0, τ, $H(\tau_0)$, and the covariate history. However, (2.53) is valid even more generally provided that τ_0 and τ are what are referred to as *stopping times* with respect to the process, and provided $\lambda(u|H(u))$ is the appropriate event intensity when the condition that the individual is "under observation" at time u (i.e. that $\tau_0 \leq u \leq \tau$) is added. Formally we require

$$\lambda(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N(t) = 1|H(t), \tau_0 \leq t \leq \tau\}}{\Delta t}\,. \tag{2.54}$$

The stopping time condition means that τ_0 and τ may be determined randomly in a way that depends on event history prior to τ_0 and τ, respectively, but not on the process after those times. In this general case (2.53) may not be the conditional probability density of the outcome "n events, at times $t_1 < \cdots < t_n$," given τ_0, τ, and $H(\tau_0)$. However, it is a "partial" likelihood, and it has been shown that it can be treated as an ordinary likelihood function for estimation or testing of model parameters (Andersen et al. 1993, Ch. 2).

A little different way of writing expressions such as (2.53) and (2.54) is based on the definition of the "at risk" process $\{Y(t), 0 \leq t\}$ introduced in Section 1.4.1, where

$$Y(t) = I(\text{process is observed at time } t)\,.$$

That is, if an event occurs at time t it is observed, and so we say the individual is "at risk" of having an observed event at time t. In the preceding discussion $Y(t) = I(\tau_0 \leq t \leq \tau)$ but the notation can also accommodate settings where individuals are observed over disjoint time intervals, as described in Section 1.4.1.

We often focus on the most common setting in which the process is observed over a single complete interval from $\tau_0 = 0$. By convention, $Y(t)$

is assumed to be left-continuous, so $Y(t) = Y(t^-)$. We can then define $\bar{N}(t) = \int_0^t Y(u)dN(u)$, which represents the observed part of the counting process. We may then consider the history of the observable process $\bar{H}(t) = \{\bar{N}(s), Y(s), 0 \leq s < t\}$, and let

$$\bar{\lambda}(t|\bar{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta\bar{N}(t) = 1|\bar{H}(t)\}}{\Delta t}$$

denote the intensity of the observable process. If $\Delta N(t)$ and $Y(t)$ are conditionally independent given the history, then $\bar{\lambda}(t|\bar{H}(t)) = Y(t)\lambda(t|H(t))$ and the censoring mechanism is said to be *conditionally independent*. Note that it is important to ensure a sufficient amount of information is incorporated from $H(t)$ into the model to justify the conditional independence assumption; typically this means including terms that affect both $\{N(t), 0 \leq t\}$ and $\{Y(t), 0 \leq t\}$. If an insufficient amount of information is included to warrant the conditional independence assumption, then a dependent censoring or dependent observation scheme is present and it is more difficult to relate the intensity of the observable process to that of the underlying process. Methods for estimating the features of the underlying process in this setting are discussed in Chapter 7, but these typically require effort to model the observation process. These approaches are commonly adopted when models for marginal features of the underlying event process are of interest.

The key feature under conditionally independent censoring schemes is that the intensity of the observable counting process is $\bar{\lambda}(t|\bar{H}(t)) = Y(t)\lambda(t|H(t))$. When $Y(t) = 0$ the process is not under observation at t and hence it is not possible to observe an event at t. Unless we state otherwise, we assume that the *observation* or *at-risk* process is conditionally independent of the event process, so that the likelihood for the observable data can be written down directly in terms of the model for the underlying process. In this case (2.53) can be rewritten as

$$L = \prod_{j=1}^{n} \lambda(t_j|H(t_j)) \cdot \exp\left\{-\int_0^\infty Y(u)\lambda(u|H(u))du\right\} \tag{2.55}$$

and it is valid for inference about $\lambda(t|H(t))$ even when τ_0 and τ in (2.53) are random. This version is also valid when an individual process is observed intermittently, as in Section 1.4. A practical constraint on the use of (2.55), however, is that the necessary information in $H(u)$ must be available for $\lambda(u|H(u))$ to be known. Problem 2.15 considers estimation based on (2.55).

It was observed in Chapter 1 that studies may use some form of retrospective observation, meaning that part of the period $[\tau_0, \tau]$ for which an individual's process history is recorded is prior to the time of selection of the individual for the study. In such cases the selection of an individual might not be independent of the process history over $[\tau_0, \tau]$; the selection plan must then be taken into account or inferences may be substantially biased. In Chapters 3 to 6 we consider only prospective studies or equivalent retrospective studies in

which the selection of an individual is independent of their event history over $[\tau_0, \tau]$. We deal with retrospective studies in which selection is not independent of event history in Chapter 7.

It has been assumed in the presentation of (2.53) that the exact times of events, as well as covariate values and the relevant history $H(\tau_0)$ needed to specify the process intensity $\lambda(t|H(t))$ for $t \geq \tau_0$, are available. Sometimes information is missing; for example, a process may be observed intermittently so that only the numbers of events between successive followup points, and not exact event times, are available. Information on events or covariates after an individual is lost to followup are of course missing. Event history prior to τ_0 may also be missing, for example, the time of the most recent event in a process that started before τ_0. The history $H(\tau_0)$ is part of the initial conditions for an individual; see Section 1.4.3. Problems can arise when essential parts of $H(\tau_0)$ are missing; Sections 4.5 and 7.3 consider this issue.

A crucial issue is whether data are missing "at random" in some sense. In some circumstances the mechanism that leads to data being missing is completely independent of the event process; in the terminology of Little and Rubin (2002) the data are then said to be MCAR, for "missing completely at random." If the missing data mechanism depends on observed external covariates that are included in the process model, then the missing data are still MCAR.

The condition that data be MCAR is too stringent to cover many relatively simple situations involving longitudinal data. For example, it does not hold for a study in which an individual is observed until some specified number of events occurs. Fortunately we need only weaker requirements which allow the probability that data are missing at t to depend upon variables in $H(t)$, although not on responses at or after t. The conditions that, as discussed above, event and observation processes are conditionally independent, and the times τ_0 and τ above are stopping times, fulfill this requirement and, when processes are observed continuously, make the likelihood (2.53) valid. The conditions are more stringent when a process is observed only intermittently, however. Suppose, for example, that an individual is scheduled to be seen at specified times $a_1 < a_2 < \cdots < a_k$. In that case $\tau = a_r^+$ for some $r \geq 1$ and if $r < k$ it means that the individual was observed at a_r but not at a_{r+1}. The stopping rule condition then allows the probability the individual is not observed at a_{r+1} to depend on $H(a_r^+)$, but not on events or new covariate values over $(a_r, a_{r+1}]$. In some settings this condition may be violated, for example, when the probability of loss to followup is associated with the number of recent event occurrences. Intermittent observation is considered in some detail in Chapter 7; until then processes are assumed to be continuously observed.

The stopping time requirements are an example of the general concept of data being "missing at random" or MAR, in the terminology of Little and Rubin (2002). This allows the probability that data are missing to depend on

variables (data) that are observed, but not on missing variables; this is weaker than the MCAR condition.

2.7 Bibliographic Notes

Models for recurrent events are discussed in many texts on point processes or stochastic processes (e.g. Cox and Isham, 1980; Daley and Vere-Jones, 1988, 2003; Parzen, 1999; Ross, 1983; Snyder and Miller, 1991), although covariates are rarely mentioned. Poisson and related processes are especially widely studied; Grandell (1997) provides many results. The symposium volume Lewis (1972) gives an excellent early overview of models, applications, and statistical methods for point processes. Theory and statistical methods for counting processes are given an authoritative treatment by Andersen et al. (1993), who emphasize modulated Markov models and multiplicative regression models. They also provide extensive references to the work of Brémaud, Jacod and others, on rigorous mathematical foundations for counting processes. Many results are also scattered across the literatures of engineering and mathematics; for examples see Segall and Kailath (1975) and, for results related to (2.9) and (2.46), Papangelou (1972), or Brown and Nair (1988). Cox and Lewis (1966) is an early but still very useful discussion of methods for analyzing recurrent event data. Nelson (2003) considers simple but useful methods based on mean functions. Karr (1991) gives a mathematical treatment of inference for point processes.

Statistical analysis of recurrent events for multiple individuals, and the inclusion of covariates and random effects, has received considerable attention since about 1980. Many references on analysis are given in later chapters, but we mention a few key ones here, emphasizing early work. Andersen and Gill (1982), Cheuvart (1988), Lawless (1987a), and Thall (1988) consider Poisson models, and Aalen and Husebye (1991), Follmann and Goldberg (1988), and Dabrowska et al. (1994) emphasize renewal models. More general multiplicative intensity-based models are considered by Gail et al. (1980) and Prentice et al. (1981). Much, but by no means all, of this early development was stimulated by advances in survival analysis for the Cox multiplicative hazards model (Cox, 1972a); Andersen et al. (1993) is the authority in this area. Therneau and Grambsch (2000) present analyses based on the Cox model for recurrent events and other event history settings.

Discrete time models for recurrent events are often based on longitudinal binary response models, which are discussed in books on longitudinal data (Diggle et al., 2002; Fahrmeir and Tutz, 2001; Fitzmaurice et al., 2004). Chamberlain (1985, Section 2) gives an insightful discussion of dependence and heterogeneity in binary processes. Borgan et al. (2005) provide an example of discrete time analysis. Models where more than one event may occur at each time point have also been considered, often under the heading of longi-

tudinal count data, or time series of counts (e.g. Diggle et al., 2002, Chapters 10 and 11).

Likelihood construction for prospective observations schemes is rigorously examined by Andersen et al. (1993, Ch. 2), where conditions for the start and stop of observation on individual processes and the validity of (2.53) are discussed. See also Heckman and Singer (1986). Aalen and Husebye (1991) give a short clear synopsis in the context of renewal processes. Berman and Turner (1992) discuss the approximation of (2.53) so as to facilitate analysis using generalized linear model software. Lawless and Zhan (1998) discuss conditions for intermittent observation and interval-count data. Little and Rubin (2002) and Gill et al. (1997) consider more general types of incompleteness in longitudinal data, and Little (1992) surveys missing event history data. Guo (1993), Hamerle (1991), Hoem (1985), Keiding (1991, 2006), and Lawless and Fong (1999) discuss the selection of individuals for event history studies, and cases where information about process history at the time of selection may be missing. Hoem (1985) and Hamerle (1991) discuss retrospective observation and selection effects in some detail.

2.8 Problems and Supplements

2.1. For a Poisson process with rate function $\rho(t)$, let $P_n(t) = \Pr(N(t) = n)$. Show that for small $\Delta t > 0$ and $n = 1, 2, \ldots,$

$$P_n(t + \Delta t) = P_n(t)(1 - \rho(t)\Delta t) + P_{n-1}(t)\rho(t)\Delta t + o(\Delta t)$$

and then prove that if $P_n'(t) = dP_n(t)/dt$, then

$$P_n'(t) = -\rho(t)\{P_n(t) - P_{n-1}(t)\} \qquad n = 1, 2, \ldots .$$

Use this and the fact that $P_0(0) = 1$ and $P_n(0) = 0$ for $n > 0$ to show that

$$P_n(t) = \frac{\mu(t)^n \exp(-\mu(t))}{n!} \qquad n = 0, 1, \ldots .$$

Extend this to prove (2.15).

[Section 2.2]

2.2. A homogeneous Poisson process is easy to simulate by using the result (2.18) for the interevent times. This can be used to simulate a nonhomogeneous process and other event processes.

a. Show that if $\{N(t), 0 \le t\}$ is a Poisson process with intensity function $\rho(t)$ then $N^*(s) = N(t)$ where $s = \int_0^t \rho(u)du = \mu(t)$ is a homogeneous Poisson process with intensity function $\rho^*(s) = 1$. This is result (iv) in Section 2.2.1. Show how this can be used to simulate the nonhomogeneous process $\{N(t), 0 \le t\}$ over the time period $[0, \tau]$. Simulate five realizations of a process with $\rho(t) = 1 + .2t$, over $[0, 5]$.

b. The result in (a) can be obtained from (2.9) in the corollary of Theorem 2.2. Use (2.9) to prove also that for $j = 1, 2, \ldots$ and $t_0 = 0$,

$$E_j = \int_{t_{j-1}}^{t_{j-1}+W_j} \lambda(t|H(t))dt$$

has a standard exponential distribution for a general event process, given t_{j-1} and $H(t_{j-1})$. This can be used to simulate an event process with intensity function $\lambda(t|H(t))$, as described following (2.46) in Section 2.4.

c. Let $\lambda(t|H(t))$ be the intensity for a general recurrent event process and suppose there are always constants λ_j^* such that

$$\lambda(t|H(t), N(t^-) = j-1, t_{j-1}) \le \lambda_j^* \qquad j = 1, 2, \ldots; \quad t_{j-1} < t\,.$$

Prove that the following rejection algorithm (e.g. Ogata, 1981; Daley and Vere-Jones, 1988, pp. 506–507) provides a simulation of the event process.
(i) Set $j = 1$, $t_0 = 0$.
(ii) Choose a suitable λ_j^*.
(iii) Generate E_j from the exponential distribution with mean $1/\lambda_j^*$, and U_j from the uniform distribution on $(0, 1)$.
(iv) If $U_j \le \lambda(t_{j-1} + E_j|H(t_{j-1}), t_{j-1})$, set $t_j = t_{j-1} + E_j$; increase j by one and return to step (ii). Otherwise, leave j unchanged, replace t_{j-1} with $t_{j-1} + E_j$, and return to step (ii).
The existence of the λ_j^* above is not as restrictive as it might look, because we are normally interested in generating a process over some finite time interval $[0, \tau]$.

[Sections 2.2, 2.4]

2.3. Let $\{N(t), 0 \le t\}$ be a Poisson process with intensity function $\rho(t)$.

a. Given that $N(\tau) = n$, find the conditional distribution of the event times $T_1, \ldots, T_n$.
b. Use this to show how you might carry out goodness-of-fit checks for a single homogeneous Poisson process that is observed over $[0, \tau]$. What kind of model checking would you suggest if you have m independent processes observed over $[0, \tau]$?

[Section 2.2]

2.4. Show that for event times $T_1, \ldots, T_n$ from a homogeneous Poisson process over $[0, \tau]$, as in Problem 2.3, the statistic $S_n = \sum_{i=1}^n T_i$ has (conditional on n) mean $n\tau/2$ and variance $n\tau^2/12$. Then argue that for n even moderately large, the conditional distribution of

$$T = \frac{(S_n - n\tau/2)}{(n\tau^2/12)^{1/2}}$$

is approximately $N(0,1)$ if the event process is a homogeneous Poisson process (HPP). Use this to test whether the following data could be considered to come from a HPP: the observations are the times (in hours) of successive equipment failures in an airplane over $\tau = 1400$ hours of operation. If there is evidence against the HPP, what type of departure from it is indicated? The t_j are 487, 505, 605, 612, 710, 715, 800, 891, 934, 1164, 1277, 1297.

[Sections 2.2, 2.6]

2.5. Suppose that $\{N(t), 0 \le t\}$ is a Poisson process with intensity function $\rho(t)$, and that $U_1, U_2, \ldots$ are i.i.d. random variables with finite mean and variance. The process $\{S(t), 0 \le t\}$, where

$$S(t) = \sum_{i=1}^{N(t)} U_i \qquad t > 0\,,$$

is called a marked or compound Poisson process (the U_i are sometimes referred to as "marks"). Such processes are used in settings where there are costs or other variables associated with an event.

a. For the homogeneous Poisson process, where $\rho(t) = \rho$, prove that $E\{S(t)\} = \rho t E(U)$, $\text{var}\{S(t)\} = \rho t E(U^2)$, and that the moment generating function (m.g.f.) of $S(t)$ is

$$M_{S(t)}(z) = \exp\{\rho t[M_U(z) - 1]\}\,,$$

where $M_U(z)$ is the m.g.f. of U. Assume here that all necessary expectations exist.

b. The moment generating functional associated with the process $\{S(t), 0 < t\}$ is defined as (assuming it exists),

$$M_S(z) = E\left\{\exp\left[\int_0^\infty z(t)dS(t)\right]\right\}\,,$$

where $z(t) \le 0$ for $t \ge 0$ and $z(t) = 0$ outside some bounded interval I. Prove that the moment generating functional for $\{S(t), 0 \le t\}$ is

$$\exp\left\{-\int_0^\infty [1 - M_U(z(t))]\rho(t)dt\right\}\,.$$

Use this to find the m.g.f. for $S(t_0)$, where $t_0 > 0$ is a specified value.

[Section 2.2]

2.6. Consider the Poisson random effects model in Section 2.2.3 in which conditional on a random effect u, an individual experiences events according to a Poisson process with intensity function $u\rho(t)$. Furthermore, suppose u has a gamma density $g(u)$ given by (2.28), with mean 1 and variance ϕ.

a. For small Δt, show that to order Δt,

$$\Pr\{N(t, t+\Delta t) = 1 | H(t)\} = \frac{\int_0^\infty u\rho(t)\Delta t \cdot \Pr\{H(t)|u\} g(u) du}{\int_0^\infty \Pr\{H(t)|u\} g(u) du}\,.$$

Then use this to prove (2.33). Plot this (random) intensity for a process with $\rho(t) = \rho$ and events which occur at time $t_1, t_2, \ldots$.

b. Show that the distribution of u, given $H(t)$, is gamma with mean $\{1 + \phi N(t^-)\}/\{1 + \phi\mu(t)\}$ and variance $\{1 + \phi N(t^-)\}/\{1 + \phi\mu(t)\}^2$. Use this to deduce (2.33) directly.

c. Suppose there is a covariate x so that $\rho(t) = \rho_0(t)e^{x'\beta}$. Determine the intensity ratio $\lambda(t|H(t), x = 1)/\lambda(t|H(t), x = 0)$ and show that it depends on t. On the same graph, plot this ratio as a function of t when $\rho(t) = \rho$ and $\phi = 0, 1, 2$, respectively.

[Section 2.2]

2.7. The mean or renewal function $\mu(t)$ of an ordinary renewal process satisfies the *integral equation*

$$\mu(t) = F(t) + \int_0^t \mu(t - x) dF(x)\,, \tag{2.56}$$

where $F(w)$ is the c.d.f. for the gap times W_j in the process.

a. Prove this by using the fact that $E\{N(t)\} = E[E\{N(t)|W_1\}]$. Numerical methods of solving integral equations provide a way to compute $\mu(t)$.

b. Show that (2.39) satisfies (2.56).

c. The mean function corresponding to the gamma gap time distribution with density function

$$f(w; \theta) = \theta^2 w e^{-\theta w} \qquad 0 < w \tag{2.57}$$

can be shown from (2.56) to be

$$\mu(t) = \frac{1}{2}\theta t - \frac{1}{4}(1 - e^{-2\theta t})\,.$$

Show directly that $\mu(t)/t \to 1/E(W_j)$ as $t \to \infty$. Plot $\mu(t)$.

[Section 2.3]

2.8. a. For a renewal counting process $\{N(t), 0 \le t\}$,

$$\frac{E\{N(t)\}}{t} \to \frac{1}{\mu} \qquad \text{as} \quad t \to \infty\,, \tag{2.58}$$

where $\mu = E(W)$ is the mean time between events. This follows from the fact that $\rho(t) \to \mu^{-1}$ as $t \to \infty$, but a simple direct proof can be based on the fact that for any $t > 0$,

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{T_{N(t)+1}}{N(t)}$$

and the fact that $T_{N(t)}$ is a sum of i.i.d. random variables. Prove this.

b. Simulate 100 renewal processes $N_j(t)$, $j = 1, \ldots, 100$ for $0 < t < 10$, where W has a gamma distribution with mean 1 and variance $\phi = 0.25$. Plot $\widehat{\mu}(t)$ in (1.1) versus t. Repeat this when W is gamma with mean 1 and variance 0.01. What do you observe?

[Section 2.3]

2.9. Consider the modulated renewal process based on (2.45) with Gaussian random effects.

a. Obtain (i) the marginal distribution of Y_{ij} and (ii) the conditional distribution of Y_{ij}, given $Y_{i1}, \ldots, Y_{i,j-1}$.
b. Investigate the same thing for the model (2.44), when u_i has a gamma distribution with density (2.28).

[Section 2.3; Aalen and Husebye, 1991]

2.10. Consider a discrete time Bernoulli process where the number of events $n(t)$ at time t ($t = 1, 2, \ldots$) is 0 (with probability $1-p$) or 1 (with probability p), and the $n(t)$ are mutually independent. Describe departures from this model which one might encounter, and outline ways in which extended models could be formulated.

[Section 2.5]

2.11. Consider Bernoulli processes as in Problem 2.10, but suppose that p varies across distinct processes according to a beta distribution with density function

$$g(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \qquad 0 < p < 1\,,$$

where $a > 0, b > 0$, and $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. Prove that

$$\Pr\{n(t) = 1 | H(t)\} = \frac{\sum_{j=1}^{t-1} n(j) + a}{t - 1 + a + b} \qquad t = 2, 3, \ldots\,.$$

[Section 2.5]

2.12. Consider a continuous time process for recurrent events in which only the numbers of events $N(a_{j-1}, a_j)$ are observable, where $0 = a_0 < a_1 < \cdots < a_k$. Give the joint distribution of the counts $N(a_{j-1}, a_j)$, $j = 1, \ldots, k$ for the cases where (i) the underlying process is Poisson with rate function $\rho(t)$, and (ii) the process is Poisson with rate function $u_i\rho(t)$, conditional on the random effect u_i, which has a gamma distribution with density (2.28).

[Section 2.2; Lawless and Zhan, 1998]

2.13. Consider the following different "rules" for terminating observation of a recurrent event process with intensity function $\lambda(t|H(t))$:

(i) Stop observation once the third event has occurred, that is, at time T_3.
(ii) Stop observation at $\tau = \min(T_3, 100)$.
(iii) Stop observation at the smallest time $t > 0$ such that $N(t-1, t) \geq 2$.

Show that (2.54) is satisfied in each case.

[Section 2.6; Aalen and Husebye, 1991]

2.14. Suppose that a clinical study is to involve observing recurrent events that occur for study individuals over the time period $0 \leq t \leq \tau$. However, a condition for including an individual in the study is that $N(-1, 0) \geq 1$; that is, they must have experienced at least one event in the time period $(-1, 0)$ before the study.

Discuss the general ramifications of this entry condition on the analysis of the study. Specifically discuss the situation where

(i) individuals' event processes before and after entry to the study are closely approximated by identical homogeneous Poisson processes, and
(ii) individuals' event processes before and after entry are closely approximated by homogeneous Poisson processes, but with rates that vary according to a random effects model where the jth individual has rate $u_j\rho$, where the u_j are i.i.d. with density (2.28).

[Section 2.6; Cook and Wei, 2003]

2.15. Suppose that the intensity function in (2.55) is specified in terms of a parameter θ. Estimation of θ can be based on a product of likelihood contributions $L(\theta)$, of the form (2.55), across independent event processes. Let $\ell(\theta) = \log L(\theta)$, and consider the likelihood score function $U(\theta) = \partial\ell(\theta)/\partial\theta$; we can typically estimate θ by solving the maximum likelihood score equation $\sum_i U_i(\theta) = 0$, where $i = 1, \ldots, m$ indexes individuals.

a. Show from (2.55) that $\ell(\theta)$ for an individual can be expressed as

$$\int_0^\infty Y(t) \log \lambda(t|H(t);\theta) dN(t) - \int_0^\infty Y(t)\lambda(t|H(t);\theta)dt\,, \qquad (2.59)$$

where $Y(t) = I(\tau_0 \leq t \leq \tau)$ indicates when the process is observed, and we define $dN(t) = N(t) - N(t^-)$ and use the Riemann–Stieltjes integral defined just after (2.4) for the first term in (2.59). Note that because $N(t)$ is a step function with jumps of size one, the first term in (2.59) is simply the sum

$$\sum_{j=1}^{n} \log \lambda(t_j|H(t_j);\theta)\,,$$

where $t_1, \ldots, t_n$ are the observed event times for the indivdual.

b. Show that $U(\theta)$ for an individual can be expressed as

$$U(\theta) = \int_0^\infty Y(t) \frac{\partial \log \lambda(t|H(t);\theta)}{\partial \theta} \{dN(t) - \lambda(t|H(t);\theta)dt\}\,, \qquad (2.60)$$

assuming one can differentiate through the integral. Prove that $E_\theta\{U(\theta)\} = 0$; that is, the score function is unbiased. To do this, use results in Section 2.1 to argue that we can notionally write

$$E\{dN(t)|H(t)\} = \lambda(t|H(t))dt\,,$$

as a consequence of

$$E\{\Delta N(t)|H(t)\} = \lambda(t|H(t))\Delta t + o(\Delta t)\,.$$

c. Consider m independent individuals, with each having events according to a homogeneous Poisson process with rate ρ. Obtain the likelihood score function and maximum likelihood estimate $\widehat{\rho}$. Assuming that individual i is observed over the prespecified time interval $[0, \tau_i]$, derive the variance of $\widehat{\rho}$.

[Sections 2.1, 2.7]

3

Methods Based on Counts and Rate Functions

3.1 Introduction

In many settings the event of interest represents a transient adverse experience for which there is little immediate impact on the event generating process. An example is a mild seizure in a study of epileptic patients. Although such events are undesirable, their occurrence does not materially affect the risk of subsequent seizures and so analyses based on event counts and the rate of event occurrence are natural.

The focus of this chapter is on methods for the analysis of recurrent event data based on rate functions and counts of events. The Poisson process is the canonical model for this setting but more general models may be formulated based on extensions of it. Models that involve only the specification of marginal rate functions $\rho(t)$ or mean functions $\mu(t)$ can also be developed. They provide a basis for the development of robust methods of inference because they don't involve assumptions regarding the underlying stochastic process, but naturally do not allow the calculation of probabilities for event occurrence. Such methods are presented in Sections 3.6 to 3.8.

Suppose m individuals are each under observation from time $t = 0$ to a censoring or stopping time. The notation introduced in Chapters 1 and 2 is used. If τ_i denotes the time at the end of observation for individual i, the left-continuous function $Y_i(t) = I(t \leq \tau_i)$ indicates whether individual i is observed at time t. The observation process is assumed to be conditionally independent of the event process in the sense of (2.54). We can also accommodate settings where individual i is observed from $\tau_{i0} > 0$ to τ_i by redefining $Y_i(t)$ as $I(\tau_{i0} \leq t \leq \tau_i)$ as in Section 2.6; results below are all valid with this change. We assume in this chapter that during periods in which an individual is under observation, the exact times of events are obtained. We can therefore observe $\{N_i(t), 0 \leq t\}$, or equivalently $\{dN_i(t), 0 \leq t\}$, over $[0, \tau_i]$, $i = 1, \ldots, m$. We let $H_i(t) = \{N_i(s) : 0 \leq s < t\}$ denote the history of the event process for individual i. For Poisson processes, the intensity function for individual i is

$$\lambda_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N_i(t) = 1\}}{\Delta t} = \rho(t) \tag{3.1}$$

at time $t \geq 0$. As noted in Section 2.2, $\rho(t)$ is also the rate function.

Suppose $\{x_i(t), 0 \leq t\}$ is a left-continuous $p \times 1$ covariate process which includes only external covariates, with $x_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))'$. As discussed in Section 2.1, the history of the process is then broadened to $H_i(t) = \{N_i(s) : 0 \leq s < t; x_i^{(\infty)}\}$, so that fixed and time-varying covariate values are assumed part of $H_i(0)$. Covariate effects may be specified quite generally, but by far the most common framework is through multiplicative models of the form $\lambda_i(t|H_i(t)) = \rho_i(t)$ with

$$\rho_i(t) = \rho_0(t;\alpha) g(x_i(t);\beta)\,, \tag{3.2}$$

where $\rho_0(t;\alpha)$ is a so-called baseline rate function applicable for subjects with $x_i(t) = 0$, and $g(x(t);\beta)$ is a nonnegative function. Frequently $g(x_i(t);\beta) = \exp(x_i'(t)\beta)$, as in (2.19), in which case $\exp(\beta_k)$ is the multiplicative effect on the intensity of a one-unit increase in $x_{ik}(t)$ and all other covariates are fixed. This is the formulation we adopt here, and probability calculations are conditional on the covariate process. Recall from Section 2.1 that although the intensity is defined given $x_i^{(\infty)}$, at t it depends only on $x_i^{(t)}$. It should also be noted that $x_i(t)$ may depend on variables measured prior to time t, as in (2.19). For example, an air pollution covariate $x_i(t)$ could be an average of pollution counts taken over the previous 48 hours.

The baseline rate function is parameterized by α and we let $\theta = (\alpha', \beta')'$ denote the full vector of parameters. In this chapter we consider both fully parametric models, where α is finite-dimensional, and semiparametric models, where α is considered infinite-dimensional. We write $\rho_i(t;\theta) = \rho_0(t;\alpha)\exp(x_i'(t)\beta)$ when it is important to make the dependence on the parameters explicit, and simply write $\rho_i(t)$ when it is not.

If the observation process is completely independent of the event occurrence process, then one can condition on $Y^{(\infty)} = \{Y_i(s), 0 \leq s : i = 1, \ldots, m\}$ and we get the resulting likelihood $L(\theta)$ from (2.7) and (3.2) based on the conditional probability density of the observed outcomes "n_i events, at times $t_{i1} < \cdots < t_{in_i}$, for individual i $(i = 1, \ldots, m)$." This gives

$$L(\theta) = \prod_{i=1}^{m} L_i(\theta)\,,$$

where

$$L_i(\theta) = \prod_{j=1}^{n_i} \{\rho_0(t_{ij})\exp(x_i'(t_{ij})\beta)\} \exp\left(-\int_0^\tau Y_i(s)\rho_0(s)\exp(x_i'(s)\beta)ds\right)\,, \tag{3.3}$$

and where $\tau = \max(\tau_1, \ldots, \tau_m)$.

If the observation process is not completely independent of the event occurrence process, then, as discussed in Section 2.6, $L(\theta)$ is a partial likelihood

and can still be treated as a standard likelihood function provided that $Y_i(t)$ depends only on information in $H_i(t)$. Maximizing $L(\theta)$ then yields the partial maximum likelihood estimator $\widehat{\theta}$, which under mild regularity conditions has the usual asymptotic properties.

In the remainder of this chapter we describe methodology for estimating rate and mean functions. Examples of the various methods are given in Section 3.8.

3.2 Parametric Maximum Likelihood for Poisson Models

In this section we develop maximum likelihood methods for Poisson processes; Appendix A reviews the general theory.

3.2.1 Score and Information Functions

Suppose $\rho_0(t;\alpha)$ is indexed by an $r \times 1$ parameter α and β is a $p \times 1$ vector of regression coefficients. The maximum likelihood score equations for θ arise from differentiating $\ell(\theta) = \log L(\theta)$ where $L(\theta)$ is a product of terms (3.3). Their general form is as given by (2.59) and (2.60) in Problem 2.15, and we use the notation introduced there, writing $\ell(\theta)$ as

$$\ell(\theta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s)[\log \rho_i(s;\theta) dN_i(s) - \rho_i(s;\theta) ds] \,.$$

If we let $U_\alpha(\theta) = \partial \ell(\theta)/\partial \alpha$ be the $r \times 1$ score vector for α and $U_\beta(\theta) = \partial \ell(\theta)/\partial \beta$ denote the $p \times 1$ score vector for β, we find that the maximum likelihood equations are

$$U_\alpha(\theta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s) \frac{\partial \log \rho_0(s;\alpha)}{\partial \alpha} \{dN_i(s) - \rho_i(s;\theta) ds\} = 0 \qquad (3.4)$$

$$U_\beta(\theta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s) x_i(s) \{dN_i(s) - \rho_i(s;\theta) ds\} = 0 \,. \qquad (3.5)$$

Let $U(\theta) = (U'_\alpha(\theta), U'_\beta(\theta))'$ denote the full score vector.

The components of the observed information are $I_{\alpha\alpha}(\theta) = -\partial U_\alpha(\theta)/\partial \alpha'$, $I_{\alpha\beta}(\theta) = -\partial U_\alpha(\theta)/\partial \beta'$, $I_{\beta\alpha}(\theta) = -\partial U_\beta(\theta)/\partial \alpha'$, and $I_{\beta\beta}(\theta) = -\partial U_\beta(\theta)/\partial \beta'$. These are given by

$$I_{\alpha\alpha}(\theta) = -\sum_{i=1}^{m} \int_0^\tau Y_i(s) \frac{\partial^2 \log \rho_0(s;\alpha)}{\partial \alpha \partial \alpha'} \{dN_i(s) - \rho_i(s;\theta) ds\}$$

$$+\sum_{i=1}^{m}\int_0^\tau Y_i(s)\left\{\frac{\partial\log\rho_0(s;\alpha)}{\partial\alpha}\frac{\partial\log\rho_0(s;\alpha)}{\partial\alpha'}\right\}\rho_i(s;\theta)ds$$

$$I_{\alpha\beta}(\theta)=\sum_{i=1}^{m}\int_0^\tau Y_i(s)\frac{\partial\log\rho_0(s;\alpha)}{\partial\alpha}\,x_i(s)\rho_i(s;\theta)ds$$

$$I_{\beta\alpha}(\theta)=\sum_{i=1}^{m}\int_0^\tau Y_i(s)\frac{\partial\log\rho_0(s;\alpha)}{\partial\alpha}\,x_i'(s)\rho_i(s;\theta)ds$$

$$I_{\beta\beta}(\theta)=\sum_{i=1}^{m}\int_0^\tau Y_i(s)x_i(s)x_i'(s)\rho_i(s;\theta)ds$$

which give the observed information matrix

$$I(\theta)=\begin{pmatrix} I_{\alpha\alpha}(\theta) & I_{\alpha\beta}(\theta)\\ I_{\beta\alpha}(\theta) & I_{\beta\beta}(\theta)\end{pmatrix}.$$

Typically, but not always, the maximum likelihood estimate $\widehat{\theta}$ is unique and satisfies the score equations (3.4) and (3.5). General optimization software (see Appendix B) can be employed to maximize $\ell(\theta)$. Good implementations make specification of derivatives optional, and will produce the Hessian matrix $-I(\widehat{\theta})$ by numerical differentiation, if requested. If one wishes to program the optimization directly, a Newton–Raphson algorithm is often quite useful; it uses an initial estimate $\tilde{\theta}^{(1)}$ and the iterative scheme

$$\tilde{\theta}^{(k+1)}=\tilde{\theta}^{(k)}+I^{-1}(\tilde{\theta}^{(k)})U(\tilde{\theta}^{(k)})\qquad k=1,2,\ldots,$$

where iterations are terminated when the difference between estimates and between values $\ell(\theta)$ at successive steps is less than prespecified tolerances. Alternatively, a Fisher-scoring algorithm may be used by replacing $I(\theta)$ with the expected information matrix $\mathcal{I}(\theta)=E\{I(\theta)\}$. The expected and observed information matrices differ in that the first term of $I_{\alpha\alpha}(\theta)$ vanishes upon taking the expectation. Regardless of the method of optimization, model based variance estimates for the maximum likelihood estimators are available from $I^{-1}(\widehat{\theta})$ or $\mathcal{I}^{-1}(\widehat{\theta})$ and for large samples ($m\to\infty$) we can act, for example, as though $(\widehat{\theta}-\theta)\sim N(0,I^{-1}(\widehat{\theta}))$. An alternative is to use likelihood ratio methods; see Appendix A. Note that whatever methods are used, numerical integration may be necessary in order to evaluate $\ell(\theta)$ and its derivatives. Good software for doing this exists, but see Berman and Turner (1992) or Lawless and Thiagarajah (1996) for algorithms that are easy to implement.

3.2.2 A General Parametric Rate Function

Models where $\rho_0(t;\alpha)$ is either of the form $\exp(\alpha_0+\alpha_1 t)$ or $\alpha_0\alpha_1 t^{\alpha_1-1}$ are often used in settings where the rate function is monotonic. A flexible general family of parametric models that includes these is given by

$$\rho_i(t;\theta) = \exp(z_i'(t)\theta)\,, \tag{3.6}$$

where $z_i(t)$ includes known functions of t and the covariates in $x_i(t)$. For example, $z_i(t) = (1, t, x_i'(t))'$ could be adopted if the log of the baseline rate function changes linearly with t and a time-dependent covariate is present. Alternatively one might specify $z_i(t) = (1, \log t, x_i'(t))'$ if the rate function changes as a power of t. By taking the exponential of the linear combination $z_i'(t)\theta$, we guarantee rate functions are positive.

For any particular form of $z(t)$ under (3.6), the log-likelihood, score vector, and observed information matrix for θ are, from (3.3),

$$\ell(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} z_i'(t_{ij})\theta - \int_0^\tau Y_i(s)\exp(z_i'(s)\theta)ds \right\}\,, \tag{3.7}$$

$$U(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} z_i(t_{ij}) - \int_0^\tau Y_i(s)z_i(s)\exp(z_i'(s)\theta)ds \right\} \tag{3.8}$$

$$I(\theta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s)z_i(s)z_i'(s)\exp(z_i'(s)\theta)ds\,. \tag{3.9}$$

It is shown in Chapter 5 (see Problem 5.1) that (3.7) is a convex function of θ. The maximum likelihood equations $U(\theta) = 0$ are usually easy to solve, say via general-purpose software or directly by Newton's method, with numerical integration used to evaluate the integrals in (3.7), (3.8), and (3.9) if necessary.

3.2.3 Time Transform Models

In survival analysis accelerated failure time regression models are a useful alternative to multiplicative hazards models in many settings (e.g. Lawless 2003a, Ch. 6). They are examples of *time transform models*, which can be used in the recurrent event setting as described in (2.23)–(2.25). In particular, suppose there are vectors of fixed covariates x_i associated with individuals $i = 1, \ldots, m$ and that given x_i, the recurrent event process $\{N_i(t), 0 \le t\}$ is Poisson with rate function of the form

$$\rho_i(t) = e^{x_i'\beta}\rho_0(e^{x_i'\beta}t)\,,$$

where $\rho_0(t)$ is a specified baseline rate function. The corresponding mean function is

$$\mu_i(t) = E\{N_i(t)|x_i\} = \mu_0(e^{x_i'\beta}t)\,,$$

where $\mu_0(t) = \int_0^t \rho_0(s)ds$.

This model is called a time transform model because the effect of the covariate x is to transform the time scale from t to $\exp(x'\beta)t$; thus the covariates define a time scale $s_i = \exp(x_i'\beta)t$ for each individual, on which the

rate function is $\rho_0(s_i)$. The exponential form $\exp(x'\beta)$ has been used for convenience, because it is flexible and gives a monotone increasing transform of t. If desired, $\exp(x'\beta)$ can be replaced with some other positive-valued function $g(x;\beta)$. The models can also be extended to accommodate time-varying covariates $x(t)$, by specifying

$$\rho_i(t) = e^{x_i'(t)\beta}\rho_0(e^{x_i'(t)\beta}t)\,.$$

As with multiplicative models, the mean function $\mu_i(t)$ is in this case a complicated function of the covariate history $x_i^{(t)}$, obtained by integration of $\rho_i(t)$.

Parametric time transform models are relatively easily handled. In this case $\rho_0(t)$ is specified parametrically as $\rho_0(t;\alpha)$ and the likelihood function based on data from m independent individuals is given by

$$L(\alpha,\beta) = \prod_{i=1}^{m}\left\{\prod_{j=1}^{n_i} e^{x_i'(t_{ij})\beta}\rho_0(e^{x_i'(t_{ij})\beta}t_{ij})\right\}\exp\{-\mu_i(\tau_i)\}\,,$$

where for simplicity we assume that individual i is observed over $[0,\tau_i]$. When covariates are fixed, the likelihood becomes

$$L\,(\alpha,\beta) = \prod_{i=1}^{m}\left\{\prod_{j=1}^{n_i} e^{x_i'\beta}\rho_0\left(e^{x_i'\beta}t_{ij}\right)\right\}\exp\left\{-\mu_0(e^{x_i'\beta}\tau_i)\right\}\,,$$

where functions $\rho_0(\cdot)$ and $\mu_0(\cdot)$ are specified by the parameter vector α. With fixed covariates, these models can be fitted using survival analysis software for accelerated failure time models; this is described in the next section.

3.2.4 Using Survival Software

Some useful parametric rate functions have the form of the hazard function for one of the common lifetime distributions. In such cases, survival analysis software that allows for delayed entry can be used for estimation and inference. This applies in particular to certain cases where there are no covariates and to cases where $\rho_i(t)$ has a time transform structure of the preceding section.

If we let $t_{i0} = 0$, (3.3) can be written as

$$\prod_{j=1}^{n_i}\left[\rho_i(t_{ij})\exp\left\{-\int_{t_{i,j-1}}^{t_{ij}}\rho_i(s)ds\right\}\right]\times\exp\left\{-\int_{t_{i,n_i}}^{\tau_i}\rho_i(s)ds\right\}\,. \qquad (3.10)$$

Expression (3.10) has the same form as the likelihood arising from a sample of n_i+1 independent survival time observations over the "at risk" intervals $(t_{i,j-1},t_{ij}]$, $j=1,\ldots,n_i+1$, where $t_{i,n_i+1}=\tau_i$, with each interval except the last ending with an event. The likelihood for the full sample is a product of

terms (3.10) over all subjects. An application to the rat tumor data of Section 1.2.1 is given in Section 3.8.1.

With time-dependent covariates we have $\rho_i(t) = \rho_0(t;\alpha)\exp(x_i'(t)\beta)$, and as can be seen by (3.10), integrals of the form $\int_s^t \rho_i(u)du$ are required. In this case values for the time-dependent covariates must be available at all time points over $[0, \tau_i)$. In some cases this is possible, but in many settings this information may be difficult to obtain. Often time-dependent covariates are assumed to be constant between periodic assessment times which we might denote $0 = s_{i0} < s_{i1} < \cdots < s_{ik_i}$. In this case, the integrals are easy to compute over the intersections of the intervals $(t_{i,j-1}, t_{ij}]$, $j = 1, \ldots, n_i + 1$ and $(s_{i,k-1}, s_{ik}]$, $k = 1, \ldots, k_i$ inasmuch as for each such intersection we have a model of a form leading to (3.10).

If covariates are fixed over the intervals $(t_{i,j-1}, t_{ij}]$ and $\rho_i(t)$ has the form of certain accelerated failure time hazard functions, then parametric survival analysis software such as `censorReg` in S-PLUS or `survreg` in R may be used. For example,

$$\rho_i(t) = \exp(x_i'\beta)\rho_0(\exp(x_i'\beta)t)$$

with $\rho_0(t) = \alpha_2\alpha_1^{-\alpha_2}(t/\alpha_1)^{\alpha_2-1}$ is a Weibull accelerated failure time model (Lawless, 2003a, Section 6.3.2) and can be handled by many survival software packages. Other models where $\rho_0(t)$ takes the form of a log-normal, log-logistic, or a gamma distribution hazard function can similarly be handled.

3.3 Poisson Models with Piecewise-Constant Rates

The specific models mentioned in Section 3.2 all have rate functions with a small number of parameters. More flexible parametric models may be obtained by using splines or piecewise-constant baseline rate functions. Here we consider the latter; they provide useful flexibility in some settings, and also have a connection with the semiparametric methods considered in the next section.

As the name implies, under such models the rate function is assumed to be constant over prespecified intervals. In particular, let $a_0 < a_1 < \cdots < a_K$ denote K cutpoints such that $a_0 = 0$ and $a_K = \tau$. The baseline rate function is then given as

$$\rho_0(t;\alpha) = \alpha_k \qquad a_{k-1} < t \le a_k\,, \tag{3.11}$$

and $\alpha = (\alpha_1, \ldots, \alpha_K)'$ is the parameter that characterizes the baseline rate. These models have rate functions with discontinuities at the cutpoints, but can provide good approximations to various shapes of functions. Here the dimension of α can become arbitrarily large but, although greater flexibility is achieved by using a larger number of pieces, models involving three to ten pieces with cutpoints evenly distributed over the event times are flexible enough for most practical situations.

Let $w_k(t) = I(a_{k-1} < t \le a_k)$, $k = 1, \ldots, K$ indicate whether $t \in (a_{k-1}, a_k]$, $n_{ik} = \sum_{j=1}^{n_i} w_k(t_{ij})$ denote the total number of events experienced by subject i in $(a_{k-1}, a_k]$, and $n_{\cdot k} = \sum_{i=1}^{m} n_{ik}$ denote the total number of events experienced by all subjects over $(a_{k-1}, a_k]$, $k = 1, \ldots, K$. Suppose $\lambda_i(t|H_i(t)) = \rho_0(t)\exp(x_i'(t)\beta)$, and $\rho_0(t)$ is given by (3.11). By (3.3), $L(\theta)$ can be rewritten as $L(\theta) = \prod_{k=1}^{K} L_k(\theta)$, where $L_k(\theta)$ is given as

$$\alpha_k^{n_{\cdot k}} \prod_{i=1}^{m} \left\{ \exp\left(\sum_{j=1}^{n_i} x_i'(t_{ij})\beta w_k(t_{ij}) - \alpha_k \int_{a_{k-1}}^{a_k} Y_i(s)\exp\left(x_i'(s)\beta\right) ds \right) \right\} . \quad (3.12)$$

The resulting log-likelihood is of the form

$$\ell(\theta) = \sum_{k=1}^{K} n_{\cdot k}\log\alpha_k + \sum_{i=1}^{m}\left\{ \sum_{j=1}^{n_i} x_i'(t_{ij})\beta - \sum_{k=1}^{K}\alpha_k \int_{a_{k-1}}^{a_k} Y_i(s)\exp(x_i'(s)\beta)ds \right\}$$

and the score vector $U(\theta) = (U_\alpha'(\theta), U_\beta'(\theta))'$ has elements

$$U_{\alpha_k}(\theta) = \frac{\partial\ell(\theta)}{\partial\alpha_k} = \frac{n_{\cdot k}}{\alpha_k} - \sum_{i=1}^{m}\int_{a_{k-1}}^{a_k} Y_i(s)\exp(x_i'(s)\beta)ds, \qquad k = 1, 2, \ldots, K ,$$

where $U_\alpha(\theta) = (U_{\alpha_1}(\theta), \ldots, U_{\alpha_K}(\theta))'$ and

$$U_\beta(\theta) = \frac{\partial\ell(\theta)}{\partial\beta} = \sum_{i=1}^{m}\left[\sum_{j=1}^{n_i} x_i'(t_{ij}) - \sum_{k=1}^{K}\alpha_k \int_{a_{k-1}}^{a_k} Y_i(s)\exp(x_i'(s)\beta)x_i'(s)ds \right] .$$

Because $\rho_i(t) = \sum_{k=1}^{K} w_k(t)\alpha_k\exp(x_i'(t)\beta)$, $U_\beta(\theta)$ can be written as in (3.5).

Solving $U_\alpha(\theta) = \partial\ell(\theta)/\partial\alpha = 0$ gives profile likelihood estimates

$$\widetilde{\alpha}_k(\beta) = \frac{n_{\cdot k}}{\sum_{i=1}^{m}\int_{a_{k-1}}^{a_k} Y_i(s)\exp(x_i'(s)\beta)ds} , \qquad k = 1, \ldots, K . \quad (3.13)$$

The profile likelihood function for β is obtained by inserting $\widetilde{\alpha}(\beta) = (\tilde{\alpha}_1(\beta), \ldots, \tilde{\alpha}_K(\beta))'$ into (3.12). Specifically, $L_P(\beta) = L(\widetilde{\alpha}(\beta), \beta)$ is

$$L_P(\beta) = \prod_{i=1}^{m}\prod_{j=1}^{n_i}\left\{ \frac{\exp(x_i'(t_{ij})\beta)}{\sum_{k=1}^{K} w_k(t_{ij})\sum_{\ell=1}^{m}\int_{a_{k-1}}^{a_k} Y_\ell(s)\exp(x_\ell'(s)\beta)ds} \right\} , \quad (3.14)$$

which in turn can be maximized to give $\widehat{\beta}$. The profile score equation is $\partial\log L_P(\beta)/\partial\beta = 0$, which may be written explicitly as

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{x_i(t_{ij}) - \frac{\sum_{k=1}^{K} w_k(t_{ij}) \int_{a_{k-1}}^{a_k} \sum_{\ell=1}^{m} Y_\ell(s)\exp(x'_\ell(s)\beta)x_\ell(s)ds}{\sum_{k=1}^{K} w_k(t_{ij}) \int_{a_{k-1}}^{a_k} \sum_{\ell=1}^{m} Y_\ell(s)\exp(x'_\ell(s)\beta)ds}\right\} = 0\,.$$

The solution to this equation is $\widehat{\beta}$, and inserting it into (3.13) gives the maximum likelihood estimate $\widehat{\alpha} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K)'$. Variance estimates for $\widehat{\theta} = (\widehat{\alpha}', \widehat{\beta}')'$ can be obtained from the inverse of the information matrix $I^{-1}(\widehat{\theta})$. If only variance estimates for $\widehat{\beta}$ are wanted they can alternatively be taken from $I_P^{-1}(\widehat{\beta})$, where $I_P(\beta) = -\partial^2 \ell_P(\beta)/\partial\beta\partial\beta'$.

When the covariates are all fixed, the likelihood $L(\theta)$ simplifies to a product of Poisson likelihoods,

$$L(\theta) = \prod_{k=1}^{K}\left\{\prod_{i=1}^{m}(\alpha_k \exp(x'_i\beta))^{n_{ik}} \exp(-S_{ik}\alpha_k \exp(x'_i\beta))\right\}, \tag{3.15}$$

where $S_{ik} = \int_0^\infty Y_i(u)w_k(u)du$ is the exposure time in $(a_{k-1}, a_k]$ for individual i. The likelihood (3.15) can be maximized using Poisson log-linear regression software where $\mu_{ik} = \exp(\log \alpha_k + x'_i\beta + \log S_{ik})$ is the mean of the response N_{ik}. The variance estimates for $\widehat{\theta}$ delivered by Poisson log-linear software (e.g. the `glm` function in S-PLUS or R) are also valid, provided that the observational conditions described earlier are met. An illustration is given in Section 3.8.1.

In the case where there are no covariates, (3.15) gives the maximum likelihood estimates $\widehat{\alpha}_k = n_{\cdot k}/S_{\cdot k}$, where $S_{\cdot k} = \sum_{i=1}^{m} S_{ik}$ is the total exposed time in $(a_{k-1}, a_k]$ across all individuals. The $\widehat{\alpha_k}$ are easily seen to be independent, with variances estimated by $\widehat{\alpha}_k/S_{\cdot k}$. The estimated mean function is

$$\widehat{\mu}(t) = \sum_{k=1}^{K} \widehat{\alpha}_k\, u_k(t) = \sum_{k=1}^{K} \widehat{\alpha}_k \int_{a_{k-1}}^{a_k} I(t \geq s)ds\,.$$

We observe that when K is large, this estimate is quite close to the Nelson–Aalen estimate (3.17) described in the next section.

Piecewise-constant models are often used in applications in demography and epidemiology, where analyses of population rates for events such as pregnancies, births, and disease occurrence are often wanted; for example, see Andersen et al. (1993, p. 408). Lawless (1998) discusses similar applications to warranty data analysis. It is important in such applications to consider the possibility of extra-Poisson variation; this is discussed in Section 3.5.3 for parametric models in general.

3.4 Nonparametric and Semiparametric Poisson Models

3.4.1 Nonparametric Inference

Methods that do not make parametric assumptions about baseline rate functions are appealing in many settings. First, we consider the case of a sample of m individuals each providing data on events realized from the same Poisson process with rate function $\rho(t)$, $0 < t$. We begin with a heuristic development of nonparametric estimation that looks at $d\mu(t) = \rho(t)dt$ as the expected value of $N(t, t + dt)$ and treats these, for a partition of the time axis, as the model parameters α in (3.4). In the absence of covariates, the Poisson estimating function for $d\mu(s)$ from (3.4) is then

$$\sum_{i=1}^{m} Y_i(s)\{dN_i(s) - d\mu(s)\}\,. \tag{3.16}$$

Setting (3.16) equal to zero and solving for $d\mu(s)$ gives $d\widehat{\mu}(s) = d\bar{N}_\cdot(s)/Y_\cdot(s)$, where $d\bar{N}_\cdot(s) = \sum_{i=1}^{m} Y_i(s)dN_i(s)$ and $Y_\cdot(s) = \sum_{i=1}^{m} Y_i(s)$ are the total number of events observed and the total number of subjects at risk over $[s, s+ds)$, respectively. Because $E\{d\widehat{\mu}(s)\} = E[E\{d\bar{N}_\cdot(s)/Y_\cdot(s)|Y_1(s), \ldots, Y_m(s)\}] = d\mu(s)$, provided that $E\{d\bar{N}_i(s)|H(s), Y_i(s) = 1\} = d\mu(s)$, the estimator arising from (3.16) is unbiased.

Because $\mu(t) = \int_0^t d\mu(s)$, we obtain an estimate for $\mu(t)$ as

$$\widehat{\mu}(t) = \int_0^t d\widehat{\mu}(s) = \int_0^t \frac{d\bar{N}_\cdot(s)}{Y_\cdot(s)} = \sum_{h:t_{(h)}\le t} \frac{d\bar{N}_\cdot(t_{(h)})}{Y_\cdot(t_{(h)})}\,, \tag{3.17}$$

where here $t_{(1)} < t_{(2)} < \cdots < t_{(H)}$ denote the H distinct event times across all individuals. The estimator (3.17) is the same as the Nelson–Aalen estimator from survival analysis but here it is viewed as a nonparametric maximum likelihood estimator of the mean function for Poisson processes.

It should be noted that, strictly speaking, (3.17) applies only for settings where $Y_\cdot(u) > 0$ for $0 \le u \le t$. In some settings there may occasionally be cases where this is not true for every $t \le \tau$; the main case is where individual i is observed for $t \ge \tau_{i0}$, and $\tau_0 = \min(\tau_{10}, \ldots, \tau_{m0}) > 0$. In that case, we can use the convention in (3.17) that $0/0 = 0$, but must remember that $\widehat{\mu}(t)$ is actually $\widehat{\mu}(\tau_0, t) = \int_{\tau_0}^t d\widehat{\mu}(s)$; there is no information about $\mu(t)$ for $0 \le t < \tau_0$. More generally, if there is a region A over $[0, t]$ where $Y_\cdot(s) = 0$ for a given dataset, then we are able to estimate only $\mu(t) - \int_A d\mu(s)$.

Because $\mathrm{var}\{d\bar{N}_\cdot(s)|Y_1(s), \ldots, Y_m(s)\} = E\{d\bar{N}_\cdot(s)|Y_1(s), \ldots, Y_m(s)\}$ is simply $Y_\cdot(s)d\mu(s)$ when individuals generate data under independent Poisson processes, we have $\mathrm{var}\{d\widehat{\mu}(s)|Y_1(s), \ldots, Y_m(s)\} = d\mu(s)/Y_\cdot(s)$. Inserting the Nelson–Aalen estimate gives $\widehat{\mathrm{var}}\{d\widehat{\mu}(s)|Y_1(s), \ldots, Y_m(s)\} = d\bar{N}_\cdot(s)/Y_\cdot^2(s)$, which leads to the variance estimate

$$\widehat{\text{var}}\{\widehat{\mu}(t)\} = \sum_{h:t_{(h)}\le t} \frac{d\bar{N}_{\cdot}(t_{(h)})}{Y_{\cdot}(t_{(h)})^2} \tag{3.18}$$

from the independent increments property of the Poisson model (see (ii) in Section 2.2.1). A large sample $100(1-\alpha)\%$ confidence interval can be constructed for $\mu(t)$ as

$$\widehat{\mu}(t) \pm z_{\alpha/2}\sqrt{\widehat{\text{var}}(\widehat{\mu}(t))}\,,$$

where z_p is the upper $100p\%$ point of the standard normal distribution. This approach does not preclude a confidence interval from including negative values. An alternative approach is to note that by Taylor series expansion or the delta method,

$$\widehat{\text{var}}\{\log\widehat{\mu}(t)\} = \widehat{\text{var}}\{\widehat{\mu}(t)\}/\widehat{\mu}^2(t)\,.$$

In this case, approximate confidence intervals may be obtained for $\log\mu(t)$ and the resulting limits may be exponentiated to give the interval for $\mu(t)$ as

$$(\exp(\log(\widehat{\mu}(t)) - z_{\alpha/2}\sqrt{\widehat{\text{var}}\{\log\widehat{\mu}(t)\}}), \exp(\log(\widehat{\mu}(t)) + z_{\alpha/2}\sqrt{\widehat{\text{var}}\{\log\widehat{\mu}(t)\}}))\,.$$

Inadmissable negative values will never be included in such intervals. This also gives interval estimates whose true coverage in finite samples is typically closer to the nominal level than the method based on the orginal scale.

Smooth estimates of a common rate function $\rho(t)$ can also be obtained; these are useful in providing insights into the shape of $\rho(t)$ in addition to what can be seen from a plot of the Nelson–Aalen estimate $\widehat{\mu}(t)$. Kernel estimation (Ramlau–Hansen, 1983) is easy to use and operates as follows. Let $K(x)$ be a bounded function which is zero outside $[-1,1]$ and integrates to 1. If $\widehat{\mu}(t)$ is the Nelson–Aalen estimator of $\mu(t)$, then a smooth nonparametric estimate of $\rho(t)$ is given by

$$\widehat{\rho}(t) = b^{-1}\int_{t-b}^{t+b} K\left(\frac{t-s}{b}\right) d\widehat{\mu}(s)\,,$$

where $b>0$ is a constant termed the *bandwidth*. We restrict ourselves here to estimating $\rho(t)$ only at t for which $[t-b, t+b]$ lies within the range of the observed event times t_{ij}, which define points at which $d\widehat{\mu}(s)>0$. Note in fact that

$$\widehat{\rho}(t) = b^{-1}\sum_{t_{ij}} K\left(\frac{t-t_{ij}}{b}\right)\frac{1}{Y_{\cdot}(t_{ij})}\,,$$

where the sum for a given value of t is over only the event times t_{ij} that satisfy $t-b\le t_{ij}\le t+b$.

Various choices of kernel function $K(x)$ on $[-1,1]$ have been proposed, including uniform and triangular probability density functions, trimmed Gaussian density functions, and the Epanechnikov kernel function $K(x) = 0.75(1-x^2)$. A Gaussian kernel function is often the default in software.

The choice of b determines the smoothness of the estimated rate function, with larger values of b giving smoother functions. Variance estimates for $\widehat{\rho}(t)$ are available (e.g. Andersen et al. 1993, Section 4.2), and there are also automated approaches to the selection of b. The most valuable aspect of $\widehat{\rho}(t)$, however, is visual, and in practice a good approach is to compute and plot estimates with a number of values of b. A point to remember is that $\widehat{\rho}(t)$ is, for fixed b, a consistent estimator not of $\rho(t)$ but rather of

$$\rho^*(t) = b^{-1} \int_{t-b}^{t+b} K\left(\frac{t-s}{b}\right) \rho(s) ds\,.$$

To achieve convergence to $\rho(t)$, it is necessary to consider asymptotics in which $b = b_m$ depends on m, with $b_m \to 0$ as $m \to \infty$ (e.g. Andersen et al., 1993, Section 4.2.2).

Kernel density estimation procedures in statistical software can be used to compute estimates $\widehat{\rho}(t)$. In S-PLUS or R, the function `ksmooth` can do this, although many other functions are also available. Venables and Ripley (2002, Section 5.6) discuss this as well as alternative approaches such as parametric spline models for $\rho(t)$; the latter have advantages when only interval counts rather than exact event times are observed (see Section 7.1).

3.4.2 Semiparametric Regression

We now consider estimation and inference for the semiparametric regression model $\rho_i(t) = \rho_0(t) \exp(x_i'(t)\beta)$ where $\rho_0(t)$, the baseline rate function, is not assumed to have any particular parametric form. As noted in Section 2.2.2, this is sometimes called the *Andersen–Gill model*. The standard estimation procedures can be derived in more than one way, and we begin by considering a profile likelihood approach.

Profile Likelihood

With the semiparametric specification we operate as in Section 3.4.1 and replace (3.4) with the following score equation, which treats $d\mu_0(t) = \rho_0(t)dt$ as a parameter,

$$\sum_{i=1}^{m} Y_i(s)\{dN_i(s) - \exp(x_i'(s)\beta) d\mu_0(s)\} = 0 \qquad 0 \le s\,. \tag{3.19}$$

Solving (3.19) for $d\mu(s)$ gives the profile likelihood estimates

$$d\widetilde{\mu}_0(s;\beta) = \frac{d\bar{N}_\cdot(s)}{\sum_{i=1}^{m} Y_i(s) \exp(x_i'(s)\beta)}\,, \tag{3.20}$$

which may be substituted into (3.5) to give a $p \times 1$ system of equations $U_\beta(\beta) = 0$, where

$$U_\beta(\beta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s) x_i(s) \left[dN_i(s) - \frac{d\bar{N}_\cdot(s)}{\sum_{\ell=1}^m Y_\ell(s) \exp(x_\ell'(s)\beta)} \exp(x_i'(s)\beta) \right] .$$

This can be rewritten as

$$U_\beta(\beta) = \sum_{i=1}^{m} \int_0^\tau Y_i(s) W_i(s;\beta) dN_i(s) \,, \tag{3.21}$$

where

$$W_i(s;\beta) = x_i(s) - \frac{\sum_{l=1}^m Y_l(s) \exp(x_l'(s)\beta) x_l(s)}{\sum_{l=1}^m Y_l(s) \exp(x_l'(s)\beta)} \,. \tag{3.22}$$

The system of equations $U_\beta(\beta) = 0$ may be solved to obtain $\widehat{\beta}$ and the semiparametric estimate of $d\mu_0(s)$ is obtained by inserting $\widehat{\beta}$ into (3.20), giving

$$d\widehat{\mu}_0(s) = d\tilde{\mu}_0(s;\widehat{\beta}) = \frac{d\bar{N}_\cdot(s)}{\sum_{i=1}^m Y_i(s) \exp(x_i'(s)\widehat{\beta})} \,. \tag{3.23}$$

The resulting estimate of $\mu_0(t)$,

$$\widehat{\mu}_0(t) = \int_0^t d\widehat{\mu}_0(s) = \sum_{h:t_{(h)} \le t} \frac{d\bar{N}_\cdot(t_{(h)})}{\sum_{i=1}^m Y_i(t_{(h)}) \exp(x_i'(t_{(h)})\widehat{\beta})} \tag{3.24}$$

is referred to as the *generalized Nelson–Aalen estimate* because it is an estimate of the baseline mean function for a regression model and reduces to (3.17) when $\widehat{\beta} = 0$. The same proviso regarding the need to have $Y_\cdot(s) > 0$ for $0 \le s \le t$ as discussed following (3.17), applies to (3.24). If interest lies in estimating the mean function for an individual with covariate path $\{x_i(s), 0 \le s \le t\}$, then this is given by

$$\widehat{\mu}_i(t) = \int_0^t \exp(x_i'(s)\widehat{\beta}) d\widehat{\mu}_0(s) \,.$$

Partial Likelihood

One can derive (3.21) in a quite different manner by considering a particular factorization of the likelihood based on (3.3).

Let $t_{(1)} < t_{(2)} < \cdots < t_{(H)}$ denote the H unique ordered event times in the sample and let $\mathcal{F}_h$ denote the set of individuals with an event at $t_{(h)}$ (i.e. $\mathcal{F}_h = \{i | t_{ij} = t_{(h)}$ for some $j = 1, \ldots, n_i\}$). The number of individuals in $\mathcal{F}_h$ is denoted f_h, $h = 1, \ldots, H$. For continuous time processes f_h should in theory equal one, but ties in event times can occur because they are recorded with finite precision. As before we let $x^{(t)} = \{x_i(s), 0 \le s \le t, i = 1, \ldots, m\}$, $x^{(\infty)} = \{x_i(s), 0 \le s, i = 1, \ldots, m\}$, and define $y^{(t)} = \{Y_i(s), 0 \le s \le t, i = 1, \ldots, m\}$. For simplicity, we suppose $y^{(\infty)}$ is independent of the event process, and write the probability of the event data conditional on $x^{(\infty)}$ and $y^{(\infty)}$ as

$$\prod_{h=1}^{H}\left[\Pr(d\bar{N}_1(t_{(h)}),\ldots d\bar{N}_m(t_{(h)})|d\bar{N}_\cdot(t_{(h)}),x^{(\infty)},y^{(t_{(h)})})\right]\times$$
$$\prod_{h=1}^{H}\left[\Pr(d\bar{N}_\cdot(t_{(h)})|x^{(\infty)},y^{(t_{(h)})})\right],$$

where as before $d\bar{N}_i(t) = Y_i(t)dN_i(t)$ and $d\bar{N}_\cdot(t) = \sum_{i=1}^{m} Y_i(t)dN_i(t)$. The likelihood contributions from the first set of terms give a partial likelihood

$$L_1(\beta) = \prod_{h=1}^{H}\left\{\frac{\exp(\sum_{i\in\mathcal{F}_h} x_i'(t_{(h)})\beta)}{\left[\sum_{\ell=1}^{m} Y_\ell(t_{(h)})\exp(x_\ell'(t_{(h)})\beta)\right]^{f_h}}\right\} \tag{3.25}$$

for β, which may be used to estimate β directly. Differentiating $\log L_1(\beta)$ with respect to β leads to a partial likelihood score equation which is in fact identical to (3.21). The partial likelihood (3.25) is also valid under conditionally independent $Y_i(t)$ processes; see Andersen et al. (1993; Section 7.2).

Variance Estimation

Because $U_\beta(\beta)$ in (3.21) is a partial likelihood score function and it has zero expectation, one can use martingale-based partial likelihood theory (see Andersen and Gill, 1982; Andersen et al., 1993, Ch. 7) to show that conditional on the covariate processes $x^{(\infty)}$, $\text{asvar}(U_\beta(\beta)) = \mathcal{I}_{\beta\beta}(\beta) = E\{U_\beta(\beta)U_\beta'(\beta)\} = E\{-\partial U_\beta(\beta)/\partial\beta'\}$. Based on $\mathcal{I}_{\beta\beta}(\beta) = E\{U_\beta(\beta)U_\beta'(\beta)\}$ we obtain, when the $Y(s)$ processes are independent of the event processes,

$$\begin{aligned}\mathcal{I}_{\beta\beta}(\beta) &= E\left\{\sum_{i=1}^{m}\int_0^\tau Y_i(s)W_i(s;\beta)dN_i(s)\times\sum_{i=1}^{m}\int_0^\tau Y_i(t)W_i'(t;\beta)dN_i(t)\right\}\\ &= \sum_{i=1}^{m}\text{cov}\left\{\int_0^\tau Y_i(s)W_i(s;\beta)dN_i(s)\,,\int_0^\tau Y_i(t)W_i'(t;\beta)dN_i(t)\right\}\\ &= \sum_{i=1}^{m}\int_0^\tau\int_0^\tau Y_i(s)Y_i(t)W_i(s;\beta)W_i'(t;\beta)\text{cov}\{dN_i(s),dN_i(t)\}\,,\end{aligned}$$

because different subjects have independent event processes. Due to the independent increments property of Poisson processes, this can be further simplified to

$$\begin{aligned}\mathcal{I}_{\beta\beta}(\beta) &= \sum_{i=1}^{m}\int_0^\tau Y_i(s)W_i(s;\beta)W_i'(s;\beta)\text{var}\{dN_i(s)\}\\ &= \sum_{i=1}^{m}\int_0^\tau Y_i(s)W_i(s;\beta)W_i'(s;\beta)d\mu_i(s)\,.\end{aligned} \tag{3.26}$$

Estimates are obtained by replacing β and $d\mu_0(t) = \rho_0(t)dt$ by their estimates. Provided $m^{-1}\mathcal{I}_{\beta\beta}(\widehat{\beta})$ converges to a positive definite limit as $m\to\infty$,

then asymptotically we can act as though $U_\beta(\beta) \sim MVN(0, \mathcal{I}_{\beta\beta}(\widehat{\beta}))$ and $(\widehat{\beta} - \beta) \sim MVN(0, \mathcal{I}^{-1}_{\beta\beta}(\widehat{\beta}))$.

Software for the Cox survival model has been adapted to also deal with the Andersen–Gill model for recurrent events. Covariance matrix estimates based on $\mathcal{I}_{\beta\beta}(\widehat{\beta})$ are available from the `coxph` function in S-PLUS or R (see Section 3.8), in particular. The software also provides estimates $\widehat{\mu}_0(t)$ as well as variance estimates for $\widehat{\mu}_0(t)$ or for $\widehat{\mu}(t|x) = \widehat{\mu}_0(t)\exp(x'\widehat{\beta})$. Illustrations are provided in Section 3.8 and Appendix C. The method of handling "ties" implicit in (3.21) and (3.25) is referred to as the "Breslow" method in S-PLUS and R.

3.4.3 Stratification

In some settings, the population individuals are sampled from is comprised of subpopulations of individuals with different intensity or rate functions. Stratification is a convenient method of accommodating differences in rate functions between such subpopulations. Although multiplicative covariate effects can also do this, stratification by subpopulation introduces population-specific baseline intensity or rate functions which do not hinge on such multiplicative effects. If strata have different baseline rate functions and regression coefficients, then analysis simply amounts to a separate treatment of each stratum. Stratification more commonly refers to the case where the baseline rate functions vary, but the same regression coefficients apply across strata, and we consider this situation here. There is typically some loss of efficiency for estimating β if one stratifies unnecessarily but with strata of moderate size this loss tends to be modest.

Let r index strata, with $r = 1, 2, \ldots, R$ and let $N_{ri}(t)$ count events for individual i in stratum r ($i = 1, \ldots, m_r$). Let $\lambda_{ri}(t|H_{ri}(t))$ denote the corresponding event intensity function, where $H_{ri}(t) = \{N_{ri}(u) : 0 \le u < t\}$ and let τ_{ri} denote the censoring time, $Y_{ri}(t) = I(t \le \tau_{ri})$ the at-risk indicator, and $x_{ri}(t)$ the covariate vector. The basic stratified Poisson model takes

$$\lambda_{ri}\left(t|H_{ri}(t)\right) = \rho_{r0}(t)\exp(x'_{ri}(t)\beta)\,,$$

where $\rho_{r0}(t)$ is the baseline rate function for stratum r and β is the vector of regression coefficients assumed to be common across strata.

The likelihood contribution from individual i in the rth stratum is a function of (α^r, β) where α^r indexes $\rho_{r0}(t)$, and is still of the form given in (3.3),

$$L_{ri}(\alpha^r, \beta) = \prod_{j=1}^{n_{ri}} \rho_{r0}(t_{rij})\exp(x'_{ri}(t_{rij})\beta) \times \exp(-\int_0^\tau Y_{ri}(s)\rho_{r0}(s)\exp(x'_{ri}(s)\beta)ds)\,, \qquad (3.27)$$

where $\tau = \max(\tau_{ri})$ and t_{rij}, $j = 1, \ldots, n_{ri}$ are the times of the $N_{ri}(\tau_{ri}) = n_{ri}$ events for individual i in stratum r. The likelihood based on data for stratum r is

$$L_r(\alpha^r, \beta) = \prod_{i=1}^{m_r} L_{ri}(\alpha^r, \beta)$$

and the full likelihood for $\alpha = (\alpha^1, \ldots, \alpha^R)'$ and β is $L(\alpha, \beta) = \prod_{r=1}^{R} L(\alpha^r, \beta)$. With parametric models, $L(\alpha, \beta)$ is maximized as described in Section 3.2.

With semiparametric models, one obtains stratum-specific partial likelihoods for β of the form (3.25). The overall partial likelihood then is a product of these, and can be maximized in the usual fashion. The corresponding score equations are given by

$$U_\beta(\beta) = \sum_{r=1}^{R} \sum_{i=1}^{m_r} \int_0^\tau Y_{ri}(s) W_{ri}(s; \beta) dN_{ri}(s)\,,$$

and

$$W_{ri}(s; \beta) = x_{ri}(s) - \frac{\sum_{l=1}^{m_r} Y_{rl}(s) \exp(x'_{rl}(s)\beta) x_{rl}(s)}{\sum_{l=1}^{m_r} Y_{rl}(s) \exp(x'_{rl}(s)\beta)}\,.$$

The information matrix for β is

$$\mathcal{I}_{\beta\beta}(\beta) = \sum_{r=1}^{R} \mathcal{I}^{(r)}_{\beta\beta}(\beta)\,,$$

where $\mathcal{I}^{(r)}_{\beta\beta}(\beta)$ is given by (3.26) but evaluated using only data from the rth stratum. An estimate of the baseline rate function for stratum r, following (3.23), is

$$d\widehat{\mu}_{r0}(s) = \frac{d\bar{N}_{r\cdot}(s)}{\sum_{i=1}^{m_r} Y_{ri}(s) \exp(x'_{ri}(s)\widehat{\beta})}\,,$$

where $d\bar{N}_{r\cdot}(s) = \sum_{i=1}^{m_r} Y_{ri}(s) dN_{ri}(s)$, $r = 1, 2, \ldots, R$. As in (3.24) we obtain the estimate of the cumulative baseline mean function for stratum r as

$$\widehat{\mu}_{r0}(t) = \int_0^t d\widehat{\mu}_{r0}(s)\,.$$

Plots of the $\widehat{\mu}_{r0}(t)$ help indicate when the $\rho_{r0}(t)$ are truly different. Tests of the hypothesis $H_0 : \rho_{r0}(t) = \rho_0(t)$, $r = 1, \ldots, R$, can also be developed. One approach is to consider the model

$$\rho_{ri}(t) = \rho_0(t) \exp(\gamma_r + x'_{ri}(t)\beta)\,,$$

where $\gamma_1 = 0$ and $\gamma_2, \ldots, \gamma_R$ are arbitrary. Testing that $\gamma_2 = \cdots = \gamma_R = 0$ provides a test of H_0. Sun and Yang (2000) provide some other tests that do not constrain the $\rho_{r0}(t)$ to be proportional to each other; parametric modeling can also achieve this.

Stratified models are handled by the S-PLUS and R function `coxph` as well as other Cox model software. In some settings it is useful to allow one or more covariate effects to be different for different strata. It is possible to fit "covariate by stratum" interactions easily in S-PLUS and R as indicated in Appendix C. When there are covariate by stratum interactions for all covariates, estimates and inferences are identical to those resulting from separate analyses of each stratum, as noted previously.

In addition to serving as a useful modeling strategy for strata defined by subpopulations, stratification can serve as a method for defining more general intensity-based models as we discuss in Chapter 5.

3.4.4 Additive Models

A class of semiparametric additive regression models has also received a good deal of attention in the literature. Originally proposed by Aalen (1980), it takes the rate function for an individual with covariate vector $x_i(t)$ to be of the form

$$\rho_i(t) = x_i'(t)\beta(t)\,,$$

where $x_i(t) = (1, x_{i1}(t), \ldots, x_{i,p-1}(t))'$ and $\beta(t) = (\beta_0(t), \ldots, \beta_{p-1}(t))'$. This model allows the regression coefficients $\beta(t)$ to be time-dependent, but if desired some may be assumed constant, with $\beta_j(t) = \beta_j$. The function $\beta_0(t)$ acts as a baseline rate function, corresponding to an individual with $x_{i1}(t) = \cdots = x_{i,p-1}(t) = 0$.

Aalen (1980) and others have developed least squares estimation of the integrated coefficients

$$B_j(t) = \int_0^t \beta_j(s)ds \qquad j = 1, \ldots, p\,.$$

The essential idea is to define the $m \times p$ matrix $X(t)$ with rows $Y_i(t)x_i'(t)$, $i = 1, \ldots, m$ and to note that $E\{dN(t)|H(t)\} = X(t)\beta(t)dt$, where $dN(t) = (dN_1(t), \ldots, dN_m(t))'$ and $H(t)$ represents the history of events, covariates, and at-risk variables at time t. Writing $dB(t) = \beta(t)dt$, this suggests the (weighted) least squares estimator

$$d\widehat{B}(t) = \{X'(t)W(t)X(t)\}^{-1} X'(t)W(t)dN(t)\,,$$

where $W(t)$ is a left-continuous diagonal $m \times m$ weight matrix. For simplicity, we assume that $X(t)$ is of full rank p for all $t \leq \tau$, where $\tau = \max(\tau_i)$ is the maximum followup time.

Variance estimates for

$$\widehat{B}(t) = \int_0^t d\widehat{B}(s) = \int_0^t \{X'(s)W(s)X(s)\}^{-1} X'(s)W(s)dN(s)$$

can be developed. It should be noted that $\widehat{B}(t)$ is a step function which may increase or decrease. It gives information about the shape of the $\beta_j(t)$ and, in particular, whether they are fixed or time-varying. Smooth estimates of $\beta(t)$ can be obtained from $\widehat{B}(t)$ by using the kernel estimation approach described in Section 3.4.2.

Martinussen and Scheike (2006, Ch. 5) provide a thorough discussion of semiparametric additive models, and have created an R package, `timereg`, which implements the methodology.

3.5 Poisson Models with Random Effects

3.5.1 Formulation

In Section 2.2.3 we discussed the idea of introducing random effects into Poisson models to accommodate heterogeneity across individuals. Here we reconsider the useful mixed Poisson model, in which the conditional "subject-specific" intensity function is of the form

$$\lambda_i(t|H_i(t), u_i) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N_i(t) = 1|H_i(t), u_i\}}{\Delta t} = u_i \rho_i(t)\,, \qquad (3.28)$$

where the u_i are unobservable independent random effects. Given u_i and covariates, $\{N_i(t), 0 \le t\}$ is a Poisson process with rate $u_i\rho_i(t)$. As in Section 2.2.3, u_i is a nonnegative random variable independent of covariates, with a distribution $G(u;\phi)$, having $E(u) = 1$ and $\text{var}(u) = \phi$.

Such models are called mixed models because they contain both random terms (i.e. u_i) and fixed parameters (i.e. α, β). Upon marginalizing over the random effect, such mixed Poisson models give $E\{N_i(t)\} = \mu_i(t)$ and $\text{var}\{N_i(t)\} = \mu_i(t) + \mu_i^2(t)\phi$ as in (2.31), and so accommodate extra-Poisson variation. Moreover if $s_1 < t_1 < s_2 < t_2$,

$$\text{cov}\{N_i(s_1, t_1),\ N_i(s_2, t_2)\} = \phi\mu_i(s_1, t_1)\mu_i(s_2, t_2)\,,$$

as in (2.32), so the assumption of independent counts over disjoint intervals does not hold marginally for mixed Poisson processes. The parameter ϕ therefore determines both the degree of extra-Poisson variation and the degree of association between counts over disjoint intervals.

Although the event process for subject i is Poisson conditional on u_i, unconditionally it is not and the full intensity function has the form

$$\lambda_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N_i(t) = 1|H_i(t)\}}{\Delta t} = \rho_i(t) E\{u_i|H_i(t)\}\,, \qquad (3.29)$$

of which (2.33) is a special case. In Problem 3.9 it is noted that $E\{u_i|H_i(t)\} = E\{u_i|N_i(t^-)\}$ under independent censoring so the full intensity function is

simply $\rho_i(t)$ multiplied by the conditional expectation of the random effect given the total number of events observed over $[0, t)$.

If u_i were observed, conditional on the realization of the independent observation process, the probability of the data $(n_i, t_{i1}, \ldots, t_{in_i}, u_i)$ for subject i would be

$$\prod_{j=1}^{n_i}(u_i\rho_i(t_{ij}))\ \exp\left\{-\int_0^\infty Y_i(s)u_i\rho_i(s)ds\right\}dG(u_i;\phi)\,. \tag{3.30}$$

Because u_i, $i = 1, \ldots, m$ are unobserved, we base inferences on the likelihood which is proportional to the marginal probability of the observable quantities $(n_i, t_{i1}, \ldots, t_{in_i})$. For individual i, $L_i(\theta, \phi)$ is

$$\int_0^\infty \left[\prod_{j=1}^{n_i} u_i\rho_i(t_{ij}) \exp\left\{-\int_0^\infty Y_i(s)u_i\rho_i(s)ds\right\}\right] dG(u_i;\phi)\,. \tag{3.31}$$

Criteria for the choice of the distribution $G(u;\phi)$ include tractability of the integral in (3.31), properties of the full intensity function, and availability of software.

When considering possible random effect distributions for (3.31) it is helpful to recall the definition of the Laplace transform of a nonnegative random variable U with distribution $G(u)$. It is defined as

$$\mathcal{L}(s) = \int_0^\infty \exp(-us)dG(u)\,,$$

which has obvious connections with the associated moment generating functions. Laplace transforms provide a convenient basis for handling moments of the random variable, sums of independent random variables, as well as likelihood functions for frailty models in survival analysis or mixed Poisson processes. To see this note that

$$\mathcal{L}^{(r)}(s) = \frac{\partial^r \mathcal{L}(s)}{\partial s^r} = \int_0^\infty (-1)^r u^r e^{-us} dG(u)\,.$$

Therefore we may write (3.31) as

$$L_i(\theta, \phi) = (-1)^{n_i} \prod_{j=1}^{n_i} \rho_i(t_{ij})\mathcal{L}^{(n_i)}(\mu_i(\tau_i);\phi)\,.$$

This implies that any distribution for nonnegative random variables with a closed-form Laplace transform leads to relatively tractable marginal likelihoods. These include the gamma, inverse Gaussian, and positive stable distributions, all of which are members of a broader family of distributions called the power variance function distributions (see Hougaard, 2000, Ch. 9 and App. A). In Sections 2.2.3 and 3.5.3 we consider the case where the random effects follow a gamma distribution.

3.5.2 Models for Zero-Inflated Data

In many settings, count or recurrent event data exhibit patterns which suggest the population is comprised of distinct subpopulations whose differences cannot be explained by available covariates. For example, datasets sometimes include more individuals with no events than would be expected from a proposed model; the data are then said to be *zero-inflated*. The zero-inflated Poisson model is a widely studied model and is defined as follows. Suppose $\{N_i(t), 0 \le t\}$ is a counting process for subject i and W_i is a latent (unobserved) random variable with $\Pr(W_i = 1) = \pi_i$ and $\Pr(W_i = 0) = 1 - \pi_i$, $i = 1, \ldots, m$. If $\{N_i(t), 0 \le t\}|W_i = 1$ is a Poisson process with rate $\rho_i(t;\theta) = \rho_0(t;\alpha)\exp(x_i'\beta)$, and $\Pr(N_i(\infty) = 0|W_i = 0) = 1$, then the marginal distribution is a mixed Poisson process which accommodates an excess number of zeros. Under this model $E\{N_i(t)\} = \mu_i(t)\pi_i$ and $\text{var}\{N_i(t)\} = \mu_i(t)\pi_i + \mu_i^2(t)\pi_i(1-\pi_i)$, where $\mu_i(t) = \int_0^t \rho_i(s)ds$.

The likelihood contribution from individual i is of the form

$$L_i(\theta,\gamma) \propto \left[\prod_{j=1}^{n_i} \rho_i(t_{ij}) \exp\left\{-\int_0^\infty Y_i(s)\rho_i(s)ds\right\}\pi_i\right]^{I(n_i>0)} \times \left[\exp\left\{-\int_0^\infty Y_i(s)\rho_i(s)ds\right\}\pi_i + (1-\pi_i)\right]^{I(n_i=0)}.$$

The form of this likelihood arises because $n_i > 0$ means W_i must be one, but zero counts may arise for two reasons: W_i could be one and a zero count was observed from the Poisson distribution by chance, or $W_i = 0$ in which case a zero occurs with probability one. We typically use a model for π_i where $g(\pi_i) = z_i'\gamma$ and $g(\cdot)$ is a common link function for binomial data.

For models with parametric baseline rates, maximization of the likelihood above can be carried out using general-purpose optimization software; see Appendix B. For semiparametric baseline rate functions, the EM algorithm offers a convenient approach. The steps are analogous to those given in Section 3.5.3 for the gamma–Poisson mixture, but with modified expressions for the E-step. Likelihoods can be slightly more challenging to fit with zero-inflated Poisson models because the likelihood function may be rather flat in certain regions. This will occur, for example, if the numbers of events observed for individuals with $W_i = 1$ are small. Moreover, care must be exercised in selecting the covariates to place in the linear predictor for the Poisson component and the binary component. Ideally, the scientific context would suggest which model a covariate should enter because there may be difficulties in estimation if covariates are placed in both components.

A more general model that accommodates zero-inflation and a continuous frailty is based on the compound Poisson model, which is formulated as follows. Let K_i be a Poisson random variable with mean κ, and let $V_1, V_2, \ldots,$

denote independent and identically distributed gamma random variables with shape and inverse scale parameters γ_1 and γ_2, respectively, and density

$$f(v;\gamma) = \gamma_2^{\gamma_1} v^{\gamma_1 - 1} \exp(-v\gamma_2)/\Gamma(\gamma_1)\,, \qquad v > 0\,.$$

Then if we define $U_i = V_1 + V_2 + \cdots + V_{K_i}$ if $K_i > 0$ and $U_i = 0$ if $K_i = 0$, U_i has a compound Poisson distribution. Note that because $\Pr(K_i = 0) = \exp(-\kappa)$ this model accommodates zero-inflation through κ, since $\Pr(U_i = 0) = \Pr(K_i = 0)$. The density for the continuous part of U_i is

$$f(u_i;\kappa,\gamma) = u_i^{-1} \exp(-(\kappa + u_i\gamma_2)) \sum_{k=1}^{\infty} \frac{(u_i\gamma_2)^{k\gamma_1}\kappa^k}{\Gamma(k\gamma_1)k!}\,, \qquad u_i > 0\,.$$

Although this density is rather complicated, the Laplace transform of U_i is relatively straightforward to evaluate and can be shown to equal

$$\mathcal{L}(s) = \exp\left\{-\kappa\left[1 - (1 + s/\gamma_2)^{-\gamma_1}\right]\right\}$$

and hence it is convenient to write marginal likelihoods and obtain the required derivatives (see Problem 3.10).

3.5.3 Negative Binomial Models

The gamma distribution (2.28) for u_i is the most common choice for (3.28), to a large degree because of the tractability of the integral in (3.31). In this case $(n_i, t_{i1}, \ldots, t_{in_i})$ arises from a negative binomial process for which, with fixed covariates, the full intensity function (3.29) is given by (2.33), which we write here as

$$\lambda_i(t|H_i(t)) = \rho_i(t)\left(\frac{1 + N_i(t^-)\phi}{1 + \mu_i(t)\phi}\right)\,.$$

Note that covariate effects which are expressed multiplicatively in (3.28), as in (3.2), are not multiplicative effects in the full intensity function. The likelihood function for α, β, and ϕ may be constructed from this intensity function using (2.7), or directly from (3.31). When $\rho_i(t)$ is of the form (3.2) the resulting likelihood contribution from individual i is

$$L_i(\theta,\phi) = \left\{\prod_{j=1}^{n_i} \frac{\rho_0(t_{ij})}{\mu_0(\tau_i)}\right\} \frac{\Gamma(n_i + \phi^{-1})}{\Gamma(\phi^{-1})} \frac{(\phi\mu_i(\tau_i))^{n_i}}{(1 + \phi\mu_i(\tau_i))^{n_i + \phi^{-1}}}\,, \tag{3.32}$$

where $\mu_0(t) = \int_0^t \rho_0(s)ds$. More generally, the likelihood is of the form (3.32) with $\rho_0(t_{ij})$ and $\mu_0(\tau_i)$ replaced by $\rho_i(t_{ij})$ and $\mu_i(\tau_i)$.

Parametric Models

Maximum likelihood estimation can proceed in a number of ways for parametric models of the form (3.2). If we define $\theta = (\alpha', \beta', \phi)'$ with $\rho_0(t)$ dependent on α, the log-likelihood from (3.32) is

$$\ell(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} [\log \rho_0(t_{ij}) - \log \mu_0(\tau_i)] + \sum_{j=0}^{n_i^*} \log(1+\phi j) \right. \qquad (3.33)$$
$$\left. + n_i \log \mu_i(\tau_i) - (n_i + \phi^{-1}) \log(1+\phi\mu_i(\tau_i)) \right\} ,$$

where $n_i^* = \max(0, n_i - 1)$. When covariates are fixed and (3.2) is of the form $\rho_0(t;\alpha)\exp(x_i'\beta)$, the elements of the score vector are

$$U_\alpha(\theta) = \frac{\partial \ell(\theta)}{\partial \alpha} = \sum_{i=1}^{m} \left\{ \left[\sum_{j=1}^{n_i} \frac{\partial \rho_0(t_{ij})/\partial \alpha}{\rho_0(t_{ij})} \right] - \frac{(1+\phi n_i)\exp(x_i'\beta)}{1+\phi\mu_i(\tau_i)} \frac{\partial \mu_0(\tau_i)}{\partial \alpha} \right\}$$

$$U_\beta(\theta) = \frac{\partial \ell(\theta)}{\partial \beta} = \sum_{i=1}^{m} \frac{n_i - \mu_i(\tau_i)}{1+\phi\mu_i(\tau_i)} x_i$$

$$U_\phi(\theta) = \frac{\partial \ell(\theta)}{\partial \phi} = \sum_{i=1}^{m} \left[\sum_{j=0}^{n_i^*} \frac{j}{1+\phi j} - \frac{\mu_i(\tau_i)(n_i + \phi^{-1})}{1+\mu_i(\tau_i)\phi} + \frac{\log(1+\phi\mu_i(\tau_i))}{\phi^2} \right]$$

and the information matrix is

$$I(\theta) = \begin{pmatrix} -\dfrac{\partial U_\alpha(\theta)}{\partial \alpha'} & -\dfrac{\partial U_\alpha(\theta)}{\partial \beta'} & -\dfrac{\partial U_\alpha(\theta)}{\partial \phi} \\ -\dfrac{\partial U_\beta(\theta)}{\partial \alpha'} & -\dfrac{\partial U_\beta(\theta)}{\partial \beta'} & -\dfrac{\partial U_\beta(\theta)}{\partial \phi} \\ -\dfrac{\partial U_\phi(\theta)}{\partial \alpha'} & -\dfrac{\partial U_\phi(\theta)}{\partial \beta'} & -\dfrac{\partial U_\phi(\theta)}{\partial \phi} \end{pmatrix} .$$

Good general-purpose optimization software will readily maximize $\ell(\theta)$ and also compute $I(\widehat{\theta})$. One may also proceed by direct Newton–Raphson for θ using the observed information matrix. It can be shown that $E\{\partial U_\alpha/\partial\phi\} = E\{\partial U_\beta/\partial\phi\} = 0$ and so a Fisher-scoring algorithm based on the expected information instead of $I(\theta)$ can make use of the orthogonality of (α, β) and ϕ. If covariates are time-varying, so that

$$\mu_i(t) = \int_0^t \rho_0(s)\exp(x_i'(s)\beta)ds ,$$

it is generally necessary to use numerical integration, and the guidelines in Section 3.2 apply.

Semiparametric Models

Semiparametric mixed Poisson models arise when $\rho_i(t) = \rho_0(t)\exp(x_i'(t)\beta)$ in (3.28) and $\rho_0(t)$ is completely unspecified. Direct maximization of the resulting likelihood is difficult so here we provide a brief sketch of an EM algorithm (Dempster et al., 1977) which facilitates estimation in this setting.

In this case we take the "complete data" likelihood for individual i as (3.30), which would apply if the random effects were observed. With the gamma distribution for u_i the overall complete data likelihood $L_C(\theta, \phi)$ is given by

$$\prod_{i=1}^{m}\left\{\left[\prod_{j=1}^{n_i}\frac{\rho_0(t_{ij})}{\mu_0(\tau_i)}\right](u_i\mu_i(\tau_i))^{n_i}\exp(-u_i\mu_i(\tau_i))\times\frac{u_i^{\phi^{-1}-1}\exp(-u_i/\phi)}{\Gamma(\phi^{-1})\phi^{\phi^{-1}}}\right\}.$$

The corresponding complete data log-likelihood, considering for simplicity the case where covariates are fixed, is of the form

$$\ell_C(\theta) = \ell_1(\theta) + \ell_2(\theta) + \ell_3(\theta)\,,$$

where

$$\ell_1(\theta) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}[\log\rho_0(t_{ij}) - \log\mu_0(\tau_i)]$$

$$\ell_2(\theta) = \sum_{i=1}^{m}[n_i(\log u_i + \log\mu_0(\tau_i) + x_i'\beta) - u_i\mu_0(\tau_i)\exp(x_i'\beta)]$$

$$\ell_3(\theta) = \sum_{i=1}^{m}\left[(\phi^{-1}-1)\log u_i - u_i/\phi - \log\Gamma(\phi^{-1}) - \phi^{-1}\log\phi\right].$$

The E-step at the kth iteration in the process for obtaining the maximum likelihood estimates involves taking the expectation of the complete data log-likelihood with respect to u_i, but based on the conditional distribution of u_i given $H_i(\tau)$, and evaluated at $\widehat{\theta}^{(k-1)}$, the parameter estimates from the previous iteration. The use of the gamma random effect distribution is particularly appealing because $G(u_i|H_i(\tau);\theta)$ is also gamma, with shape $\phi^{-1} + N_i(\tau_i)$ and scale $\phi/(1+\phi\mu_i(\tau_i))$ and so the required expectations have closed form. Specifically

$$E\{u_i|H_i(\tau_i);\widehat{\theta}^{(k-1)}\} = \frac{1+n_i\widehat{\phi}^{(k-1)}}{1+\widehat{\mu}_i^{(k-1)}(\tau_i)\widehat{\phi}^{(k-1)}}$$

and

$$E\{\log u_i|H_i(\tau_i);\widehat{\theta}^{(k-1)}\} = \Psi(1/\widehat{\phi}^{(k-1)} + n_i) - \log(1/\widehat{\phi}^{(k-1)} + \widehat{\mu}_i^{(k-1)}(\tau_i))\,,$$

where $\Psi(\cdot)$ is the digamma function.

Let $Q_r(\theta;\widehat{\theta}^{(k-1)}) = E\{\ell_r(\theta)|H_i(\tau_i);\widehat{\theta}^{(k-1)}\}$, $r = 1,2,3$, and note that at the M-step, maximizing $Q_1(\theta;\widehat{\theta}^{(k-1)}) + Q_2(\theta;\widehat{\theta}^{(k-1)})$ is equivalent to maximizing the likelihood under a semiparametric specification with no random effects (see Section 3.4.2) with an offset $\log\widetilde{\gamma}_i^{(k-1)}$, where $\widetilde{\gamma}_i^{(k-1)} = E\{u_i|H_i(\tau_i);\widehat{\theta}^{(k-1)}\}$. The term $Q_3(\theta;\widehat{\theta}^{(k-1)})$ may easily be maximized using general-purpose optimization software. The iteration continues until the difference in estimates at successive iterations drops below a desired tolerance. Interval estimation and testing is most conveniently carried out by using likelihood ratio statistics (Appendix A).

The above algorithm is essentially the one given by Klein (1992) and Nielsen et al. (1992) for fitting semiparametric frailty models for clustered survival data. The `coxph` function in S-PLUS and R accommodates random effects by the `frailty(id)` option, as illustrated in Section 3.8; however, it uses a different algorithm to obtain estimates and variance estimates. An alternative approach that is convenient and extends to other random effects models is to use (3.31) and (3.33) with a piecewise-constant baseline rate function, as in Section 3.3. This makes the model parametric but by letting the number of pieces in the baseline rate become large, we obtain estimates close to the semiparametric estimates.

3.6 Robust Methods for Rate and Mean Functions

3.6.1 Nonparametric Estimation

Full specification of a model for recurrent events via intensity functions is often desirable, particularly when interest lies in extrapolations or predictions, or simply when it is desired to have a comprehensive understanding or description of an event process. However, when comparing groups of individuals, or when assessing the effects of fixed covariates, methods that focus on marginal features such as the rate or mean function are often sufficient. An advantage of restricting attention to such marginal features is that it is often possible to relax model assumptions and therefore achieve greater robustness for inferences. For example, in some applications there are only a few individuals with more than one or two events. In this case there is naturally less information about the event processes and it may be difficult to postulate models and carry out diagnostic tests for model fit. However, it may be possible and sufficient simply to assess the effects of covariates on mean functions. This approach has the advantage of easy interpretability, and we show here that simple robust methods can be developed by building upon methods for Poisson processes. An important point, however, is that this methodology requires the observation process $\{Y_i(t), 0 \le t\}$ and the event processes to be independent.

To see this, note that provided $\{Y_i(t), 0 \le t\}$ and $\{dN_i(t), 0 \le t\}$ are independent, then in the absence of covariates

$$E[Y_i(s)\{dN_i(s) - d\mu(s)\}] = 0\,,$$

and hence the solution to (3.16) is valid for $d\mu(s)$ regardless of the form of the underlying event processes. More specifically, the estimate of the mean function (3.17) is unbiased because $E\{d\widehat{\mu}(s)|Y_1(s),\ldots,Y_m(s)\} = d\mu(s)$ regardless of the underlying process. The estimates $d\widehat{\mu}(s)$ and (3.17) are Poisson maximum likelihood estimates, but are also therefore valid quite generally. In order to make use of this result, we require a variance estimate that is also valid generally.

A robust variance estimate for $\widehat{\mu}(t)$ is obtained by noting that

$$\begin{aligned}\mathrm{var}\{\sqrt{m}(\widehat{\mu}(t) - \mu(t))\} &= m \cdot \mathrm{var}\left\{\int_0^t \frac{d\bar{N}_\cdot(u)}{Y_\cdot(u)}\right\} \\ &= m \cdot \sum_{i=1}^m \int_0^t \int_0^t \frac{Y_i(u)}{Y_\cdot(u)} \frac{Y_i(v)}{Y_\cdot(v)}\, \mathrm{cov}\{dN_i(u), dN_i(v)\}\,.\end{aligned}$$

It can be shown that

$$m \sum_{i=1}^m \int_0^t \int_0^t \frac{Y_i(u)}{Y_\cdot(u)} \frac{Y_i(v)}{Y_\cdot(v)}\, [dN_i(u) - d\widehat{\mu}(u)]\, [dN_i(v) - d\widehat{\mu}(v)]$$

is a consistent estimate for this variance provided $Y_\cdot(u)/m \to p(u) > 0$ as $m \to \infty$ for all u in $[0,t]$ (Lin et al., 2000). This can be rewritten as

$$\widehat{\mathrm{var}}\{\sqrt{m}(\widehat{\mu}(t) - \mu(t))\} = m \sum_{i=1}^m \left\{\int_0^t \frac{Y_i(u)}{Y_\cdot(u)} \left[dN_i(u) - \frac{d\bar{N}_\cdot(u)}{Y_\cdot(u)}\right]\right\}^2 . \quad (3.34)$$

Note that when $\tau_i \geq t$ for all $i = 1,\ldots,m$, the mean function estimate $\widehat{\mu}(t)$ reduces to the sample mean $\bar{N}_\cdot(t)/m$ and (3.34) becomes

$$\widehat{\mathrm{var}}\{\sqrt{m}(\widehat{\mu}(t) - \mu(t))\} = \frac{1}{m} \sum_{i=1}^m \{N_i(t) - \bar{N}_\cdot(t)/m\}^2\,,$$

which is recognizable as the usual robust sample variance estimate. Finally, we remark that as pointed out in Section 3.4.1, it is frequently preferable to construct confidence intervals on the log scale and then exponentiate the limits. The delta method can be used to obtain the corresponding expressions based on the robust variance estimates.

3.6.2 Parametric Estimation

The robust methods of Section 3.6.1 also apply to fully parametric models that may include covariates. Suppose the rate function is of the form

$$\rho_i(t;\alpha,\beta) = \rho_0(t;\alpha)\exp(x_i'(t)\beta)\,,$$

where α is a finite-dimensional parameter. Letting $\theta = (\alpha', \beta')'$ and writing $\rho_i(t;\theta)$ for $\rho_i(t;\alpha,\beta)$, the Poisson process likelihood score equations are given by (3.4) and (3.5). It is obvious that when the at risk processes $\{Y_i(t), 0 \le t\}$ are independent of the event processes, the score function $U_\alpha(\theta)$ of (3.4) and $U_\beta(\theta)$ of (3.5) have zero expectation, provided only that the specification $\rho_i(t;\theta)$ is correct; a Poisson process is therefore not required.

It follows from standard large sample theory for estimating functions (Appendix A) that the Poisson maximum likelihood estimates $\widehat{\theta} = (\widehat{\alpha}', \widehat{\beta}')'$ obtained by solving $U_\alpha(\theta) = 0$, $U_\beta(\theta) = 0$ are, under mild conditions, asymptotically normal as $m \to \infty$. More specifically, if $U_i(\theta) = (U_{i\alpha}(\theta)', U_{i\beta}(\theta)')'$ is the likelihood score vector for the ith individual, $\sqrt{m}(\widehat{\theta} - \theta)$ is asymptotically normal with zero mean and a covariance matrix $V(\theta)$ that is estimated consistently by

$$\widehat{\text{asvar}}(\widehat{\theta}) = I(\widehat{\theta})^{-1} B(\widehat{\theta}) I(\widehat{\theta})^{-1}\,, \tag{3.35}$$

where $I(\theta) = -\partial U(\theta)/\partial\theta'$ and $B(\theta) = \sum_{i=1}^m U_i(\widehat{\theta}) U_i(\widehat{\theta})'$.

If Poisson process software exists for a specific rate function model $\rho_i(t;\theta)$, it can be used to obtain $\widehat{\theta}$ and $I(\widehat{\theta})$, which is the observed information matrix. The matrix $B(\widehat{\theta})$ can be obtained from the so-called score residual vectors $U_i(\widehat{\theta})$. More generally, $\widehat{\theta}$ can be obtained by maximizing the Poisson log-likelihood function $\ell(\theta)$ at the start of Section 3.2, using optimization software. This will also return $I(\widehat{\theta})$ so that only $B(\widehat{\theta})$ has to be computed separately.

3.6.3 Robust Semiparametric Methods

We now consider the semiparametric multiplicative model where the rate function for $\{N_i(t), 0 \le t\}$ given the external covariate vectors $x_i(t)$ is

$$E\{dN_i(t) | Y_i(t), x_i^{(\infty)}\} = \rho_0(t)dt \exp(x_i'(t)\beta)\,, \qquad t > 0\,.$$

Recall that the partial or profile score for β in (3.21) and the estimate of $\mu_0(t)$ in (3.24) were derived under the assumption that the events were generated according to a Poisson process. Provided the observation process is completely independent of the event process, however, (3.21) is more generally an unbiased estimating function for β. Therefore, any software for the maximization of a Cox partial likelihood function can be used to obtain $\widehat{\beta}$. Furthermore, under the assumption that the rate functions are proportional, (3.24) remains a consistent estimator of the baseline mean function and is obtainable from Cox model software. Variance estimation requires estimating function results, which we now describe.

The expression for $\text{var}\{U_\beta(\beta)\} = I_{\beta\beta}(\beta)$ that was obtained in Section 3.4 used the independent increments property of the Poisson model. More generally however, we may write $\text{var}\{\sqrt{m}^{-1} U_\beta(\beta)\}$ as

$$\frac{1}{m}\sum_{i=1}^m \int_0^\tau \int_0^\tau Y_i(u) Y_i(v) W_i(u;\beta) W_i'(v;\beta) \text{cov}\{dN_i(u), dN_i(v)\}\,, \tag{3.36}$$

and note that

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^\tau\int_0^\tau Y_i(u)Y_i(v)W_i(u;\widehat{\beta})W_i'(v;\widehat{\beta})d\widehat{M}_i(u)d\widehat{M}_i(v) \tag{3.37}$$

gives a consistent estimate of $\widehat{\text{asvar}}\{\sqrt{m}^{-1}U_\beta(\beta)\}$, where $d\widehat{M}_i(u) = dN_i(u) - d\widehat{\mu}_i(u)$ and $d\widehat{\mu}_i(u) = d\widehat{\mu}_0(u)\exp(x_i'(u)\widehat{\beta})$ (Lin et al., 2000). This can be rewritten as $m^{-1}\sum_{i=1}^m \widehat{B}_i\widehat{B}_i'$, where

$$\widehat{B}_i = \int_0^\tau Y_i(u)W_i(u;\widehat{\beta})d\widehat{M}_i(u)\,.$$

If β_0 denotes the true parameter value, by Taylor series expansion one obtains that

$$\sqrt{m}(\widehat{\beta}-\beta_0) = [-m^{-1}\partial U_\beta(\beta_0)/\partial\beta_0]^{-1}[\sqrt{m}^{-1}U_\beta(\beta_0)] + o_p(1)\,.$$

As $m\to\infty$, $\sqrt{m}^{-1}U_\beta(\beta_0)$ converges to $MVN(0,\mathcal{B}(\beta_0))$ in distribution, where $\mathcal{B}(\beta_0) = m^{-1}E\{U_\beta(\beta_0)U_\beta'(\beta_0)\}$. Moreover $-m^{-1}\partial U_\beta(\beta_0)/\partial\beta_0$ converges in probability to the $p\times p$ matrix $\mathcal{A}(\beta_0) = m^{-1}E\{-\partial U_\beta(\beta_0)/\partial\beta_0'\}$. Thus $\text{asvar}\{\sqrt{m}(\widehat{\beta}-\beta_0)\} = \mathcal{A}^{-1}(\beta_0)\mathcal{B}(\beta_0)[\mathcal{A}^{-1}(\beta_0)]'$. Empirical estimates of $\mathcal{A}(\beta_0)$ and $\mathcal{B}(\beta_0)$ are obtained by using empirical averages instead of expectations and replacing unknown quantities with their estimates. Therefore we obtain

$$\widehat{\text{asvar}}\{\sqrt{m}(\widehat{\beta}-\beta_0)\} = \widehat{A}^{-1}(\widehat{\beta})\widehat{B}(\widehat{\beta})[\widehat{A}^{-1}(\widehat{\beta})]'\,, \tag{3.38}$$

where $\widehat{B}(\widehat{\beta})$ is given by (3.37) and $\widehat{A}(\widehat{\beta}) = -m^{-1}\partial U_\beta(\widehat{\beta})/\partial\widehat{\beta}'$ is, from (3.21),

$$\sum_{i=1}^m\int_0^\tau Y_i(u)\left[\frac{S^{(2)}(\widehat{\beta};u)}{S^{(0)}(\widehat{\beta};u)} - \frac{S^{(1)}(\widehat{\beta};u)}{S^{(0)}(\widehat{\beta};u)}\left\{\frac{S^{(1)}(\widehat{\beta};u)}{S^{(0)}(\widehat{\beta};u)}\right\}'\right]dN_i(u)\,, \tag{3.39}$$

where

$$\begin{aligned}
S^{(0)}(\beta;u) &= \sum_{i=1}^m Y_i(u)\exp(x_i'(u)\beta)\,,\\
S^{(1)}(\beta;u) &= \sum_{i=1}^m Y_i(u)\exp(x_i'(u)\beta)x_i(u)\,, \quad \text{and}\\
S^{(2)}(\beta;u) &= \sum_{i=1}^m Y_i(u)\exp(x_i'(u)\beta)x_i(u)x_i'(u)\,.
\end{aligned}$$

Note that under Poisson likelihood or partial likelihood analyses, $\mathcal{A}(\beta_0) = \mathcal{B}(\beta_0)$. More generally, however, this does not hold and robust "sandwich type" variance estimates (3.38) are required for valid inference. The Poisson variance estimate used in Section 3.4.2 was $\mathcal{A}(\widehat{\beta})^{-1}$, but this may be seriously

biased for other processes. The estimate given here is valid under a Poisson model, but offers protection against departures from the Poisson process.

When covariates are fixed or piecewise-constant, the integrals in (3.37) and (3.39) are simply sums, and the covariance matrix estimate for $\widehat{\beta}$ is readily obtained. The joint asymptotic distribution of $\widehat{\beta}$ and $\widehat{\mu}_0(t)$ may also be obtained. This is useful, for example, where estimation of mean functions $\mu_0(t)\exp(x_0'\beta)$ is of interest. An alternative way to obtain a variance estimate for $\widehat{\mu}_0(t)\exp(x_0'\widehat{\beta})$ is to recenter covariates as $x_{\text{new}} = x - x_0$, so that the vector x_0 of interest becomes the zero vector. Then one merely needs the variance estimates for $\widehat{\mu}_0(t)$. The methods given here for robust estimation of $\mu_0(t)$ and β for the multiplicative regression model $\rho(t; x_i) = \rho_0(t)\exp(x_i'\beta)$ are available via the S-PLUS or R function `coxph`. Section 3.8 provides an illustration.

3.6.4 Robust Methods with Semiparametric Variances

As discussed in Section 3.6.3, consistent estimates are obtained for the parameters of the mean function in the semiparametric setting from the Poisson equations (3.19) and (3.21), and robust variance estimates ensure valid inferences if the form of the mean function is correctly specified. It is sometimes of interest, however, to formulate a parametric covariance structure to obtain insight into the nature of the association structure among the counts. One approach is to adopt the covariance structure implied by a mixed Poisson model given by (2.31) and (2.32). In this case, if $\mu_i(\tau_i)$ is estimated as in Section 3.4.2, a moment type estimate of ϕ is given by

$$\widehat{\phi} = \frac{\sum_{i=1}^{m}[(n_i - \widehat{\mu}_i(\tau_i))^2 - \widehat{\mu}_i(\tau_i)]}{\sum_{i=1}^{m}\widehat{\mu}_i^2(\tau_i)}\,. \tag{3.40}$$

A generally better estimator, with high efficiency under mixed Poisson processes with gamma random effects, is $\widehat{\phi}$ obtained as the solution to (Dean, 1991),

$$\sum_{i=1}^{m}\left\{\frac{(n_i - \widehat{\mu}_i)^2 - \widehat{\mu}_i(1+\phi\widehat{\mu}_i)}{(1+\phi\widehat{\mu}_i)^2}\right\} = 0\,,$$

where for simplicity we write $\widehat{\mu}_i$ for $\widehat{\mu}_i(\tau_i)$. In Section 6.4.2 we discuss this approach for multitype recurrent event data, where likelihoods for fully specified multivariate random effect models are generally intractable.

3.6.5 Methods Based on Multivariate Failure Time Data

Another marginal approach for analyzing recurrent events which has received considerable attention in the clinical trial arena is based on methods of analyzing multivariate failure time data developed by Wei et al. (1989). Here we make some brief remarks on the formulation of these models followed by comments on their utility for recurrent events.

Suppose each subject is at risk of K different types of clinical events which may represent, for example, different types of infections in immunology studies, different sites of metastases in cancer studies, and so on. Let T_k denote the time from randomization to the occurrence of the type-k event for subject i, $k = 1, \ldots, K$, observed over $[0, \tau_i]$, $i = 1, \ldots, m$. Marginal Cox regression models may be formulated for the failure times $T_1, \ldots, T_K$ with hazard functions

$$h_k(t|x_i) = h_{k0}(t) \exp(x_i' \beta_k) \qquad k = 1, \ldots, K\,, \tag{3.41}$$

where we adopt the same $p \times 1$ covariate vector for each event type but allow the baseline hazard functions and regression coefficients to differ. In the context of a parallel group clinical trial one would include a treatment indicator, denoted x_{i1}, in the covariate vector along with any other covariates that are considered important.

In the setting of multivariate failure time data there is no natural ordering of the event times and subjects are therefore considered "at risk" for all events from the time of randomization. Estimation of $\beta = (\beta_1', \ldots, \beta_K')'$ may be carried out under a working independence assumption, which means that the partial likelihood one would use if interest were only in the events of type k can be maximized to give $\widehat{\beta}_k$, and $\widehat{H}_{k0}(t)$ is estimated by a generalized Nelson–Aalen estimate similar to (3.24). Specifically, if $U_{ik} = \min(T_{ik}, \tau_i)$, $\delta_{ik} = I(T_{ik} \le \tau_i)$, and $Y_{ik}(s) = I(s \le U_{ik})$, then the partial likelihood

$$L_k(\beta_k) = \prod_{i=1}^{m} \left\{ \frac{\exp(x_i' \beta_k)}{\sum_{\ell=1}^{m} Y_{\ell k}(U_{ik}) \exp(x_\ell' \beta_k)} \right\}^{\delta_{ik}}$$

is maximized to give $\widehat{\beta}_k$. Robust estimates of the covariance matrix $\mathrm{var}(\sqrt{m}(\widehat{\beta} - \beta))$ are given by Wei et al. (1989) and may be used to construct global tests of treatment effect by taking optimally weighted linear combinations of the K estimated treatment coefficients.

There are several issues to consider when applying this approach to recurrent event data. First, in the recurrent event setting, the first, second, third, and subsequent event times are typically taken to be analogous to the failure times for the different event types in the multivariate setting. One must therefore specify the maximum number of events to be analyzed to correspond to the number of distinct event types. This may require discarding events occurring in subjects having more events than are to be analyzed. Second, subjects are considered "at risk" for their kth event even before experiencing their $(k-1)$st event. This is a natural consequence of adopting an approach based on marginal failure time models, but does not coincide with reality in applications to recurrent event data. Third, there is no plausible underlying recurrent event model for which the proportional hazards assumption would hold for each event time. Robust covariance matrix estimates provide some protection against model misspecification, and tests of the null hypothesis of no treatment effect remain valid, but it is difficult to interpret the event-specific

or global measures of treatment effect in this setting. The Andersen–Gill or rate function approach with robust variance estimation, as described in Section 3.6.3, does not have such problems and is a preferable approach for the marginal analysis of recurrent event data. Estimation and testing of treatment effects with recurrent event data are discussed further in Sections 3.7.5 and 8.4.

3.7 Some Useful Tests for Rate Functions

Tests for a time trend and tests that compare two or more groups of processes are often of interest. In addition, model assessment is critical to ensure that assumptions underlying analyses are plausible in light of the available data. In this section we describe some tests for trend, model assessment, and comparisons. The methods use expanded models which include the "null" model to be assessed. These expanded models can be fitted and the null model tested against them. We focus here on score or pseudo-score tests which allow us to test the null model without actually fitting the expanded model.

3.7.1 Tests for Trend

In many applications it is of interest to test for a trend in the rate of occurrence of events. For example, in process or equipment reliability, it is important to monitor whether the rate of failure is increasing over time in order to identify problems or plan maintenance. In studies of chronic disease such as asthma, one may be interested in testing whether exacerbation rates increase with the time since disease onset. For Poisson processes this amounts to testing that the process is time-homogeneous. A simple and convenient framework for developing tests of this sort is to consider expanded models which accommodate trend.

In the absence of covariates, for example, one could adopt a model of the form

$$\rho(t;\alpha) = \exp(\alpha_1 + \alpha^* t) \qquad t \geq 0\,,$$

where $\alpha = (\alpha_1, \alpha^*)'$. Suppose m independent individual processes are observed, with process i observed over the interval $[0, \tau_i]$, $i = 1, \ldots, m$. Then (3.3) leads to the log-likelihood function

$$\ell(\alpha) = n_{\cdot}\alpha_1 + \alpha^* \sum_{i=1}^{m}\sum_{j=1}^{n_i} t_{ij} - \frac{\exp(\alpha_1)}{\alpha^*}\sum_{i=1}^{m}(\exp(\alpha^*\tau_i) - 1)\,,$$

where $n_{\cdot} = \sum_{i=1}^{m} n_i$. The maximum likelihood estimate $\widehat{\alpha}$ may be found by solving the equations $\partial\ell/\partial\alpha = 0$ or by maximization of $\ell(\alpha)$ some other way.

In the context of this model, the trend test is a test of the null hypothesis $H_0 : \alpha^* = 0$. This can be carried out by using a Wald, likelihood ratio, or score

test, but in some situations it is possible to develop a simple test based on a conditional likelihood for α^*. In particular, if the τ_i are completely independent of the event occurrences then $n_\cdot = \sum_{i=1}^m n_i$ has a Poisson distribution with mean $\sum_{i=1}^m (e^{\alpha^* \tau_i} - 1) e^{\alpha_1}/\alpha^*$, and an easy calculation shows that the distribution of the full data $\{(n_i, t_{i1}, \ldots, t_{in_i}), i = 1, \ldots, m\}$ given $n_\cdot$ is $(n_\cdot)!$ times

$$L_c(\alpha^*) = \frac{(\alpha^*)^{n_\cdot} \prod_{i=1}^m \prod_{j=1}^{n_i} \exp(\alpha^* t_{ij})}{\left(\sum_{i=1}^m \exp(\alpha^* \tau_i) - m\right)^{n_\cdot}} .$$

Thus, $L_c(\alpha^*)$ is a conditional likelihood for α^* and may be used for inference purposes. In particular, a score test of $H_0 : \alpha^* = 0$ may be based on the score statistic $U_c(0)$, where $U_c(\alpha^*) = \partial \log L_c(\alpha^*)/\partial \alpha^*$, and $U_c(0)$ is defined by the limit of $U_c(\alpha^*)$ as $\alpha^* \to 0$. We obtain

$$U_c(\alpha^*) = \sum_{i=1}^m \sum_{j=1}^{n_i} t_{ij} + n_\cdot \left\{ \frac{1}{\alpha^*} - \frac{\sum_{i=1}^m \tau_i \exp(\alpha^* \tau_i)}{\sum_{i=1}^m \exp(\alpha^* \tau_i) - m} \right\} ,$$

and thus

$$U_c(0) = \sum_{i=1}^m \sum_{j=1}^{n_i} t_{ij} - \frac{n_\cdot}{2} \frac{\sum_{i=1}^m \tau_i^2}{\sum_{i=1}^m \tau_i} . \tag{3.42}$$

The variance of (3.42) can be obtained by using the conditional variance formula and the fact that (see Problem 2.4) when $\alpha^* = 0$ the process is a homogeneous Poisson process, and given n_i, the times T_{ij} $(j = 1, \ldots, n_i)$ are distributed as the ordered observations from a random sample of size n_i from the uniform distribution on $[0, \tau_i]$. Thus

$$\begin{aligned} \mathrm{var}\{U_c(0)|n_\cdot\} &= E\left[\mathrm{var}\left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} t_{ij} | n_\cdot \right\} \right] + \mathrm{var}\left[E\left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} t_{ij} | n_\cdot \right\} \right] \\ &= E\left\{ \sum_{i=1}^m n_i^2 \tau_i^2 / 12 \right\} + \mathrm{var}\left\{ \sum_{i=1}^m n_i \tau_i / 2 \right\} . \end{aligned}$$

Using the fact that when $\alpha^* = 0$, and $n_\cdot$ is fixed, the counts n_i, $i = 1, \ldots, m$, have a multinomial distribution with parameters $n_\cdot$ and $p_i = \tau_i/\tau_\cdot$, where $\tau_\cdot = \sum_{i=1}^m \tau_i$, we find that

$$\mathrm{var}\{U_c(0)|n_\cdot\} = \frac{n_\cdot}{3\tau_\cdot} \sum_{i=1}^m \tau_i^3 - \frac{n_\cdot}{4\tau_\cdot^2} \left\{ \sum_{i=1}^m \tau_i^2 \right\}^2 . \tag{3.43}$$

A test of the hypothesis $H_0 : \alpha^* = 0$ can be carried out using the statistic $U_c(0)/[\mathrm{var}\{U_c(0)|n_\cdot\}]^{1/2}$, which is asymptotically standard normal when H_0 is true. The limiting distribution applies when $n_\cdot \to \infty$ and so the test can be applied even to a single process $(m = 1)$ that is observed over a sufficiently long period of time.

The preceding test can be used quite generally to test for an absence of trend in a Poisson process if the τ_i are independent of the event processes, although it possesses good power mainly against alternatives in which the rate function is monotonic. It is important to note that it also requires that the process be Poisson. In many settings it would be preferable to use the robust estimating function methodology of Section 3.6.2 to test for an absence of trend. In particular, if we consider the family of rate functions $\rho(t;\alpha) = \exp(\alpha_1 + \alpha^* t)$ then the hypothesis $H_0 : \alpha^* = 0$ can readily be tested using the Wald statistic $\widehat{\alpha}^*/\text{s.e.}(\widehat{\alpha}^*)$, where $\text{s.e.}(\widehat{\alpha}^*) = [\widehat{\text{asvar}}(\widehat{\alpha}^*)]^{1/2}$ and $\widehat{\text{asvar}}(\widehat{\alpha}^*)$ is the robust variance estimate obtained from (3.35). This test requires that we obtain the estimates of α_1 and α^*, however. A test that does not require this can be based on asymptotic theory for estimating functions (Appendix A) for the case where $m \to \infty$. This leads to the statistic (see Problem 3.13)

$$Z = U_c(0)/\text{s.e.}(U_c(0))\,, \tag{3.44}$$

where $U_c(0)$ is as in (3.42) and $\text{s.e.}(U_c(0)) = [\widehat{\text{asvar}}\{U_c(0)\}]^{1/2}$ is given in Problem 3.13. Large values of $|Z|$ provide evidence of trend; $p-$values are obtained from the standard normal distribution because Z is asymptotically $(m \to \infty)$ normal under H_0.

The trend tests above are for the case where the m individual processes are assumed to be identically distributed. It is also easy to develop a test for the case where the alternative hypothesis is

$$\rho_i(t) = \exp(\alpha_i + \alpha^* t) \qquad i = 1, \ldots, m\,,$$

or, more generally, the case where $\rho_i(t) = \exp\{\alpha_i + \alpha^* g(t)\}$, when the m processes are Poisson. A simple score test of $H_0 : \alpha^* = 0$ is obtained by considering the distribution of the event times t_{ij} $(j = 1, \ldots, n_i)$ conditional on the n_i $(i = 1, \ldots, m)$. A development similar to that leading to (3.42) gives the conditional score statistic

$$U_c(0) = \sum_{i=1}^{m} I(n_i > 0) \sum_{j=1}^{n_i} (t_{ij} - \frac{1}{2}\tau_i) = \sum_{i=1}^{m} n_i(\bar{t}_i - \tau_i/2)\,, \tag{3.45}$$

where $\bar{t}_i = \sum_{j=1}^{n_i} t_{ij}/n_i$, assuming $n_i \geq 1$. Similarly, we find that conditional on $n_1, \ldots, n_m$,

$$\text{var}\{U_c(0)\} = \sum_{i=1}^{m} n_i \tau_i^2/12$$

under H_0, so a test for trend can be based on $U_c(0)/\text{s.e.}(U_c(0))$, which is asymptotically normal under H_0. This test can also be developed in the robust estimating function framework; see Problem 3.13, part d.

3.7.2 Tests for Multiplicative Covariate Effects

Multiplicative models provide a convenient parsimonious summary of covariate effects but it is important that the assumption of multiplicative effects be

carefully assessed. With fixed covariates taking on a relatively small number of unique combinations, plots of $\log \widehat{\mu}_{0j}(t)$ versus t, $j = 1, 2, \ldots, J$ should give J roughly parallel lines, where j indexes the unique covariate configurations and $\widehat{\mu}_{0j}(t)$ is the Nelson–Aalen estimate (3.17) based only on those individuals with this configuration. For other situations, model expansion provides tests of fit.

A flexible way to generalize the standard multiplicative model is to consider alternatives of the form

$$\rho_i(t) = \rho_0(t) \exp(x_i'(t)\beta(t)) \tag{3.46}$$

in which the effects of the covariates possibly change over time. Such models may be easily fit by introducing auxiliary covariates which are functions of time, or so-called "defined" time-dependent covariates. For example, let $g(t)$ denote a known scalar function and consider the model

$$\rho_i(t) = \rho_0(t) \exp(x_i'(t)\beta + g(t)x_i'(t)\theta) = \rho_0(t) \exp(x_i'(t)(\beta + g(t)\theta)),$$

where $\theta = (\theta_1, \ldots, \theta_p)'$. When $\theta = 0$, the usual multiplicative model is retrieved but generally, if controlling for $x_{i1}(t) \ldots x_{ij-1}(t), x_{ij+1}(t) \ldots x_{ip}(t)$, the relative rate of events for an individual with $x_{ij}(t)$ one unit higher than another individual, is $\beta_j + g(t)\theta_j$. A test of $H_0 : \theta_j = 0$ is therefore a test of the proportional rates assumption for $\{x_j(t), 0 \leq t\}$. It may be based on a model with $\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p$ estimated, or a model with $\theta_k = 0$, $k \neq j$, where all other covariates are assumed to act multiplicatively. An omnibus test of $H_0 : \theta = 0$ can also be carried out. Very similar tests, which are implemented in S-PLUS and R, are discussed in Section 3.7.3.

Different choices of $g(t)$ will lead to different evidence against $H_0 : \theta_j = 0$, and some common choices are $g(t) = t$ and $g(t) = \log t$. One may also use different functions for each of the covariates by considering models $\rho_i(t) = \rho_0(t) \exp(x_i'(t)\beta + z_i'(t)\theta)$ where $z_i(t) = (z_{i1}(t), \ldots, z_{ip}(t))'$ with $z_{ij}(t) = g_j(t)x_{ij}(t)$, $j = 1, \ldots, p$.

Maximum partial likelihood estimation can be inconvenient because values for all covariates at all event times are needed, and so some software restricts time-varying covariates to be piecewise-constant. Let us therefore make $g(t)$ piecewise-constant, by considering a partition $0 < a_1 < \cdots < a_q$ and defining $g_k(t) = I(a_{k-1} < t \leq a_k)$, $k = 2, \ldots, q$. Let $z_{ijk}(t) = x_{ij}(t)g_k(t)$, $z_{ij}(t) = (z_{ij2}(t), \ldots, z_{ijq}(t))'$, and let $z_i(t) = (z_{i1}'(t), \ldots, z_{ip}'(t))'$ be a $1 \times p(q-1)$ vector. Also, let $\theta_j = (\theta_{j2}, \ldots, \theta_{jq})'$, and $\theta = (\theta_1', \ldots, \theta_p')'$ be a $p(q-1) \times 1$ vector of parameters. We now consider the model

$$\rho_i(t) = \rho_0(t) \exp(x_i'(t)\beta + z_i'(t)\theta).$$

In this case the relative rate at time t associated with a one-unit increase in $x_j(t)$ while controlling for all other covariates is $\beta_j + \sum_{k=2}^{q} g_k(t)\theta_{jk}$. Likelihood ratio tests of $H_0 : \theta_j = 0$ are easy to carry out in this setting for either parametric or semiparametric modls.

One difficulty with this approach is the selection of the number and location of the cutpoints $a_1, \ldots, a_q$. The greater the number of cutpoints the more flexible is the alternative model. However, a larger number of parameters then need to be estimated. Choosing a moderate number of points, even two or three, is usually advisable to facilitate estimation and interpretation of effects in the event the null hypothesis is rejected. The location of the cutpoints should be chosen such that a comparable number of event times fall into the resulting intervals, to ensure satisfactory power to detect trends. Of course there may also be scientific rationale for locating cutpoints at particular times and ideally these locations are prespecified rather than chosen post hoc according to plots of the data. Finally, note that in some settings it may be desirable to allow q to vary across covariates.

3.7.3 Generalized Residuals, Martingales, and Assessment of Fit

The preceding section describes how proportionality assumptions for specific covariates can be tested. Other model checks can similarly be made by expanding the model of interest, as we discuss later in this section. Another approach, which lends itself to informal graphical assessments of fit, is through types of generalized residuals which we now describe.

The quantities defined by (2.46) can, for the case of several processes $i = 1, \ldots, m$, with intensities $\rho_i(t)$, be written as

$$E_{ij} = \int_{T_{i,j-1}}^{T_{ij}} \rho_i(u)du \qquad j = 1, \ldots, n_i + 1\,, \tag{3.47}$$

where T_{i0} and T_{i,n_i+1} are the start and stop times for observation of the process $\{N_i(t), 0 \le t\}$ and $T_{i1} < \cdots < T_{in_i}$ are the times at which events before T_{i,n_i+1} occur. Under the conditions in Section 2.1, the E_{ij} $(j = 1, \ldots, n_i)$ are standard exponential random variables and the quantities

$$R_{ij} = E_{i1} + \cdots + E_{ij} \qquad j = 1, \ldots, n_i\,,$$

can be viewed as the occurrence times of events in a homogeneous Poisson process. That is, conditional on the values of external covariates, we can view

$$s = R_i(t) = \int_0^t \rho_i(u)du \tag{3.48}$$

as a transformed time scale, and the processes $N_i^*(s) = N_i(R_i^{-1}(s))$ are, for $s \ge 0$, homogeneous Poisson processes.

Informal model checks can be based on generalized residuals $\widehat{E}_{ij}$ and $\widehat{R}_{ij}$ that are obtained by replacing the T_{ij} in (3.47) with the observed t_{ij} and $\rho_i(u)$ with the maximum likelihood estimate $\widehat{\rho}_i(u)$. For sufficiently large samples, and under mild conditions, the $\widehat{E}_{ij}$ should be similar to independent standard exponential random variables, if the specifications $\rho_i(u)$ are correct.

These residuals are mainly useful with parametric models, and can supplement simple comparisons of $N_i(t)$ and $\widehat{R}_i(t)$, which in the case of a Poisson process is just the estimated mean function $\widehat{\mu}_i(t)$. For example, a plot of the Nelson–Aalen estimate $\widehat{\Lambda}^*_{NA}(s)$ based on the transformed times $s_{ij} = \widehat{R}_{ij}$ should be roughly linear with slope one if the proposed model is adequate, and departures suggest deficiencies in the time-trend specification. An alternative is to consider the scaled values $\widehat{u}_{ij} = \widehat{R}_{ij}/\widehat{R}_{i,n_i+1}$ $(j = 1, \ldots, n_i)$, which under the assumed model should look like the order statistics in a random sample from the Uniform(0,1) distribution. Departures from uniformity in a probability plot of the $\widehat{u}_{ij}$ suggest misspecification of the baseline rate function or of terms involving time-varying covariates. It should be noted that terms in $\rho_i(t)$ that are constant with respect to time disappear in the $\widehat{u}_{ij}$.

Plots of the $\widehat{E}_{ij}$ can also be made. Plots of successive values $(\widehat{E}_{i,j-1}, \widehat{E}_{ij})$ that show association suggest non-Poisson behavior, such as dependence on previous events. A comparison of the $\widehat{E}_{ij}$ with the standard exponential distribution using a probability plot can detect extra-Poisson variation or carry-over effects. These plots are also useful with more general intensity-based models, and we defer an illustration of their use to Section 5.2.3.

A complementary set of methods for model checking is provided via martingales. They give a powerful framework for motivating the construction of estimating functions and deriving asymptotic results for event history analysis (see Fleming and Harrington, 1991; Andersen et al., 1993). Here we show how they can be exploited to derive results useful for assessment of model fit.

The *counting process martingale* for subject i is defined as

$$M_i(t) = \bar{N}_i(t) - \int_0^t Y_i(u)\rho_i(u)du\,.$$

The process $\{M_i(t), 0 \le t\}$ has expectation zero and uncorrelated increments. One may also define for $s < t$,

$$M_i(s,t) = \int_s^t dM_i(u) = \bar{N}_i(s,t) - \int_s^t Y_i(u)\rho_i(u)du\,.$$

The corresponding *martingale residual process* is obtained by replacing unknown quantities with their respective estimates to give, for example,

$$\widehat{M}_i(s,t) = \int_s^t Y_i(u)\{dN_i(u) - \widehat{\rho}_i(u)du\}\,.$$

Various plots and tests based on martingale residuals have been proposed for assessing the functional form of $\rho_i(t)$, although many of these are not well understood in terms of their power to detect model departures. One approach due to Grambsch and Therneau (1994) gives tests that are similar to ones in Section 3.7.2, and is implemented in the S-PLUS or R function `cox.zph`. It uses a statistic based on Schoenfeld residuals, defined below, to test the

proportionality assumptions in multiplicative semiparametric models where $\rho_i(t) = \rho_0(t)\exp(x_i'(t)\beta)$. To begin we introduce the notion of martingale transforms.

The contribution from individual i to the log-likelihood function (2.7) or (2.59) can be written in the case of a Poisson process as

$$\int_0^\infty Y_i(u)\log\rho_i(u)dN_i(u) - \int_0^\infty Y_i(u)\rho_i(u)du\,.$$

The score vector (3.5) for β in the multiplicative model in (3.3) is obtained by differentiation. It may be written as a stochastic integral as in (2.60), in which case we can view the score vector as a "score process". This stochastic integral represents a *martingale transform* of a $p\times 1$ vector $W_i(s;\beta)$. The same is true for the semiparametric multiplicative model, where the partial likelihood score vector (3.21) has components $i = 1,\ldots,m$ that can be rewritten as

$$U_i(\beta,t) = \int_0^t W_i(s;\beta)dM_i(s) = \int_0^t \{x_i(s) - \bar{x}(s;\beta)\}dM_i(s)\,,$$

where $W_i(s;\beta)$ is given by (3.22), and for convenience we define

$$\bar{x}(s;\beta) = \frac{\sum_{i=1}^m Y_i(s)x_i(s)\exp(x_i'(s)\beta)}{\sum_{i=1}^m Y_i(s)\exp(x_i'(s)\beta)}\,,$$

which is a $p\times 1$ vector of the weighted means of the covariates among those at risk at time $s > 0$. We write $U_i(\beta,t) = (U_{i1}(\beta,t),\ldots,U_{ip}(\beta,t))'$ to indicate the individual elements of $U_i(\beta,t)$ corresponding to $\beta_1,\ldots,\beta_p$.

If $t_{(1)} < \cdots t_{(K)}$ represent the K unique sorted event times, note that one may further decompose $U_{ij}(\beta,t)$ as a sum over the observed event time intervals $k = 1,\ldots,K$, as

$$U_{ijk}(\beta) = \int_{t_{(k-1)}}^{t_{(k)}} W_{ij}(s;\beta)dM_i(s)\,,$$

where $t_{(0)} = 0$.

The *Schoenfeld residual* at $t_{(k)}$ for β_j is defined as $r_{jk} = \sum_{i=1}^m U_{ijk}(\beta)$, $j = 1,\ldots,p$, and the full $p\times 1$ vector of Schoenfeld residuals at $t_{(k)}$ is $r_k = (r_{1k},\ldots,r_{pk})'$. Furthermore, let

$$V(s;\beta) = \frac{\sum_{i=1}^m Y_i(s)\exp(x_i'(s)\beta)[x_i(s) - \bar{x}(s;\beta)][x_i(s) - \bar{x}(s;\beta)]'}{\sum_{i=1}^m Y_i(s)\exp(x_i'(s)\beta)}$$

denote the $p\times p$ weighted variance matrix for $x_i(s)$.

Suppose the null model that we wish to check is of the form $\rho_i(t) = \rho_0(t)\exp(x_i'\beta)$ and the alternative is given by $\rho_i(t) = \rho_0(t)\exp(x_i'\beta(t))$ with $\beta(t) = (\beta_1(t),\ldots,\beta_p(t))'$, $\beta_j(t) = \beta_j + (g_j(t) - \bar{g}_j)\theta_j$, and $\bar{g}_j = \sum_{k=1}^K g_j(t_{(k)})$, $j = 1,\ldots,p$. Grambsch and Therneau (1994) show that if $r_k^* = V^{-1}(t_{(k)};\beta)r_k$,

then $E(r^*_{jk}) \approx \theta_j(g_j(t_k) - \bar{g}_j)$. This suggests plots of r^*_{jk} versus functions of time may be useful for revealing possible departures from the null hypothesis of proportionality for covariate x_j. Smooths of such plots help reveal trends and suggest alternative functional forms for covariate effects. Such plots can be created using the S-PLUS or R function `cox.zph` as illustrated in Section 5.4.

Plots of these residuals are not always easily understood, however, and it is useful to combine them to form a test statistic sensitive to departures from the assumed null model in the direction of the specified alternative. Specifically, Grambsch and Therneau (1994) suggest that if $G_k = \text{diag}(g_j(t_{(k)}) - \bar{g}_j, j = 1, \ldots, p)$ then $\tilde{\theta} = \sum_{k=1}^K G^*_k r_k$ estimates θ, where $G^*_k = \widehat{\Sigma}^{-1}\ G_k$ and $\widehat{\Sigma} = [\widehat{A} - \widehat{B}\widehat{C}^{-1}\widehat{B}']^{-1}$ with $\widehat{A} = \sum_{k=1}^K G_k V(t_{(k)}; \widehat{\beta}) G_k$, $\widehat{B} = \sum_{k=1}^K G_k V(t_{(k)}; \widehat{\beta})$ and $\widehat{C} = \sum_{k=1}^K V(t_{(k)}; \widehat{\beta})$. Here $\widehat{\Sigma}$ is a consistent estimator of the covariance matrix for $\tilde{\theta}$ and so one degree of freedom tests of the proportionality assumption for x_j can be based on $\tilde{\theta}_j / \widehat{\Sigma}^{1/2}_{jj}$, which is asymptotically standard normal under the null model. Global tests for the null model may be based on a quadratic form $\tilde{\theta}' \widehat{\Sigma}^{-1} \tilde{\theta}$, which is asymptotically χ^2_p under the null hypothesis $\rho_i(t) = \rho_0(t) \exp(x_i' \beta)$. Both kinds of tests are available in the output from `cox.zph`, with a choice of functions $g(t)$ being available. Such tests usually yield results that are close to those for the Wald or likelihood ratio tests based on the discussion following (3.46).

Martingale residuals can also be utilized to provide other model checks. For example, residuals $\widehat{M}_i(t)$ for specified values of t can be plotted against the values of fixed covariates x_i; such plots should be consistent with the fact that $E\{M_i(t)\} = 0$ if the model is correct. The usefulness of such plots is limited if individuals tend to have very few events. If there are sufficiently large numbers of events for each individual, plots of successive increments $\widehat{M}_i(a_{j-1}, a_j)$ for a sequence of times $0 = a_0 < a_1 < \cdots < a_n$ can also be useful, and might indicate extra-Poisson variation or association in the successive event counts.

Formal goodness-of-fit tests can be based on residuals described in this section. One example is the quadratic statistic defined above for testing proportionality of covariate effects. More generally, suppose that we take a parametric base model with intensity functions $\rho_i(t)$ and consider the expanded model with intensity functions

$$\lambda_i(t|H(t)) = \rho_i(t) \exp(z_i'(t)\gamma)\,,$$

where $z_i(t)$ may include functions of covariates, time, or prior event history. We now consider score tests, examples of which were considered earlier in Section 3.7.1 for testing trend. Here, the idea is to develop tests of the Poisson process itself. The derivative with respect to γ of the log-likelihood function's contribution from individual i is, from (2.7) or (3.5),

$$S_{i\gamma}(\gamma, \alpha) = \int_0^\infty Y_i(u) z_i(u) \{dN_i(u) - \lambda_i(u|H(u))du\}\,,$$

where α parameterizes the $\rho_i(t)$. To test the hypothesis $H : \gamma = 0$, and thus check the base Poisson model, we consider the score statistic $S(0) = \sum_{i=1}^{m} S_{i\gamma}(0, \widehat{\alpha})$ with $\widehat{\alpha}$ estimated from the Poisson model $\rho_i(t; \alpha)$. This gives

$$S(0) = \sum_{i=1}^{m} \int_0^{\infty} z_i(u) d\widehat{M}_i(u)\,, \tag{3.49}$$

where

$$\widehat{M}_i(t) = \bar{N}_i(t) - \int_0^t Y_i(u)\widehat{\rho}_i(u)du \qquad i = 1, \ldots, m\,,$$

and $\widehat{\rho}_i(u) = \rho_i(u; \widehat{\alpha})$.

By appropriate choice of $z_i(u)$ in (3.49) we can test for various types of model departures without having to fit an expanded model. Values of $S(0)$ away from zero indicate evidence against the Poisson model. Often it is satisfactory to consider various scalar covariates $z(u)$, so that $S(0)$ is a scalar. Applying asymptotic theory for score tests (see Appendix A), it can be shown that if the Poisson model $\rho_i(t; \alpha)$ is correct, the score statistic $S(0)$ is asymptotically normal as $m \to \infty$, with variance estimated consistently by

$$V(0) = \sum_{i=1}^{m} \int_0^{\infty} Y_i(u) z_i^2(u) \widehat{\rho}_i(u) du - A' B A\,,$$

where $B = I(\widehat{\alpha})^{-1}$ is the asymptotic covariance matrix for $\widehat{\alpha}$ in the Poisson model, and

$$A = \sum_{i=1}^{m} \int_0^{\infty} Y_i(u) z_i(u) \left[\frac{\partial \log \widehat{\rho}_i(u)}{\partial \widehat{\alpha}} \right] \widehat{\rho}_i(u) du\,.$$

For m sufficiently large, $S(0)/V(0)^{1/2}$ is close to standard normal so $p-$values may readily be obtained. An alternative is to obtain $p-$values by simulation.

In cases where m is small but the total number of events is large, the normal approximation for $S(0)$ may still be reasonable; this can be investigated by simulation. To obtain $p-$values, simulation is recommended unless previous investigation has shown the normal approximation to be accurate.

Finally, we observe that if $z_i(u)$ is constant over intervals $(t_{i,j-1}, t_{ij}]$ between successive events, the statistic (3.49) is a weighted sum of the residuals $\widehat{E}_{ij}$ defined earlier. In particular,

$$S(0) = \sum_{i=1}^{m} \sum_{j=1}^{n_i+1} w_{ij}(1 - \widehat{E}_{ij})\,,$$

where $z_i(u) = w_{ij}$ for $t_{i,j-1} < u \leq t_{ij}$.

3.7.4 Tests for Extra-Poisson Variation

Sometimes the event processes for individuals may be Poisson, but heterogeneity in rate functions may exist that is not accounted for by covariates, creating extra-Poisson variation in the numbers of events. We now describe a simple test for this that can be used with parametric models.

Consider the multiplicative conditionally Poisson model given in (3.28) and note that if $\rho_i(t) = \rho_0(t;\alpha)\exp(x_i'\beta)$ it can be written as

$$\lambda_i(t|H_i(t), u_i) = u_i\rho_0(t)\exp(x_i'\beta)\,,$$

with u_i taken to have mean zero and variance ϕ. The parameter ϕ reflects the extent of extra-Poisson variation, with $\phi = 0$ giving the Poisson model. A likelihood ratio test of $H_0 : \phi = 0$ is relatively easy to carry out when a model for the random effects is specified, as in Section 3.5. However, because the value of $\phi = 0$ is on the boundary of the parameter space, the limiting distribution of the likelihood ratio statistic W under H_0 has probability mass of 0.5 at 0 and is distributed as $0.5\chi_1^2$ for $W > 0$. A simple alternative is to use a score test, which requires that only the Poisson process model be fitted.

Here we assume that the times τ_i $(i = 1, \ldots, m)$ are independent of the event processes, and condition on their values. The test below is based on the likelihood function arising from the distribution of the counts $N_i(0, \tau_i)$ as discussed in Section 2.2.3. The contribution from subject i to the likelihood score statistic for testing $H_0 : \phi = 0$ is given by

$$(n_i - \mu_i(\tau_i))^2 - \mu_i(\tau_i)\,, \tag{3.50}$$

which has expectation zero under the Poisson model. The standardized form of this test statistic for $\phi = 0$ is (e.g. Dean and Lawless, 1989)

$$\frac{\sum_{i=1}^m [(n_i - \widehat{\mu}_i(\tau_i))^2 - \widehat{\mu}_i(\tau_i)]}{[2\sum_{i=1}^m \widehat{\mu}_i^2(\tau_i)]^{1/2}}\,, \tag{3.51}$$

where $\widehat{\mu}(\tau_i)$ is the Poisson estimate $\mu_i(\tau_i;\widehat{\alpha})$. This asymptotically $(m \to \infty)$ follows a standard normal distribution under the null hypothesis. Simulation results suggest this statistic is somewhat slow to reach its limiting distribution and Dean and Lawless (1989) have proposed small sample corrections for the case where only the counts $N_i(\tau_i) = n_i$ are observed. An alternative approach is to obtain a $p-$value by simulation. This is appealing because data need only be generated under the Poisson model. Ng and Cook (1999a) consider analogous tests for recurrent events where event times are known, based on piecewise-constant models for baseline rate functions.

3.7.5 Two-Sample Test Statistics Based on Rates

We are often interested in comparing the occurrence of events in two or more groups. In industrial settings one may be interested in comparing rates of

warranty claims from cars manufactured in different periods or at different plants. Medical studies may be directed at examining event rates for patients under different therapeutic conditions.

One natural framework for making comparisons of two groups, 1 and 2, is via models $\rho_i(t) = \rho_0(t)\exp(x_i\beta)$ where $x_i = 1$ if individual i is in group 2 and $x_i = 0$ otherwise. Wald tests of $H_0 : \beta = 0$ may be carried out under parametric or semiparametric Poisson models of Sections 3.2 and 3.4, respectively, and such tests will be most powerful for detecting differences within the class of multiplicative Poisson models (i.e. with proportional rate functions). More general types of group differences may be reflected by models of the form

$$\rho_i(t) = \rho_0(t)\exp(x_i\beta + \gamma x_i g(t))\,,$$

where, as discussed in Section 3.7.2, $g(t)$ is a specified function of time. Based on such models, two degree of freedom tests of the hypothesis of no group differences may be obtained for $H_0 : \beta = \gamma = 0$. It is most common to consider such tests with $\rho_0(t)$ treated nonparametrically.

Simple graphical displays based on nonparametric estimates of mean functions can provide insight into the differences in two groups. For example, plots of the log cumulative mean functions for the two groups should be roughly parallel if the simple multiplicative model is appropriate, whereas plots of the untransformed mean functions should be roughly parallel if an additive model is appropriate.

Score tests also have appeal in this setting because of their simplicity. Consider two groups of subjects as before, with mutually independent counting processes. Now let $\{N_{ki}(t), 0 \le t\}$ denote the counting process for subject i in group k, and let $Y_{ki}(t) = 1$ if subject i in group k is at risk at time t, where $k = 1$ and $k = 2$ denote the two groups. The rate and mean functions for subjects in group k are given by $E\{dN_{ki}(t)\} = \rho_k(t)dt$ and $E\{N_{ki}(t)\} = \mu_k(t)$ respectively, $i = 1, \ldots, m_k$.

A family of test statistics mentioned by Lawless and Nadeau (1995) is based on

$$U(\tau) = \int_0^\tau w(s)\{d\widehat{\mu}_2(s) - d\widehat{\mu}_1(s)\}\,, \tag{3.52}$$

where $\tau = \max(\tau_{ki})$ is the maximum followup time,

$$w(s) = \frac{Y_{1\cdot}(s)Y_{2\cdot}(s)a(s)}{Y_{\cdot\cdot}(s)}\,, \tag{3.53}$$

and $a(s)$ is a fixed weight function. It is straightforward to show that if $a(s) = 1$, then (3.52) can be obtained as a score test statistic of the null hypothesis $H_0 : \rho_2(t) = \rho_1(t)$ versus the alternative $H_A : \rho_2(t) = \rho_1(t)\exp(\beta)$ from (3.21), with $x_i = 1$ for subjects in group 2 and $x_i = 0$ for subjects in group 1. More generally, (3.52) arises as a score test of $H_0 : \rho_2(t) = \rho_1(t)$ versus $H_A : \rho_2(t) = \rho_1(t)\exp(\beta a(t))$, which accommodates nonproportionality.

Under the assumption that the observations are generated according to a Poisson process, one can show that

$$m^{-1}\widehat{\text{var}}_P\{U(\tau)\} = \frac{1}{m}\int_0^\tau (w(s))^2 \left\{ \frac{d\widehat{\mu}_1(s)}{Y_{1\cdot}(s)} + \frac{d\widehat{\mu}_2(s)}{Y_{2\cdot}(s)} \right\} \tag{3.54}$$

is a consistent estimate for $\text{var}(\sqrt{m}^{-1}U(\tau))$ based on the expected information function corresponding to (3.26). The standardized form of this statistic, $\bar{U}_P^2(\tau) = (U(\tau))^2/\widehat{\text{var}}_P\{U(\tau)\}$, is asymptotically χ_1^2 under H_0 and large absolute values give evidence against the null hypothesis.

Following the arguments of Section 3.6 one can show that more generally, a consistent variance estimator for $\sqrt{m}^{-1}U(\tau)$ is $m^{-1}\widehat{\text{var}}_R\{U(\tau)\}$ given by

$$\frac{1}{m}\sum_{k=1}^{2}\sum_{i=1}^{m_k}\left[\int_0^\tau w(s)\frac{Y_{ki}(s)}{Y_{k\cdot}(s)}\,\{dN_{ki}(s) - d\widehat{\mu}_k(s)\}\right]^2 . \tag{3.55}$$

This variance estimator is robust to departures from Poisson assumptions and as $Y_{1\cdot}(s)$ and $Y_{2\cdot}(s)$ become large over $[0,\tau]$, the pseudo-score statistic $\bar{U}_R^2(\tau) = (U(\tau))^2/\widehat{\text{var}}_R\{U(\tau)\}$ approaches a χ_1^2 random variable under H_0 for a wide class of underlying point processes. Large observed values of $\bar{U}_R^2(\tau)$ provide evidence against H_0. A point to note, however, is that the followup times τ_{ki} must here be independent of the event processes.

Many statistics of the form (3.52) with $a(s) > 0$ in (3.53) will be effective against departures from H_0 in which the mean functions do not cross. This is analogous to the situation for "weighted log-rank" tests of equality for lifetime distributions (e.g. Andersen et al., 1993, Ch. 5; Fleming and Harrington, 1991, Ch. 7), where the tests are primarily effective when the hazard functions and survivor functions of the distributions do not cross. Because the test statistic with $a(s) = 1$ in (3.52) and (3.53) arises as a pseudo-score test of $H_0 : \beta = 0$ in the model where $\rho_i(t) = \rho_0(t)\exp(x_i\beta)$ it is powerful against alternatives of this form. An alternative weight function given by $a(s) = t - s$ can be shown to generate a statistic that is c times the area between the empirical mean functions if there is no censoring over $[0,\tau]$ and c is a positive real constant satisfying $c = Y_{1\cdot}(s)Y_{2\cdot}(s)/Y_{\cdot\cdot}(s)$, $0 \le s \le t$. In settings where censoring occurs and $Y_{1\cdot}(s)Y_{2\cdot}(s)/Y_{\cdot\cdot}(s)$ varies with time, the statistic does not retain this interpretation exactly, but it remains powerful for scenarios in which the mean functions are nonproportional, but do not cross.

If $\mu_1(t)$ and $\mu_2(t)$ are expected to cross, other procedures are preferable. For example, if the difference in the mean functions is fairly systematic, it may be possible to use a group indicator variable along with a time-dependent covariate to develop a test based on a more general multiplicative model. Here we restrict consideration to models of the form $\rho_i(t) = \rho_0(t)\exp(\beta_1 x_i + \beta_2 x_i t)$, noting that this allows for a time trend in the ratio of the rate functions. Thus if $\beta = (\beta_1, \beta_2)'$, a test of $\beta = 0$ would be more powerful for these types of

departures from $H_0 : \mu_1(t) = \mu_2(t)$, $t \geq 0$. The pseudo-score vector resulting from this model can be written as

$$U(\tau) = \int_0^\tau w(s)\{d\widehat{\mu}_1(s) - d\widehat{\mu}_2(s)\}\,,$$

where $w(s) = (w_1(s), w_2(s))' = a(s)\, Y_{1\cdot}(s)\, Y_{2\cdot}(s)/Y_{\cdot\cdot}(s)$ with $a(s) = (1, s)'$. A two degree of freedom pseudo-score test of $H_0 : \mu_1(t) = \mu_2(t)$, $t \geq 0$ based on this model can then be constructed because for large m,

$$\bar{U}(\tau) = U'(\tau)\widehat{\text{var}}^{-1}\{U(\tau)\}U(\tau) \sim \chi^2_2$$

under the null hypothesis, where the covariance matrix $\text{var}\{U(\tau)\}$ is estimated either under a Poisson model or robustly, according to whether the processes are Poisson.

3.8 Applications and Illustrations

3.8.1 Rat Mammary Tumor Data

Consider the data on the development of tumors in rats discussed in Section 2.2.4. Here we provide some illustrative analyses based on models described in this chapter. First we consider separate analyses of the two treatment groups based on parametric Poisson models. We also consider nonparametric Poisson models (Section 3.4.1), robust nonparametric estimation (Section 3.6.1), and two-sample tests (Section 3.7.5). In order to assess treatment effect, parametric and semiparametric Poisson regression models are then discussed along with semiparametric mixed Poisson models (Section 3.5.3) and robust semiparametric methods (Section 3.6.3). S-PLUS or R data frames, code, and output for these analyses are provided in Appendix C.

Separate Poisson Analyses of the Control and Treatment Arms

Let $\mu_k(t)$ denote the mean function for group k, where $k = 1$ for the treated group and $k = 2$ for the control group. Figure 3.1 gives plots of $\log \widehat{\mu}_k(t)$ versus t and versus $\log t$, $k = 1, 2$, where $\widehat{\mu}_k(t)$ is the Nelson–Aalen estimate given by (3.17). The latter plots are roughly linear suggesting parametric rate functions

$$\rho_k(t) = \alpha_{2k}\alpha_{1k}^{\alpha_{2k}} t^{\alpha_{2k}-1} \tag{3.56}$$

are reasonable. Because this has the same form as a Weibull hazard function, as mentioned in Section 3.2.3, such a model can be fit using software for survival analysis which can accommodate left truncation (also known as "delayed entry"). Here we use the S-PLUS function `censorReg` but note that there the parameterization adopted is for a location-scale formulation (Lawless, 2003a,

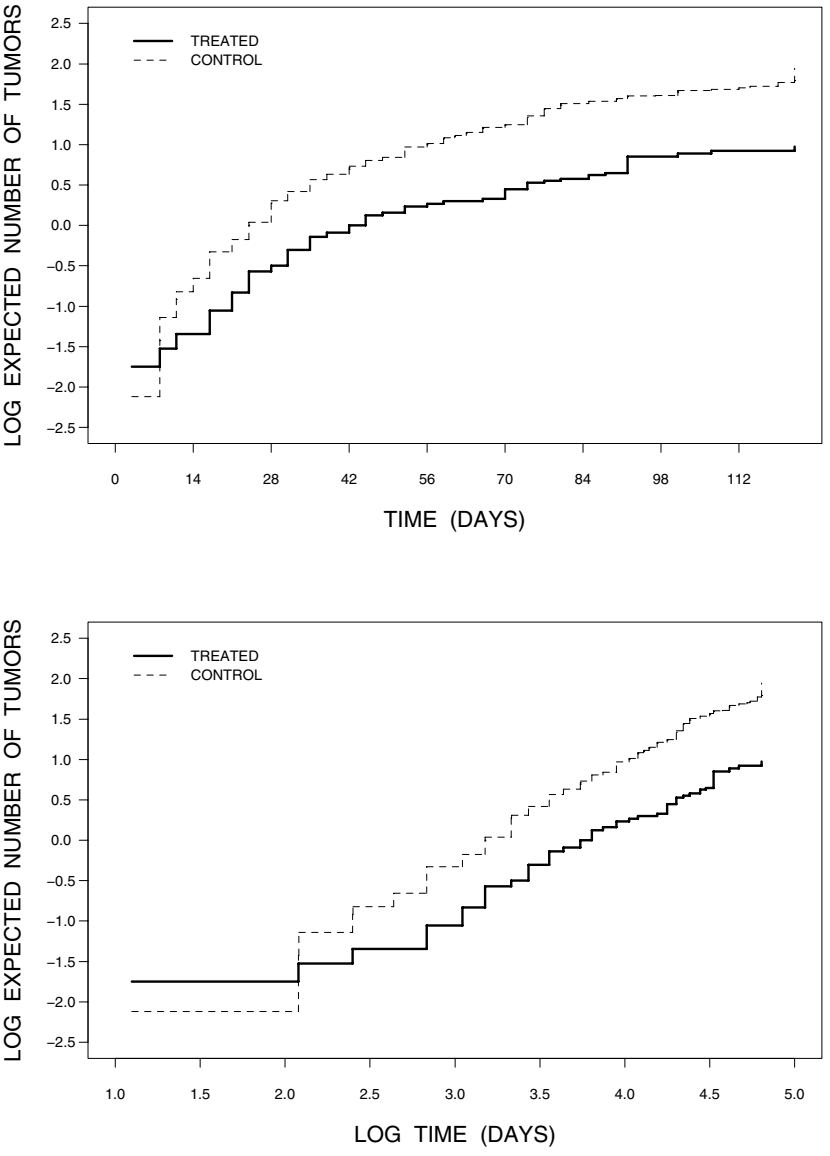


Fig. 3.1. Plots of $\log(\widehat{\mu}_k(t))$ versus t and $\log(\widehat{\mu}_k(t))$ versus $\log t$ for treated and control groups based on Nelson–Aalen estimates of mean functions.

Section 5.2) so that for group k, the scale parameter is α_{2k}^{-1} in this notation and the intercept is $-\log \alpha_{1k}$. For the control arm this gives estimates (s.e.) $\widehat{\alpha}_{12} = 0.042$ (0.0065) and $\widehat{\alpha}_{22} = 1.09$ (0.0891), giving $\widehat{\mu}_2(t) = (0.042t)^{1.09}$, which is plotted in Figure 3.2 along with the corresponding estimate for the treatment arm.

We next consider use of a piecewise-constant rate function with cutpoints at 30, 60, and 90 days. Recall from (3.15) that the likelihood arising from the piecewise model is a product of likelihoods for a time-homogeneous Poisson model. The estimated mean functions for the control and treated rats based on this model are also given in Figure 3.2.

The Nelson–Aalen estimates (3.17) of the mean functions for the treated and control rats shown in Figure 3.2 can be used as a basis for judging fit of the parametric models. There is close agreement between the piecewise-constant and Nelson–Aalen estimates over the course of observation for both groups. The Weibull rate function gives a slightly lower estimate of the expected number of tumors over the latter half of the followup period for the control rats. The data frame and code for estimating the rate functions under the Weibull and piecewise-constant models, as well as for the nonparametric estimates, are given in Appendix C.

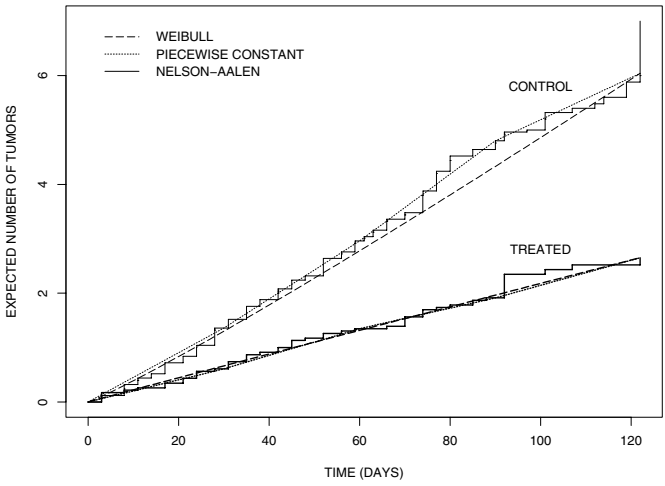


Fig. 3.2. Estimated mean functions for the treated and control groups obtained under the assumptions of Weibull and piecewise-constant (cutpoints at 30, 60, and 90 days) parametric rate functions, and Nelson–Aalen estimates.

Two-Sample Tests and Tests for Extra-Poisson Variation

Tests of the hypothesis $H_0 : \mu_1(t) = \mu_2(t)$ versus $H_A : \mu_1(t) \neq \mu_2(t)$ can be carried out using (3.52) and (3.54) or (3.55). These give observed statistics $\bar{U}_P^2(122) = 30.54$ and $\bar{U}_R^2(122) = 14.65$ for the standardized statistics based on (3.54) and (3.55), respectively. The statistic based on (3.54) is asymptotically χ_1^2 under H_0, if a Poisson model is correct, but the statistic based on (3.55) is asymptotically χ_1^2 under H_0 more generally. The $p-$values based on χ_1^2 are $p < 0.0001$ and $p = 0.0001$, respectively, both indicating strong evidence against the hypothesis. The fact that the two statistics differ suggests that the Poisson assumption should be questioned, so the robust test is preferred.

Figure 3.3 contains two sets of pointwise 95% confidence limits for the nonparametric estimates of the mean functions. The first set is based on the Poisson assumption and is given by exponentiating the limits of a 95% confidence interval for $\log \mu(t)$, as described following (3.18) of Section 3.4.1. The second set is based on the robust variance (3.34) and is again obtained by exponentiating limits for $\log \mu(t)$ computed as in Section 3.4.1. The wider robust limits reflect the presence of extra-Poisson variation, especially in the control group.

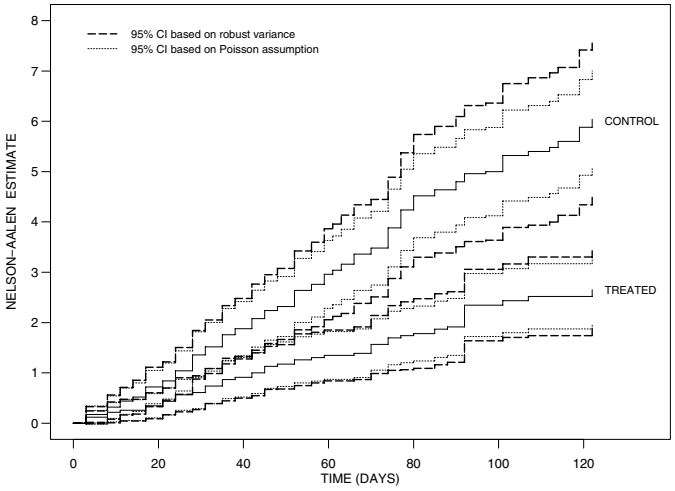


Fig. 3.3. Estimated mean functions and pointwise Poisson-based and robust 95% confidence intervals for the treated and control groups.

The statistic (3.51) for testing for extra-Poisson variation depends on the estimated mean at the common end of followup $\tau = 122$, which is $\widehat{\mu}_2(\tau) = 6.040$ for the control arm and $\widehat{\mu}_1(\tau) = 2.652$ for the treated arm. The statistic

(3.51) gives an observed value of 5.27 and a $p-$value $\Pr(\chi_1^2 > 5.27^2) < 0.001$ suggesting, as one would expect based on the above results, evidence of extra-Poisson variation.

Regression Analyses

The plots of $\log \widehat{\mu}_k(t)$, $k = 1, 2$ in Figure 3.1 are not only useful to assess the plausibility of a parametric model for each group, but also the multiplicative effect of the treatment variable. As these plots are roughly parallel, a multiplicative model is plausible. To this end we let $x_i = 1$ if subject i is in the treated arm and $x_i = 0$ if they are in the control arm and fit a multiplicative model of the form $\rho_i(t) = \rho_0(t) \exp(x_i'\beta)$ with $\rho_0(t) = \alpha_2 \alpha_1^{\alpha_2} t^{\alpha_2 - 1}$. This is done using the S-PLUS function `censorReg`; see Appendix C. The maximum likelihood estimates of α_1 and α_2 are 0.045 and 1.06, respectively. The estimated treatment effect (s.e.) is -0.823 (0.152). This gives an estimated relative rate of tumors of $\exp(-0.823) = 0.44$ for rats in the treated versus control arms (95% CI (0.33, 0.59), $p < 0.0001$), indicating a strong treatment effect. The estimated relative rate in the treated versus control rats for a piecewise-constant model (see Appendix C) is also 0.44 (95% CI (0.33, 0.59)), $p <$0.0001).

A semiparametric model can be fit using the `coxph` function in S-PLUS as described in Section 3.4 and Section C.1.3 of Appendix C. We obtain the maximum likelihood estimate $\widehat{\beta} = -0.816$ (s.e.=0.152) which gives an estimated relative rate 0.44 for rats in the treated versus control arms (95% CI (0.33, 0.60); p<0.0001). This is in close agreement with the previous estimates from the Weibull and the piecewise-constant model.

A semiparametric mixed Poisson model can also be fit using the `coxph` function with the `frailty(id)` option (Appendix C); this is advisable given the evidence of extra-Poisson variation from the test above. The estimate of the treatment coefficient arising from this model is similar to those from the previous analyses, but the standard error is larger and the associated confidence intervals wider because this model accommodates extra-Poisson variation. Specifically, the estimated relative rate is 0.44 (95% CI (0.29, 0.67); $p <$0.0001).

Finally, the semiparametric robust analysis can be carried out using the `cluster(id)` option for the `coxph` function. This analysis also gives an estimate $\widehat{\beta} = -.82$ but a standard error of .20, which leads to a similar confidence interval as the one obtained from the mixed Poisson model ($RR = 0.44$; 95% CI (0.29, 0.67)). This is coincidental, although it is often found that a mixed Poisson process analysis gives results close to the robust analysis for the regression coefficient.

Table 3.1 summarizes the results for the different models fitted. Note that the findings from these analyses are consistent with the results of the earlier two-sample tests in that we obtain strong evidence against the null hypotheses of no treatment difference.

Table 3.1. Estimates of treatment coefficient from several multiplicative rate function models for rat tumorgenicity data.

Model	Baseline Rate	EST.	S.E.	RR	95% CI	$p-$value
Poisson	Weibull	-0.82	0.15	0.44	(0.33, 0.59)	<0.0001
Poisson	Piecewise	-0.82	0.15	0.44	(0.33, 0.59)	<0.0001
Poisson	Nonparametric	-0.82	0.15	0.44	(0.33, 0.60)	<0.0001
Mixed Poisson	Weibull	-0.82	0.21	0.44	(0.29, 0.67)	<0.0001
Mixed Poisson	Nonparametric	-0.82	0.21	0.44	(0.29, 0.67)	<0.0001
Robust	Nonparametric	-0.82	0.20	0.44	(0.30, 0.65)	<0.0001

3.8.2 A Trial of Treatment for Herpes Simplex Virus

Romanowski et al. (2003) report on a 48-week multicenter open-label randomized two-period crossover trial of patients with a documented history of genital herpes simplex virus infection (type 1 or 2). A main objective was to compare the two types of therapy in terms of the occurrence of outbreaks of symptoms. Patients were randomized to two sequence groups. In sequence group A patients took 500 mg of valacyclovir once per day for 24 weeks with a view to suppressing outbreaks of symptoms (so-called suppressive therapy). When an outbreak occurred during suppressive therapy, the dose was increased to 500 mg twice a day for five days or until symptoms resolved. Period 2 of sequence group A was also 24 weeks in duration, and during this time patients took 500 mg doses of valacyclovir twice daily, but only for the treatment of outbreaks (so-called episodic therapy). Patients in sequence group B followed the episodic and suppressive regimens in periods 1 and 2, respectively. Plots of the Nelson–Aalen estimates of the mean functions for the number of outbreaks for patients in sequence groups A and B are given in Figure 3.4. The time variable t is time on study, in days.

Because this is a crossover design we define two time-dependent covariates. The first relates to the so-called "direct" effect of suppressive therapy. We set $x_{i1}(t) = 1$ if subject i is on suppressive therapy at time t and $x_{i1}(t) = 0$ otherwise. The actual day of crossover from the period 1 to period 2 regimen varies somewhat from subject to subject, so we let τ_i^c denote the time of crossover and τ_i the end of followup for subject i, $i = 1, \ldots, m$. In this case, $x_{i1}(t) = I(t \leq \tau_i^c)$ if subject i is in sequence group A and $x_{i1}(t) = I(\tau_i^c < t \leq \tau_i)$ if subject i is in sequence group B. To allow for a possible carry-over effect of suppressive therapy in period 2 for sequence group A, we let Δ represent a duration of time over which the residual effect of suppressive therapy in period 1 could affect the recurrence rate in period 2, where only episodic therapy is given; this is taken to be four weeks in these analyses. Thus we define $x_{i2}(t) = I(\tau_i^c < t \leq \tau_i^c + \Delta)$ if subject i is in sequence group A and $x_{i2}(t) = 0$ for $t \geq 0$ for patients in sequence group B. In addition to the variables related to treatment, we also examine the effects of fixed covariates for age, sex,

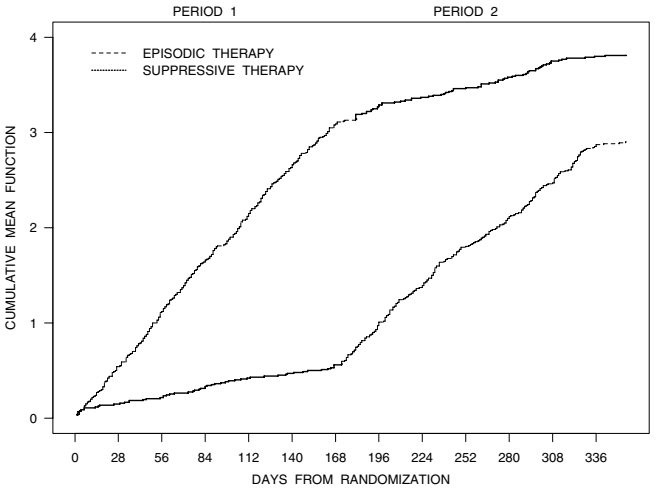


Fig. 3.4. Estimated mean functions for the cumulative expected number of outbreaks of symptoms for patients in sequence groups A and B from Romanowski et al. (2003).

virus type (I or II), and the number of recurrences in the year before entry to the study. We fitted semiparametric multiplicative regression models (3.2) with $g(x_i(t);\beta) = \exp(x_i'(t)\beta)$. The at-risk indicators $Y_i(t)$ were set to zero during short periods in which an individual was experiencing an outbreak, because they were not considered at risk for a new outbreak until the present one was resolved. We note that this poses a problem for the robust analysis below, because $Y_i(t)$ is not fully independent of the event (outbreak) process. However, the effect appears small in this setting. The results of model fits are reported in Table 3.2 for Poisson, negative binomial (mixed Poisson), and robust rate function models (see Sections 3.4.2, 3.5.3, and 3.6.3, respectively).

From Figure 3.4 it is clear that there is a dramatic effect of suppressive therapy for preventing the occurrence of outbreaks during period 1. Moreover, there appears to be some residual effect of the suppressive therapy received during period 1, in period 2. This can be seen by the fact that the slope of the mean function for patients on episodic therapy in period 2 is not quite as great as the slope of the mean function for patients on episodic therapy during period 1. In the absence of a carry-over effect, one would expect these mean functions to have the same slope. There is indeed an indication that the lower event rate for episodic treatment in period 2 persists well beyond four weeks.

Table 3.2 shows similar estimates and standard errors for treatment effects across the three models, but standard errors for the fixed covariates are

Table 3.2. Results of Poisson, mixed Poisson, and robust analysis of data from Romanowski et al. (2003).

	Poisson		Mixed Poisson		Robust	
Covariate†	EST.	S.E.	EST.	S.E.	EST.	S.E.
Direct Effect ($x_{i1}(t)$)	-1.578	0.102	-1.595	0.103	-1.578	0.108
Carryover Effect ($x_{i2}(t)$)	-0.285	0.249	-0.232	0.252	-0.285	0.211
Age	0.0006	0.004	0.0008	0.006	0.0006	0.007
Sex	-0.148	0.083	-0.159	0.126	-0.148	0.121
Virus Type	0.188	0.079	0.197	0.124	0.188	0.114
Previous Recurrences	0.074	0.018	0.074	0.028	0.074	0.025

† Age is measured in years, sex is male versus female, virus type is type II versus type I, and previous recurrences is the number of recurrences in the past year.

smaller under the Poisson model, suggesting that there is some extra-Poisson variation. The robust model reveals a substantial reduction in the rate of outbreaks with suppressive therapy ($RR = 0.21$, 95% CI (0.17, 0.26) $p < 0.0001$) and a greater rate of outbreaks for each additional number of recurrences in the previous year ($RR = 1.07$, 95% CI (1.02, 1.13), $p = 0.0034$). There is a slight suggestion of a trend towards a higher rate of outbreaks among individuals with infection from type II herpes simplex virus ($RR = 1.21$, 95% CI (0.97, 1.51), $p = 0.0986$) and no effect of age or sex is indicated. The estimate of the four-week carry-over effect ($RR = 0.75$) is consistent with the findings from the mean function plots but is not statistically significant.

3.8.3 Fitting and Prediction from a Software Debugging Model

Section 1.2.2 discussed a testing and debugging process for a large software system, in which different individuals tested the system and identified faults that were to be removed. Figure 1.3 shows a plot of $N(t)$, the cumulative number of faults detected, as a function of cumulative staff days of testing (t). The testing took place over 160 calendar days and involved 1336.7 staff days, during which 870 faults were detected.

It is of interest to fit a model to the process $\{N(t), 0 \leq t\}$, in order to predict the number of new faults that would be detected if testing were to continue. This is a key consideration in deciding when to cease testing or, perhaps, in deciding how many staff hours to devote to testing over a calendar time period. The introduction of new code or modifications of existing code may introduce new faults, and so in developing an intensity function model for the process $\{N(t), 0 \leq t\}$, we use the number of lines $C(t)$ of code changed or added up to time t as an internal time-varying covariate. The test data, given in Appendix D, show the values of $N(t)$ and $C(t)$ at time points that correspond to the ends of calendar days. Note that t considered here, however,

is cumulative staff days of testing; this is the most relevant time scale for model building and prediction.

We consider models where the fault detection intensity takes the form

$$\lambda(t|H(t)) = \alpha\theta e^{-\theta t} + \beta\theta e^{-\theta t} \int_0^t \frac{dC(u)}{e^{-\theta u}} \,. \tag{3.57}$$

Models of this type are discussed by Lawless (2006) and others, and have been shown to fit data of the type here. The data in Appendix D give the numbers of faults $N_j = N(t_{j-1}, t_j)$ detected in time intervals $[t_{j-1}, t_j), j = 1, \ldots, k$. For simplicity we assume that code changes all occur at the interval endpoints, and denote $C_j = dC(t_j)$ as the number of lines of code changed at t_j. Then, the intensity function (3.57) takes the form

$$\lambda(t|H(t)) = \theta e^{-\theta t} \left(\alpha + \beta \sum_{l=1}^{j-1} c_\ell e^{\theta t_l} \right) \qquad t_{j-1} \le t < t_j \,. \tag{3.58}$$

Let us fit the model (3.58) by maximum likelihood. The partial likelihood based on $N_1, \ldots, N_k$ (where $k = 160$ and $t_k = 1336.7$) is obtained from the terms involving $\Pr(N_\ell|H_\ell)$ in $\Pr(N_\ell, C_\ell, \ell = 1, \ldots, k)$, giving

$$L \propto \Pr(N_1) \prod_{l=2}^{k} \Pr(N_\ell|H_\ell) \,, \tag{3.59}$$

where $H_\ell = (N_1, \ldots, N_{\ell-1}; C_1, \ldots, C_{\ell-1})$. We note from (3.58) that for $j = 1, \ldots, k$ the distribution of N_j given H_j is Poisson with mean

$$\begin{aligned} E(N_j|H_j) = \mu_j &= \int_{t_{j-1}}^{t_j} \lambda(t|H(t))dt \\ &= (e^{-\theta t_{j-1}} - e^{-\theta t_j}) \, (\alpha + \beta \sum_{\ell=0}^{j-1} c_\ell e^{\theta t_\ell}) \,, \end{aligned} \tag{3.60}$$

where $C_0 = 0$, $t_0 = 0$.

By (3.59), the likelihood function is therefore a constant times

$$L(\alpha, \beta, \theta) = \prod_{j=1}^{k} e^{-\mu_j} \mu_j^{N_j} \,, \tag{3.61}$$

where μ_j is given by (3.60). This is readily maximized using general optimization software (Appendix B). We find from this that the maximum likelihood estimates are $\widehat{\alpha} = 19.1745$, $\widehat{\beta} = .003045$ and $\widehat{\theta} = .001704$. Figure 3.5 shows the cumulative fault counts shown in Figure 1.3, along with the fitted cumulative intensity or mean function given by $\widehat{\mu}(0) = 0$,

$$\widehat{\mu}(t_j) = \widehat{\mu}_1 + \ldots + \widehat{\mu}_j \qquad j = 1, \ldots, k$$

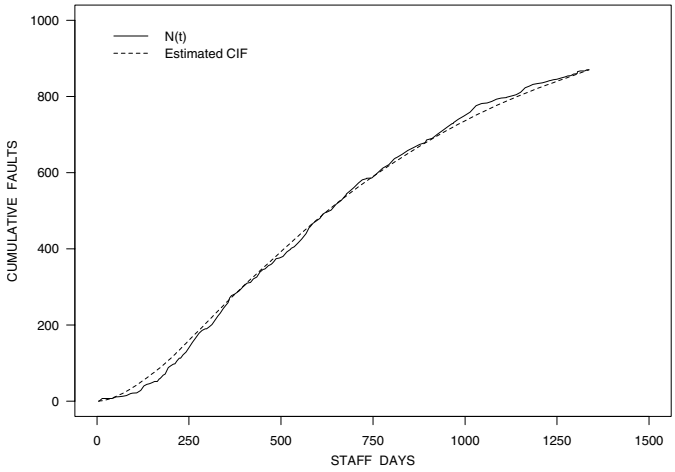


Fig. 3.5. Cumulative software faults detected during testing and cumulative intensity function from fitted model.

and linear interpolation between t_{j-1} and t_j $(j = 1, \ldots, k)$.

Figure 3.5 shows quite good agreement between the actual numbers of faults detected and the fitted cumulative mean function. A model with one additional parameter could, if desired, be fitted to deal with the slightly greater degree of curvature seen in the data than in the current model. Note that the situation here is different from others discussed in this chapter in two main respects: (i) there is only a single recurrent event process, and (ii) the model for the process has an intensity function involving an internal covariate (the number of lines of code modified). The model is, strictly speaking, not a Poisson process but the detection counts in successive time periods are, however, conditionally Poisson and the intensity (3.58) is conditionally (on the number of prior code lines changed) Poisson. We should properly think of the estimated "mean function" in Figure 3.5 as an estimated cumulative intensity, but increments $\widehat{\mu}(t_j) - \widehat{\mu}(t_{j-1})$ $(j = 1, \ldots, k$ with $t_0 = 0)$ are properly interpreted as the estimated expected number of new faults detected over the time interval $[t_{j-1}, t_j)$, given previous testing and code change history. The curve $\widehat{\mu}(t)$ can also be interpreted as a mean function if we consider the code changes C_j as part of the random process of fault detection and removal.

The previous discussion indicates that prediction of the number of faults detected beyond a current time point can be based on the Poisson process. In particular, consider prediction of the number of faults that would be found in the period $[t_k, \tau)$ beyond the final time $t_k = 1336.7$ and denote this $N_{k+1} = N(t_k, \tau)$. Prediction of this random variable is helpful in estimating

the number of faults remaining and in deciding whether it is worthwhile economically to prolong testing. Given the history $H(t_k)$, the future count N_{k+1} has a Poisson distribution under the model (3.58), with mean (and variance)

$$\mu_{k+1} = (e^{-\theta t_k} - e^{-\theta\tau})(\alpha + \beta \sum_{\ell=0}^{k} c_\ell e^{\theta t_l}) \,. \tag{3.62}$$

A naive "plug-in" method of prediction is to assume that N_{k+1} is Poisson with mean $\widehat{\mu}_{k+1}$, obtained by inserting the estimates of θ, α, and β in (3.62). For example, setting $\tau = \infty$ allows us to predict the number of new faults that would be detected if testing were prolonged indefinitely; this can be thought of as a surrogate for the number of remaining faults that would "appear" in repeated usage of the software. With $\tau = \infty$, we obtain $\widehat{\mu}_{k+1} = 196.8$; a two-sided 95% prediction interval is found as (170, 225), based on the .025 and .975 quantiles for the Poisson distribution with mean 196.8.

The Poisson assumption inherent in (3.58) can be assessed by noting that if the model is correct, the $N_j = N(t_{j-1}, t_j)$ are conditionally Poisson (μ_j), and are uncorrelated, because $E\{(N_j - \mu_j)(N_\ell - \mu_\ell)\} = 0$ for $j \neq \ell$. Model checks can be based on Poisson residuals $\widehat{z}_j = (N_j - \widehat{\mu}_j)/\widehat{\mu}_j^{1/2}$, which should be roughly uncorrelated with mean 0 and variance 1. These are analogous to martingale residuals discussed in Section 3.7.3; in the present setting the residuals do not suggest any major deficiencies in the model.

3.8.4 Comparing Warranty Claim Histories

Section 1.2.4 discussed data on warranty claims for 38,401 automobiles manufactured over a one-year period. It is useful with manufactured products to compare the warranty claims for units manufactured in different places or time periods, because that may identify quality or reliability variations and lead to opportunities for improvement. A good method of analysis is to compare mean curves for the numbers or total costs of claims across manufacturing periods, using the methodology of Section 3.6. In such analysis, the time variable t could be the age of the vehicle (the number of days since sale) or the distance driven (in miles or km). In the latter case, an interesting issue arises due to the fact that mileage accumulation data are not available for vehicles with no warranty claims, and we defer discussion of this to Chapter 7. We consider here the case where age is the time variable; this is relevant for assessing the number of claims made even though in North America there are both age and distance limits to warranty coverage. In the analysis below we ignore the issue of reporting delay that was mentioned in Section 1.2.4; its inclusion does not alter the results much. The cars in the database were mostly sold over a period of approximately one year, and data are included only for claims made up to 550 days after the first vehicle was sold, so followup times τ_i for the cars range from about 185 days to 550 days.

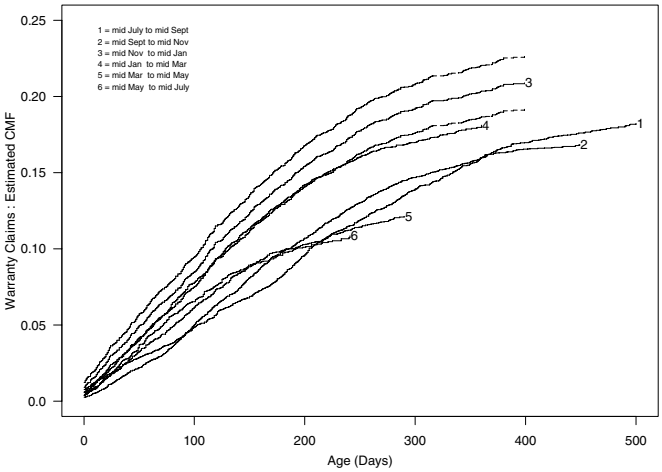


Fig. 3.6. Estimated cumulative mean functions (C.M.F.) for warranty claims, for vehicles grouped according to date of production.

Figure 3.6 shows estimated mean functions $\widehat{\mu}(t)$, given by the Nelson–Aalen estimate (3.17), for the cumulative number of claims for vehicles manufactured in six two-month time periods: mid-July to mid-September for Year 1 to mid-May to mid-July of Year 2. This one-year period is the time during which the vehicles of the given model year were manufactured. The estimates for the later production periods are truncated because the closing date for the database used here was before the most recently sold cars had completed their one-year, 12,000-mile warranties. The six curves all display a decreasing rate of claims; this is due to the nature of the car system represented here (unidentified for proprietary reasons) and also to the fact that with increasing age, more and more cars are no longer under warranty because they have exceeded the 12,000-mile distance limit. We note as well that some mean functions extend beyond the age of 365 days, because a few claims were allowed after the warranty had technically run out.

Figure 3.6 indicates that one manufacturing period (mid-November to mid-January) has a substantially higher level of claims that the others. It seems unlikely that the claim occurrence processes for the individual vehicles are identical Poisson processes, due to heterogeneity and other factors, and so the best way to compare the mean curves of the different production periods is to use the robust methodology of Sections 3.6 and 3.7. One approach is to test the equality of the mean functions by using indicator covariates and the multiplicative semiparametric model with rate function $\rho_i(t) = \rho_0(t) \exp(x_i'\beta)$. The mean functions do not appear exactly proportional to one another from

Figure 3.6, but this test still possesses good power to detect differences when the mean functions do not depart too radically from proportionality. Defining $x_{ij} = I(\text{vehicle } i \text{ is in production period } j)$ for $j = 2, \ldots, 6$, we obtain the following results using the S-PLUS function `coxph`. Estimates of regression coefficients for production periods $2, \ldots, 6$, with robust standard errors in brackets, were -0.025 (0.057), 0.213 (0.060), 0.052 (0.055), -0.357 (0.069) and -0.513 (0.075), respectively. A robust pseudo-score test of equality of the six rate functions, that is, of $H : \beta_j = 0$ $(j = 2, \ldots, 6)$ gives the observed value 152.64, which on χ^2_5 gives a very small $p-$value and indicates strong evidence against equality. A test that is tailored more to nonproportional mean functions could be developed as described at the end of Section 3.7.5. However, a test formulated after examination of Figure 3.6 could be criticised as biased, and most would prefer the conservative procedure above.

A final remark should be noted. In the warranty claim context here, the curves in Figure 3.6 in fact represent all of the data for the earlier production periods, and one might question whether tests or estimates that treat the data as realizations of a random process are relevant. We feel that they are, because in order to assess whether claim rates vary across factors such as production periods, it is conceptually useful to view the data as output from processes (which they are). In a formal test we see whether the differences seen could be due to random variation around identical underlying processes. An alternative would be to consider a permutation test (e.g. Freedman et al., 1989); that is more complex and leads to essentially the same result.

3.9 Bibliographic Notes

Poisson models have a long history of use in the analysis of data or recurrent events; see Section 2.7. There is a large literature in reliability (e.g. Ascher and Feingold, 1984; Rigdon and Basu, 2000), much of which deals with single processes without covariates. Lawless (1987a) considers parametric and semiparametric methods for regression analysis of Poisson process data. Nelson (1988, 1995) discusses nonparametric estimation of the mean function for general processes. Borgan and Hoem (1988) consider essentially the same idea. Aalen (1978) provides the properties of the Nelson–Aalen estimate (3.17) in the Poisson case. Andersen and Gill (1982) consider the semiparametric regression models of Section 3.4, and use counting process theory to derive asymptotic distributional results, generalizing methods which were introduced by Cox (1972a) for survival analysis. Andersen et al. (1993) give a comprehensive review of this area; see especially their Chapter 7. They also consider additive models; see their Section 7.4. McKeague and Sasieni (1994) give insightful discussion on additive models. Martinussen and Scheike (2006) discuss both multiplicative and additive models in detail, emphasizing survival with time-varying covariates and time-varying regression coefficients, $\beta(t)$. Many of their methods also apply to recurrent events. Therneau and Grambsch (2000)

consider examples involving multiplicative Cox models with recurrent events. Semiparametric time-transform models, analogous to the parametric models in Section 3.2.3, are considered by Lin et al. (1998) and Ghosh (2004). Methods based on piecewise-constant rate functions and multiplicative models are often found in demography and epidemiology; Holford (1980) and Andersen et al. (1993, p. 408) exemplify this approach. Andersen et al. (1993, Ch. 6 and Section 7.6) give a careful discussion of large sample theory for parametric models.

Lawless (1987a, b) discusses inference under negative binomial mixed Poisson models based on likelihood and quasi-likelihood estimation. Gaver and O'Muircheartaigh (1987) discuss mixed Poisson models in reliability. Klein (1992), Nielsen et al. (1992), Murphy (1995), Parner (1998), and Therneau and Grambsch (2000) consider random effects in the semiparametric multiplicative model. Aalen (1992) considers the compound Poisson distribution for random effects in Section 3.5.2. Grandell (1997) gives many results about mixed Poisson processes. Dean (1991) discusses estimating functions for random effects variance parameters in mixed Poisson models. Nielsen and Dean (2005) employ regression splines to model baseline rate functions in Poisson, mixed Poisson, and robust rate function models, thus providing an alternative to the kernel estimation methods of Section 3.4.1.

The robust methods for parametric or semiparametric regression analysis of rate and mean functions in Section 3.6 were developed by Lawless and Nadeau (1995), and Lin et al. (2000) provide rigorous asymptotics. Pepe and Cai (1993) consider similar ideas for estimation of rate functions. Nadeau and Lawless (1998) discuss questions of efficiency. Scheike (2002) considers additive, and Scheike and Zhang (2003) consider multiplicative-additive, rate function models; see also Martinussen and Scheike (2006).

Methods for assessing plausibility of assumptions for counting processes can be based on techniques developed for survival analysis and good reviews are given by Andersen et al. (1993), Grambsch and Therneau (1994), Therneau and Grambsch (2000), and Martinussen and Scheike (2006). Various tests based on martingale residual processes have been proposed but many are awkward to deal with and interpret. We prefer the techniques in Section 7.3 and other model expansion techniques. Pena (1998) gives a treatment of score tests; see also Hjort (1990a). Fisher (1950) develops tests for extra-Poisson variation. Testing for extra-Poisson variation through the introduction of random effects can be based on likelihood ratio tests, but score tests are often appealing as in Liang (1987), Dean and Lawless (1989), and Dean (1992). Ng and Cook (1999a) consider small sample corrections in the context of weakly parametric models.

A wide class of two-sample tests for rate functions may be borrowed from the survival analysis literature including unweighted "log-rank" tests, weighted tests, and two degree of freedom tests; see Cook and Lawless (1991) and Cook et al. (1996).

3.10 Problems and Supplements

3.1. Consider a Poisson process regression model with fixed covariates where $\rho_i(s) = \rho_0(s;\alpha)\exp(x_i'\beta)$. Let $\mu_0(t) = \int_0^t \rho_0(u)du$ and $\mu_i(t) = \int_0^t \rho_i(u)du$, $i = 1, 2, \ldots, m$.

a. Show that under completely independent censoring the likelihood contribution to (3.2) from subject i can be decomposed as

$$L_i(\theta) = L_{i1}(\alpha)L_{i2}(\theta)\,,$$

where $\theta = (\alpha', \beta')'$, $L_{i1}(\alpha) = \prod_{j=1}^{n_i} \rho_0(t_{ij})/\mu_0(\tau_i)$ arises from the conditional distribution of the event times given the total number of events and $L_{i2}(\theta) = (\mu_i(\tau_i))^{n_i}\exp(-\mu_i(\tau_i))$ is based on the likelihood contribution from the event counts.

b. Suppose the baseline rate function $(\rho_0(t))$ was known. Obtain the score vector and information matrix for β based on $L_{i2}(\beta)$.

c. Show that analyses based only on event counts n_i are fully efficient under time-homogeneous Poisson models (i.e. when $\rho(t) = \alpha$) and more generally provided $\tau_i = \tau$, $i = 1, 2, \ldots, m$.

[Section 3.1; Lawless, 1987a; Dean and Balshaw, 1997]

3.2. Consider the tumorgenicity data introduced in Section 1.2.1, given in Table 1.1 and analyzed in Section 3.8.1. Fit a parametric Poisson regression model to these data with a baseline rate function having the form of a hazard function for a log-logistic distribution (Lawless, 2003a, Section 1.3.4).

[Section 3.2]

3.3. Consider a regression model $\rho_i(t) = \rho_0(t;\alpha)\exp(x_i'\beta)$ in which the baseline rate function has a Weibull form (i.e. $\rho_0(t;\alpha) = \alpha t^{\alpha-1}$), and $x_i = (1, x_{i1}, \ldots, x_{ip})'$. Suppose this model is fit in S-PLUS using `censorReg` and interest lies in constructing a confidence interval for the mean function for a population with $x = x_0$. Describe how to construct a 95% confidence interval for the mean function $t^\alpha\exp(x_0'\beta)$ at time t, based on the information-based asymptotic covariance matrix in Section 3.2.1.

[Section 3.2]

3.4. a. Obtain (3.14) by substituting the profile maximum likelihood estimates given by (3.13) into the full likelihood in (3.12).

b. Argue that as $m \to \infty$ the profile likelihood in (3.14) approaches the partial likelihood in (3.25).

[Section 3.3]

3.5. Derive a score test for trend where the null model is time-homogeneous and the alternative is the piecewise-constant model of Section 3.3. The null hypothesis in this case is $H_0 : \alpha_k = \alpha,\ k = 1, \ldots, K$.

[Sections 3.3 and 3.7]

3.6. Show that differentiation of $\log L_1(\beta)$ where $L_1(\beta)$ is given by (3.25), leads to the expression for the profile score function in (3.21).

[Section 3.4]

3.7. Interest frequently lies in carrying out formal tests that the rate functions in two groups are the same. Suppose x_i is a treatment indicator such that $x_i = 0$ if subject i is in the control group and $x_i = 1$ if in a treated group, and consider the multiplicative model $\rho_i(s) = \rho_0(s)\exp(\beta x_i)$. The null hypothesis is then given by $H_0 : \beta = 0$. Show that if $U_\beta(\beta)$ given by (3.21) is evaluated at $\beta = 0$, then with suitable changes to the notation, this can equivalently be written as (3.52), with $a(u) = 1$ in (3.53).

[Sections 3.4 and 3.7]

3.8. Consider a parametric mixed Poisson regression model. Obtain $\partial U_\alpha(\theta)/\partial\phi$ and $\partial U_\beta(\theta)/\partial\phi$ and show that $E(-\partial U_\alpha(\theta)/\partial\phi) = E(-\partial U_\beta(\theta)/\partial\phi) = 0$. This finding makes a Fisher-scoring algorithm (a Newton–Raphson type algorithm based on $\mathcal{I}(\theta) = E(I(\theta))$ instead of $I(\theta)$) somewhat more convenient for estimation of α and β. Extend the methods for piecewise-constant rate functions in Section 3.3 to the mixed Poisson setting.

[Section 3.5]

3.9. Consider a mixed Poisson regression model as described in Section 3.5.

a. Show that under independent censoring $E(u_i|H_i(t)) = E(u_i|N_i(t^-))$.
b. Derive $E(u_i|H_i(t))$ when u_i has an inverse Gaussian distribution with density

$$g(u;\phi) = \frac{1}{\sqrt{2\pi u^3\phi}}\exp(-(u-1)^2/(2u\phi))\,,$$

where $u > 0$ and $\phi > 0$.

[Section 3.5]

3.10. Let $V_1, V_2, \ldots$ be independent and identically distributed gamma random variables with shape and inverse scale parameters γ_1 and γ_2 and density given by

$$f(v;\gamma) = \frac{\gamma_2^{\gamma_1} v^{\gamma_1 - 1}\exp(-v\gamma_2)}{\Gamma(\gamma_1)}\,.$$

If K is a Poisson random variable with mean κ, then define $U = V_1 + V_2 + \cdots + V_K$ if $K > 0$ and $U = 0$ otherwise.

a. Derive the density for the positive component of U, given in Section 3.5.2.

b. Show that the Laplace transform of U is

$$\mathcal{L}(s) = \exp(-\kappa[1 - (1 + s/\gamma_2)^{-\gamma_1}]) .$$

c. Show that $E(U_i) = \kappa\gamma_1/\gamma_2$ and $\text{var}(U_i) = \kappa\gamma_1(\gamma_1 + 1)/\gamma_2^2$.

[Aalen, 1992; Section 3.5.2]

3.11. In clinical studies subjects are frequently observed for a baseline period of observation before being randomized to treatment groups. Let N_{i1} denote a count observed for subject i over the baseline observation period and $\{N_{i2}(t), 0 \le t\}$ denote the process generating the events after randomization. Let n_{i1} denote the observed baseline count and $t_{i1}, \ldots, t_{in_{i2}}$ denote the times of the n_{i2} events observed over the followup period $[0, \tau_i]$ for subject i, $i = 1, \ldots, m$. Suppose that $N_{i1}|u_i \sim$ Poisson with mean $u_i\mu_1$ and $\{N_{i2}(t), 0 \le t\}|u_i$ is a Poisson process with conditional rate given by $u_i\rho_0(t)\exp(x_i\beta)$, $0 \le t$.

a. Obtain the conditional probability of n_{i2} events at $t_{i1} < \cdots < t_{in_{i2}}$ over $[0, \tau_i]$ given $n_i = n_{i1} + n_{i2}$.
b. Show that if $\rho_0(t) = \alpha$ then a fully efficient estimate of β can be obtained by logistic regression.

[Section 3.5]

3.12. Consider a K group problem. Let the rate function for group k be $\rho_k(s) = \rho_0(s)\exp(x'\beta)$, where $x = (x_2, \ldots, x_K)'$ with $x_j = 1$ if an individual is in group j and 0 otherwise, $j = 2, \ldots, K$; $k = 1, \ldots, K$.

a. Derive the score components for a test of the null hypothesis $H_0 : \rho_k(s) = \rho(s)$, $k = 1, \ldots, K$ under the assumption that the events are generated by Poisson processes.
b. Derive an information-based covariance matrix with which to construct a $K - 1$ degree of freedom test.
c. Derive a robust covariance matrix on which one could base a test of the null hypothesis.

[Sections 3.6 and 3.7]

3.13. Consider the family of rate functions

$$\rho(t; \alpha_1, \alpha^*) = \exp\{\alpha_1 + \alpha^* g(t)\} \qquad t \ge 0 \tag{3.63}$$

where $g(t)$ is a specified function. Testing $H_0 : \alpha^* = 0$ within this family gives a test for trend, as discussed in Section 3.7.1.

a. Assuming that individual process i ($i = 1, \ldots, m$) is observed over $[0, \tau_i]$, show that the Poisson process estimating functions (3.4) based on (3.63) are

$$U_1(\alpha_1, \alpha^*) = n_\cdot - e^{\alpha_1} \sum_{i=1}^{m} \int_0^{\tau_i} e^{\alpha^* g(t)} dt$$
$$U_2(\alpha_1, \alpha^*) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} g(t_{ij}) - e^{\alpha_1} \int_0^{\tau_i} g(t) e^{\alpha^* g(t)} dt\,.$$

As noted in Section 3.6.2, these score functions are unbiased in general, provided the τ_i are independent of the event process.

b. Let $\tilde{U}_2(0) = U_2(\tilde{\alpha}_1(0), 0)$, where $\tilde{\alpha}_1(0)$ is the estimate of α_1 obtained by solving $U_1(\alpha_1, 0) = 0$. Show that

$$\tilde{U}_2(0) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} g(t_{ij}) - \frac{n_\cdot}{\tau_\cdot} \sum_{i=1}^{m} \int_0^{\tau_i} g(t) dt\,.$$

Note that $\tilde{U}_2(0)$ equals (3.42) in the special case where $g(t) = t$.

c. Show that $\tilde{U}_2(0)$ can be written as $\sum_{i=1}^m \tilde{U}_{2i}(0)$, where (with $\rho = \exp(\alpha_1)$)

$$\tilde{U}_{2i}(0) = \int_0^{\tau_i} (g(t) - G_\cdot/\tau_\cdot)(dN_i(t) - \rho dt)\,,$$

where $G(\tau) = \int_0^\tau g(u)du$, $G_\cdot = \sum_{i=1}^m G(\tau_i)$ and $\tau_\cdot = \sum_{i=1}^m \tau_i$. Thus obtain $\widehat{\text{asvar}}\{\tilde{U}_2(0)\}$ under H_0 as $\sum_{i=1}^m [\tilde{U}_{2i}(0)]^2$ with ρ estimated by $n_\cdot/\tau_\cdot$.

d. Consider the family of rate functions $\rho_i(t) = \exp(\alpha_i + \alpha^* t)$ leading to (3.45) and let $\alpha = (\alpha_1, \ldots, \alpha_m)'$. Show that $\tilde{U}_2(0) = U_2(\tilde{\alpha}(0), 0)$ gives (3.45) and that $\tilde{U}_2(0)$ can be written as $\sum_{i=1}^m \tilde{U}_{2i}(0)$, where

$$\tilde{U}_{2i}(0) = \int_0^{\tau_i} (t - \tau_i/2) dN_i(t) = n_i(\bar{t}_i - \tau_i/2)\,.$$

Thus obtain the robust variance estimate $\widehat{\text{asvar}}\{\tilde{U}_2(0)\} = \sum_{i=1}^m [\tilde{U}_{2i}(0)]^2$.

[Section 3.7.1]

3.14. Describe four ways one could test for trend in the tumorgenicity data of Section 3.8.1.

[Section 3.7]

3.15. Consider a mixed Poisson regression model given by (3.27) with $\rho_i(t) = \rho_0(t) \exp(x_i'\beta)$ and $\rho_0(t) = \exp(\alpha_1 + \alpha^* t)$. Derive the conditional distribution of $(t_{i1}, \ldots, t_{in_i})$ given τ_i and n_i, and use this to develop a score test for trend (i.e. $H_0 : \alpha^* = 0$).

[Section 3.7]

3.16. a. Describe three ways in which one could test for extra-Poisson variation.
b. Carry out these tests for the control arm of the tumorgenicity data of Section 3.8.1 and discuss your findings.

[Section 3.7]

3.17. Nelson (1995) and Lawless and Nadeau (1995) discuss data on the times (in days of service) at which diesel engine valve seats were replaced on 41 locomotives. The data are shown in Table 3.3; the lengths of time in service are similar except for unit 28, which has a substantially shorter time. Each diesel engine had 16 valves, but the data do not indicate which ones were replaced, so we take the 41 engines as the units and consider the process of the number of replacements $N(t)$ up to t days in service, for each unit.

a. Fit Poisson process models with parametric rate functions $\rho(t;\alpha,\beta) = \exp(\alpha+\beta t)$ and $\rho(t;\alpha,\beta) = \alpha\beta(t\alpha)^{\beta-1}$ to the data. Plot the data first to see why these models might be reasonable. Assess model fit using diagnostics in Section 3.7.3.
b. Develop pointwise confidence bands for the mean number of replacements $\mu(t) = E\{N(t)\}$, up to t days in service, based on the parametric models. Also obtain nonparametric confidence limits based on the methodology of Section 3.4.1.
c. Check for extra-Poisson variation by fitting processes with gamma random effects, as described in Section 3.5. Fit parametric models and also fit a semiparametric model by using the S-PLUS or R function `coxph` with the `frailty` option. There are no covariates here so use ~1 in the formula specification.
d. Unit 28 is potentially influential because it has three replacements and a considerably shorter time in service than other units. Assess its effect by refitting models with this unit excluded.
e. Consider how you could predict the number of replacement valves needed for a future time period.

[Sections 3.2, 3.5, 3.7]

3.18. Lee (1980) gave the data in Table 3.4, which shows the times (in thousands of hours of operation) of unscheduled maintenance events on the number 4 engine of the submarine U.S.S. Grampus. Plot the data using a Nelson–Aalen estimate, which is here just the cumulative number of events. Fit Poisson process models to the data and assess whether a homogeneous Poisson process provides a satisfactory description. Consider tests of a homogeneous Poisson process, including the score test of Section 3.7.1, and diagnostic checks based on residuals introduced in Section 3.7.3. For a nonhomogeneous Poisson process that you select, carry out similar diagnostic checks.

[Section 3.7]

Table 3.3. Times in days of valve seat replacements in 41 diesel locomotives.

Unit	Replacement times	Service Time	Unit	Replacement times	Service Time
1		761	22		593
2		759	23	573	589
3	98	667	24	165, 408, 604	606
4	326, 653, 653	667	25	249	594
5		665	26	344, 497	613
6	84	667	27	265, 586	595
7	87	663	28	166, 206, 348	389
8	646	653	29		601
9		651	30	410, 581	601
10	92	653	31		611
11	258, 328, 377, 621	650	32		608
12	61, 539	648	33		587
13	254, 276, 298, 640	644	34	367	603
14	76, 538	642	35	202, 563, 570	585
15	635	641	36		587
16	349, 404, 561	649	37		578
17		631	38		578
18		596	39		586
19	120, 479	614	40		585
20	323, 447	582	41		582
21	139, 139	589			

Table 3.4. Times (in thousands of hours of operation) of unscheduled maintenance events for submarine U.S.S. Grampus number 4 engine.

0.860	1.258	1.317	1.442	1.897	2.011	2.122	2.439
3.203	3.298	3.902	3.910	4.000	4.247	4.411	4.456
4.517	4.899	4.910	5.676	5.755	6.137	6.221	6.311
6.613	6.975	7.335	8.158	8.498	8.690	9.042	9.330
9.394	9.426	9.872	10.191	11.511	11.575	12.100	12.126
12.368	12.681	12.795	13.399	13.668	13.780	13.877	14.007
14.028	14.035	14.173	14.173	14.449	14.587	14.610	15.070
16.000							

4
Analysis of Gap Times

4.1 Renewal Processes and Related Methods of Analysis

Modeling and analysis of the gaps or waiting times between successive events is attractive in certain settings. An important one is where an individual or system is restored to a similar physical state after each event; for example, following a repair or equipment replacement a system might be returned to an "as new" state.

Renewal processes were introduced in Section 2.3. They have the property that the gaps between successive events are independent and identically distributed. This is a very strong condition and such processes are mainly useful when an individual is physically renewed in some sense after each event, or when events are due to external processes that are regenerative. However, by extending renewal processes in various ways we can obtain other models that are more widely applicable to the study of gap times.

We begin by describing methodology based on renewal processes, and then extend the methodology in subsequent sections. We assume for the time being that individual i is observed over the time interval $[0, \tau_i]$ and that $t = 0$ corresponds to the start of the event process. Situations where this is not the case are discussed in Section 4.5. The event intensity function is of the form (2.37); allowing for fixed covariates x_i means that the gaps W_{ij} between events have hazard function $h(w|x_i)$. If n_i events are observed at times $0 < t_{i1} < \cdots < t_{in_i} \leq \tau_i$, let $w_{ij} = t_{ij} - t_{i,j-1}$ $(j = 1, \ldots, n_i)$ and $w_{i,n_i+1} = \tau_i - t_{in_i}$, where $t_{i0} = 0$. These are the observed gap times for individual i with the final time being possibly censored. The likelihood function from m independent individuals is of the form

$$L = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} h(w_{ij}|x_i) \exp(-H(w_{ij}|x_i)) \right\} \exp\left(-H(w_{i,n_i+1}|x_i)\right) , \qquad (4.1)$$

where $H(w|x) = \int_0^w h(u|x)du$ is the cumulative hazard function for W_{ij}, given x_i. If τ_i is a prespecified followup time or a time that is determined independently of the event process, then (4.1) is obtainable as $\Pr\{W_{i1} = w_{i1}, \ldots, W_{i,n_i} = w_{i,n_i}, W_{i,n_i+1} > w_{i,n_i+1}\}$, where for convenience we use $\Pr(\cdot)$ to denote either a probability or probability density. As discussed in Section 2.6, (4.1) is also valid for inference in settings where the τ_i are randomly determined according to a scheme that may involve prior events; it is a partial likelihood in that case.

Let $f(w|x) = h(w|x)\exp\{-H(w|x)\}$ and $S(w|x) = \exp\{-H(w|x)\}$ denote the density and survivor functions for W_{ij} given x. In terms of these functions the likelihood (4.1) is

$$L = \prod_{i=1}^{m} \prod_{j=1}^{n_i} f(w_{ij}|x_i) \cdot S(w_{i,n_i+1}|x_i)\,, \tag{4.2}$$

which is the familiar likelihood function for a random sample involving failure times w_{ij} $(j = 1, \ldots, n_i)$ and right censoring times w_{i,n_i+1} for $i = 1, \ldots, m$. If $w_{i,n_i+1} = 0$, that is, if observation terminates after the n_ith event, the term $S(w_{i,n_i+1}|x_i)$ in (4.2) disappears. Standard survival analysis methods and software can therefore be used for model fitting and inference.

We next summarize a few key survival analysis methods. For detailed treatments see Lawless (2003a) or other books on lifetime or survival data.

(i) Parametric lifetime distributions such as Weibull, log-normal, and log-logistic distributions can be used. When fixed covariates are present, corresponding accelerated failure time models can be used. An accelerated failure time (AFT) model for a response time W is one for which $Y = \log W$ has a location-scale distribution of the form

$$Y = \beta_0 + x'\beta + \sigma\epsilon\,,$$

where $x = (x_1, \ldots, x_k)'$ is a covariate vector, $\beta = (\beta_1, \ldots \beta_k)'$ is a vector of regression coefficients, $\sigma > 0$ is a scale parameter, and ϵ is a random variable whose distribution is independent of x. Common AFT models are those for which ϵ has standard extreme value, logistic, and normal distributions, respectively; this corresponds to $T = \exp(Y)$ having Weibull, log-logistic, and log-normal distributions. Software for AFT models is widely available; in S-PLUS the functions `survReg` or `censorReg` can be used and in R the function is `survreg`.

This model readily handles cases where the covariate values are fixed within gaps but vary across gaps (i.e., the covariate corresponding to W_{ij} is x_{ij}). In this case the AFT model takes $Y_{ij} = \log W_{ij}$ and the distribution of Y_{ij} given x_{ij} can be represented as

$$Y_{ij} = \beta_0 + x'_{ij}\beta + \sigma\epsilon_{ij}\,, \tag{4.3}$$

where the ϵ_{ij} terms are i.i.d. random variables. Covariates that vary within gaps are harder to handle with AFT models, but are easily dealt with by the Cox model in (iii) below.

(ii) The Kaplan–Meier nonparametric estimate of $S(w)$ and the Nelson–Aalen estimate of $H(w)$ can be used when there are no covariates. These estimates are given by (e.g. Lawless 2003a, Section 3.2)

$$\widehat{S}_{KM}(w) = \prod_{\ell: w_\ell^* \le w} \left(1 - \frac{d_\ell}{n_\ell}\right) \tag{4.4}$$

$$\widehat{H}_{NA}(w) = \sum_{\ell: w_\ell^* \le w} \frac{d_\ell}{n_\ell}\,, \tag{4.5}$$

where the w_ℓ^* are the distinct values among the w_{ij} $(i = 1, \ldots, m;\ j = 1, \ldots, n_i)$ and where

$$d_\ell = \sum_{i=1}^{m} \sum_{j=1}^{n_i} I(w_{ij} = w_\ell^*) \quad \text{and} \quad n_\ell = \sum_{i=1}^{m} \sum_{j=1}^{n_i+1} I(w_{ij} \ge w_\ell^*)\,. \tag{4.6}$$

Variance estimates for (4.4) and (4.5) are

$$\widehat{\mathrm{var}}(\widehat{S}_{KM}(w)) = \widehat{S}_{KM}(w)^2 \sum_{\ell: w_\ell^* \le w} \frac{d_\ell}{n_\ell(n_\ell - d_\ell)}$$

$$\widehat{\mathrm{var}}(\widehat{H}_{NA}(w)) = \sum_{\ell: w_\ell^* \le w} \frac{d_\ell}{n_\ell^2}\,.$$

Kaplan–Meier and Nelson–Aalen estimates may be obtained from the S-PLUS functions `survfit` and `kaplanMeier`, and the R function `survfit`.

(iii) The Cox semiparametric multiplicative hazards model in which the hazard function for W_{ij} given x_{ij} is of the form

$$h(w|x_{ij}) = h_0(w) \exp(x_{ij}'\beta) \tag{4.7}$$

can be fitted by the usual partial likelihood method (e.g. Lawless 2003a, Ch.7). This model also deals with time-varying covariates, in which case x_{ij} is replaced by $x_{ij}(t)$ in (4.7), but care is needed to "align" the times t for individuals in the risk set for w_{ij}; see Section 4.2.4. For the Cox model, when covariates are fixed across intervals between events, β is estimated by maximizing the likelihood function

$$L(\beta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left\{ \frac{\exp(x_{ij}'\beta)}{\sum_{l=1}^{m} \sum_{k=1}^{n_\ell+1} I(w_{lk} \ge w_{ij}) \exp(x_{lk}'\beta)} \right\}. \tag{4.8}$$

Once $\widehat{\beta}$ is obtained from (4.8), $H_0(w) = \int_o^w h_0(u)du$ is estimated by

$$\widehat{H}_0(w) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \frac{I(w_{ij} \le w)}{\sum_{l=1}^{m} \sum_{k=1}^{n_l+1} I(w_{lk} \ge w_{ij}) \exp(x'_{lk}\widehat{\beta})} \right\}. \qquad (4.9)$$

The Cox model is handled by the functions `coxph` and `cox.zph` in S-PLUS and R.

It should be noted that the likelihood function (4.8) for β does not come from the usual partial likelihood development used for lifetime data (e.g. Lawless 2003a, Section 7.1). That approach does not work here because individuals can contribute multiple gap times w_{ij} and so the likelihood and score functions are not expressible as processes to which martingale-type arguments can be applied. However, it has been shown (e.g. Dabrowska et al. 1994) that estimation of β can be based on (4.8) in the renewal process setting, and $L(\beta)$ can be used in the standard way to provide asymptotic variance estimates, likelihood ratio tests, and so on. Problem 4.3 outlines a way of validating (4.8) and (4.9).

As remarked, the assumption that gap times W_{ij} are independent and identically distributed when no covariates are present is very strong, and it is important to consider diagnostic checks in any given situation. Failures to account for association among gap times may lead to substantial bias in dealing with gap times after the first; we consider this issue in Section 4.4. The same applies to regression models such as (4.3) and (4.7), where covariates x_{ij} are present; the assumption that the W_{ij} are conditionally independent is critical. An important way of model checking is by fitting models that include renewal processes as a special case. This approach is considered in Sections 4.2 and 4.3, where extensions of renewal processes are presented. The independence assumption can also be checked informally when no covariates are present by, for example, looking at scatter plots of successive gap times, $w_{i,j}$ versus $w_{i,j+1}$ $(j = 1, 2, \ldots)$ within individuals; there should be an absence of trend if the renewal assumption is valid. If the gap times are independent, then informal checks on the assumption of a common distribution can also be made by comparing separate empirical distributions for the different gaps. These checks were illustrated in Section 2.3.2.

When covariates are present it is best to fit a variety of models in order to assess gap time independence or to compare distributions of different gap times. Indeed, for reasons discussed in Section 2.3, the independence of successive gap times should be assessed this way as well as through informal plots, even when no covariates are present; this is illustrated in Section 4.3.

Example 4.1.1 Bowel Motility Cycles

Section 2.3.2 discussed data from a study on muscular activity (motility) of the small bowel in humans. The observations are the gap times w_{ij} between successive bowel activity cycles for 19 subjects. We continue to assume, as in Section 2.3.2, that the duration of a "fed state" bowel activity period, which precedes the "fasting state" cycles discussed here, is not associated with the cycle durations.

The pairs of successive gap times $(w_{ij}, w_{i,j+1})$ were plotted in Figure 2.1, and show no evidence of association. A "lag 2" plot of points $(w_{ij}, w_{i,j+2})$ gives a similar result, and it seems reasonable to assume that the gap times within a subject are independent. We consider formal tests based on extended models in Sections 4.3.1 and 4.3.2, and they likewise show no evidence against the independence assumption.

Separate Kaplan–Meier estimates (4.4) for the gap time survivor function for first, second, and third and higher gaps were also shown in Figure 2.2 of Section 2.3.2. This provides a check on the assumption that successive gap times are identically distributed. Figure 2.2 suggests that the distribution for first gaps is somewhat different than the distributions for later gaps; second and subsequent gaps tend to be shorter.

Formal tests of the assumption of a common distribution can be carried out. For example, under the independence assumption we could test the equality of the distribution of first and subsequent gap times by using a log-rank test or another nonparametric test (e.g. Lawless 2003a, Section 8.1). A log-rank test of the equality of the first and second gap times, as implemented in the S-PLUS function `survdiff`, for example, gives a $p-$value of 0.12, thus indicating only very mild evidence against the hypothesis of equality.

An alternative approach is to use parametric models. Exploratory probability plots suggest log-normal models (i.e. normal models for $Y_{ij} = \log W_{ij}$) may be appropriate. Models with independent Y_{ij} and, respectively,

(i) $Y_{ij} \sim N(\mu, \sigma^2)$ for all i, j ;
(ii) $Y_{i1} \sim N(\mu_1, \sigma_1^2)$, $Y_{ij} \sim N(\mu_2, \sigma_2^2)$ for $j = 2, 3, \ldots$;

give the following results,

(i) $\widehat{\mu} = 4.51$, $\widehat{\sigma} = 0.55$, $l_{\max} = \log L(\widehat{\mu}, \widehat{\sigma}) = -73.12$.
(ii) $\widehat{\mu}_1 = 4.75$, $\widehat{\sigma}_1 = 0.40$, $\widehat{\mu}_2 = 4.45$, $\widehat{\sigma}_2 = 0.56$, $l_{\max} = \log L(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1, \widehat{\sigma}_2) = -69.12$.

A likelihood ratio (LR) test of model (i) versus model (ii) gives the observed LR statistic $D = 2 \log L(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1, \widehat{\sigma}_2) - 2 \log L(\widehat{\mu}, \widehat{\sigma}) = 8.0$. Using the large sample result (see Appendix A) that the distribution of D is asymptotically χ_2^2 under the null hypothesis of model (i) (because there are 2 parameters in (i) and 4 in (ii)), we get an approximate $p-$value of $\Pr(\chi_2^2 \geq 8.0) = .018$, thus indicating fairly strong evidence against the assumption that first and subsequent gaps have the same distribution.

4.2 Extensions of Renewal Models

The assumption of independent gap times is untenable in most situations. More general models can be formulated through the sequence of conditional distributions

$$F_j(w|x_{ij}, w_i^{(j-1)}) = \Pr(W_{ij} \le w|x_{ij}, w_i^{(j-1)}) \qquad j = 1, 2, \dots ,$$

where $w_i^{(j-1)} = (w_{i1}, \dots , w_{i,j-1})'$ and x_{ij} is a vector of covariates associated with the jth gap time for individual i. This format allows various types of dependence on previous event history to be considered, including elapsed time $w_{i1} + \cdots + w_{i,j-1}$ up to the $(j-1)$st event. Models can also be formulated to deal with covariates that vary within the gaps between events. In this section we consider several models of this form, which naturally include the renewal models of the preceding section as special cases. These models are used in applications in Section 4.3.

4.2.1 Conditional Analysis of Successive Gap Times

Statistical analysis under the general framework above can be based on regression models for survival times or durations. The two dominant families of such models (Lawless 2003a, Chs. 6, 7) are the proportional or multiplicative hazards models (4.7) and the accelerated failure time models (4.3), which were introduced in Section 4.1. We consider modeling and analysis under each family and then provide an illustration.

For parametric models the likelihood function from a set of m independent processes is an extension of (4.2):

$$L = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} f_j(w_{ij}|z_{ij}) \right\} S_{n_i+1}(w_{i,n_i+1}|z_{i,n_i+1}) , \tag{4.10}$$

where z_{ij} is a vector that models dependence of W_{ij} on x_{ij} and $w_i^{(j-1)}$, and $f_j(w|z_{ij})$ and $S_j(w|z_{ij})$ are the density and survivor functions for W_{ij}, given z_{ij}. If z_{ij} does not depend on $w_i^{(j-1)}$, the model is a renewal process with independent but not identically distributed gap times.

A multiplicative hazards model takes the hazard function for W_{ij} given x_{ij} and $w_i^{(j-1)}$ to be of the form

$$h_{ij}(w) = h_{0j}(w) \exp(z'_{ij}\beta_j) \qquad j = 1, 2, \dots , \tag{4.11}$$

where for convenience we write $h_{ij}(w)$ for $h_j(w|x_{ij}, w_i^{(j-1)})$. It is possible to include time-varying covariates in the hazard functions (4.11) but we defer a discussion of this until Section 4.2.4. It is also possible to constrain the baseline hazard functions $h_{0j}(w)$ or the regression parameters β_j to be identical.

The most common form of analysis based on (4.11) is the semiparametric analysis of Section 4.1 (iii), in which the baseline hazard functions $h_{0j}(w)$ are treated nonparametrically. In the case where the $h_{0j}(w)$ and β_j terms in (4.11) are distinct, this leads to separate estimation for each gap $j = 1, 2, \ldots$, based on functions similar to (4.8) and (4.9). That is, β_j is obtained by maximizing the likelihood function

$$L_j(\beta_j) = \prod_{i=1}^{m} \left\{ \frac{\exp(z'_{ij}\beta_j)}{\sum_{l=1}^{m} \delta_{lj} I(w_{lj} \geq w_{ij}) \exp(z'_{\ell j}\beta_j)} \right\}^{\delta_{i,j+1}}, \tag{4.12}$$

and $H_{0j}(w) = \int_0^w h_{0j}(u)du$ is estimated by

$$\widehat{H}_{0j}(w) = \sum_{i=1}^{m} \left\{ \frac{\delta_{i,j+1} I(w_{ij} \leq w)}{\sum_{l=1}^{m} \delta_{lj} I(w_{lj} \geq w_{ij}) \exp(z'_{lj}\widehat{\beta}_j)} \right\}. \tag{4.13}$$

The variable δ_{ij} equals zero if individual i does not experience a $(j-1)$st event and equals one otherwise.

As noted previously for (4.8), the justification for the validity of (4.12) as a likelihood function, and for (4.13), is different than for standard survival analysis settings; see Dabrowska et al. (1994) and Lawless et al. (2001). The bottom line, however, is that standard methods and software can be applied to (4.12) and (4.13).

Models (4.11) in which the $h_{0j}(w)$ and β_j terms are the same for $j = 1, 2, \ldots$ are also easily handled. In this case, the likelihood (4.8) and estimator (4.9) for $H_0(w)$ apply. Models where the functions $h_{0j}(w)$, $j = 1, 2, \ldots$, are different but $\beta_j = \beta$ for $j = 1, 2, \ldots$, are often called stratified Cox models as in Section 3.4.3; they too are handled by standard survival analysis software.

It may be of interest in some settings to test the hypothesis that the β_j terms in (4.11) are equal; this is easily done with a likelihood ratio test based on the likelihood functions (4.12). It may also be of interest to test that the baseline hazard functions $h_{0j}(t)$, $j = 1, 2, \ldots$ are equal, with or without an assumption that the β_j terms are. This is discussed in Section 3.4.3. It is also easy to carry out such tests with parametric models. Problem 4.2 describes parametric analysis for multiplicative hazards models.

Accelerated failure time (AFT) models are likewise easily handled. The AFT models analogous to (4.11) define $Y_{ij} = \log W_{ij}$ and are of the form (4.3) with

$$Y_{ij} = \beta_{0j} + z'_{ij}\beta_j + \sigma_j \epsilon_{ij} \qquad j = 1, 2, \ldots, \tag{4.14}$$

where $\epsilon_{1j}, \epsilon_{2j}, \ldots, \epsilon_{mj}$ are i.i.d. random variables with a fully specified distribution $G_j(\epsilon)$. Models for which $G_j(\epsilon)$ is a standard normal, logistic, or extreme value distribution are often used (Lawless 2003a, Ch. 6), and maximum likelihood estimation based on (4.10) can be carried out using survival

analysis software. When the parameters $(\beta_{0j}, \beta_j, \sigma_j)$ are distinct for each of $j = 1, 2, \ldots$ the analysis involves separate treatments for each gap. As mentioned previously, an assumption of equality of parameters for $j = 1, 2, \ldots$ may sometimes be of interest.

Sometimes analysts want to consider the marginal distribution of second or subsequent gap times, that is, conditional on covariates but not on previous gap times. If the W_{ij}, $j = 1, 2, \ldots$, are conditionally independent given the covariate vectors x_{ij} $(j = 1, 2, \ldots)$ then ignoring previous events and doing a separate analysis for each gap is valid. However, if this is not the case, then separate analyses that ignore dependence on previous gap times can be badly biased, because dependence among W_{ij}, $j = 1, 2, \ldots$, can create dependent censoring for W_{i2} and subsequent gap times. For example, if the observation period for individual i is $[0, \tau_i]$ then the potential censoring time for W_{i2} is $\tau_i - \min(W_{i1}, \tau_i)$. This is not independent of W_{i1} and so marginal analysis of W_{i2} without conditioning on W_{i1} involves dependent censoring. Estimates of the marginal distribution of W_{ij} given x_{ij} can in principle be obtained from conditional models (4.11) or (4.14) that have been fitted to the data. The expressions for the marginal distribution will generally be rather complicated, and the marginal effect of covariate x_{ij} on W_{ij} will not be easily interpreted. Models in which the joint distribution of gap times within an individual is specified more symmetrically than in (4.11) or (4.14) provide simpler interpretations. They are considered in Sections 4.2.2 and 4.2.3.

An exception to the problem above occurs when τ_i is defined so that some event, say the Jth, always occurs; in that case the first J gap times are observed for each individual, and none of them is ever censored.

4.2.2 Models with Random Effects

An alternative to forming models conditional on event history is to use individual-specific i.i.d. random effects to induce associations among gap times. The simplest such models assume that given a random effect u_i, the gap times W_{ij} $(j = 1, 2, \ldots)$ for individual i are independent. Because the random effects are unobserved, we base inference on the marginal likelihood obtained by integrating out the random effect from the joint distribution of the gap times and random effect for each subject. The likelihood function in this case takes the form (see (4.2))

$$L = \prod_{i=1}^{m} \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(w_{ij}|x_{ij}, u_i) \cdot S(w_{i,n_i+1}|x_{ij}, u_i) dG(u_i)\,, \tag{4.15}$$

where $G(u)$ is the common distribution function for the u_i and x_{ij} is a covariate vector associated with W_{ij}. Such models are rather special in the sense that, with fixed covariates (i.e. $x_{ij} = x_i$), they give an exchangeable correlation structure among the gap times for an individual. They can be useful

however, in situations where the gap times arising from a particular individual are independent, but unobservable factors create heterogeneity in the gap time distributions across individuals. In this case gap times within individuals are more similar than gap times from different individuals, and we say that unconditionally, within individuals, the gap times are correlated.

Many random effects models require special treatment, but two types are rather easily handled by standard survival analysis software. The first are so-called proportional hazards frailty models (e.g. Hougaard, 2000; Therneau and Grambsch, 2000) in which W_{ij} has conditional hazard function

$$h_{ij}(w|u_i) = u_i h_{0j}(w) \exp(x'_{ij}\beta_j) \qquad j = 1, 2, \ldots . \tag{4.16}$$

Semiparametric models for which the $h_{0j}(w)$ are unspecified can be fitted using the "`frailty`" option in the S-PLUS or R function `coxph`. Covariate effects on each W_{ij}, conditional on the unobserved random effect u_i, are thus given. However, obtaining an estimate of the marginal distribution for W_{ij} given x_{ij} is awkward because it involves integrating over u_i in the conditional model (4.16). The resulting model will not, in general, have a multiplicative form for the covariate effects.

A second family of models that is easily handled is the log-normal family for which $Y_{ij} = \log W_{ij}$ and the distribution of Y_{ij} given u_i and covariates x_{ij} is given by

$$Y_{ij} = \beta_{0j} + x'_{ij}\beta_j + u_i + \epsilon_{ij}\,, \tag{4.17}$$

where the ϵ_{ij} $(j = 1, 2, \ldots; i = 1, 2, \ldots, m)$ are i.i.d. $N(0, \sigma^2)$ and the u_i are i.i.d. $N(0, \sigma_u^2)$. For this model $\mathrm{var}(Y_{ij}) = \sigma^2 + \sigma_u^2$ and $\mathrm{cov}(Y_{ij}, Y_{ik}) = \sigma_u^2$ for $k \neq j$ so that $\mathrm{corr}(Y_{ij}, Y_{ik}) = \sigma_u^2/(\sigma^2 + \sigma_u^2)$, conditional on the covariate values. Furthermore, the marginal distribution of Y_{ij}, given x_{ij}, is $N(\beta_{0j} + x'_{ij}\beta_j, \sigma^2 + \sigma_u^2)$.

One way to fit these models is to note that they can be written in the symmetric form

$$Y_{ij} = \beta_{0j} + x'_{ij}\beta_j + e_{ij}\,, \tag{4.18}$$

where $e_{ij} = u_i + \epsilon_{ij}$ and for J gaps the vector $e_i = (e_{i1}, \ldots, e_{iJ})'$ is multivariate normal with mean 0 and covariance matrix Σ_J with diagonal entries $\sigma^2 + \sigma_u^2$ and off-diagonal entries σ_u^2. This allows the likelihood function to be written down easily when the last gap time for each individual is uncensored, and maximum likelihood estimates can be obtained using multipurpose optimization software. When the final gap times are censored a better approach is to rewrite (4.18) in terms of the univariate normal distributions for Y_{ij} given x_{ij} and $y_i^{(j-1)}$. This is described explicitly in Section 4.3.1.

4.2.3 Joint Gap Time Distributions

The random effects models in the preceding section do not always provide marginal distributions for gap times that are of a simple form. Another approach is to consider multivariate distributions for specified sets of gap times

$W_{i1}, \ldots, W_{iJ}$, where $J \geq 2$. There is a substantial literature on multivariate lifetime distributions that can be utilized (e.g. Joe, 1997; Hougaard, 2000). A useful approach is to consider models of the form

$$\begin{aligned} S(w_1, \ldots, w_J) &= \Pr(W_1 > w_1, \ldots, W_J > w_J) \\ &= C(S_1(w_1), \ldots, S_J(w_J); \phi)\,, \end{aligned} \tag{4.19}$$

where $S_j(w_j)$ is the marginal survivor function for W_j, ϕ is a vector of parameters, and $C(u_1, \ldots, u_J; \phi)$ is a J-variate cumulative distribution function with uniform (0,1) marginal distributions. This ensures that $S_j(w_j) = S(0, \ldots, w_j, \ldots, 0)$ is indeed the marginal survivor function for W_j. The function $C(\cdot)$ is often referred to as a *copula*, and the form of $C(\cdot)$ and the parameter ϕ determine the association among $W_1, \ldots, W_J$. A major advantage of this approach is that specific types of distributions can be used for the marginal survivor functions $S_j(w_j)$ $(j = 1, 2, \ldots)$, according to modeling needs.

An important family of distributions of the type (4.19) was introduced by Clayton (1978). If we incorporate covariates, so that

$$S_{ij}(w_j) = \Pr(W_{ij} > w_j | x_{ij})\,,$$

these have joint survivor function of the form

$$S(w_1, \ldots, w_J | x_i) = \left\{ \sum_{j=1}^{J} S_{ij}(w_j)^{-\phi} - (J-1) \right\}^{-\phi^{-1}}, \tag{4.20}$$

where $\phi > 0$ and x_i represents the vectors $x_{i1}, \ldots, x_{iJ}$. Different models are obtained by specifying different parametric or semiparametric distributions for $S_{ij}(w_j)$ $(j = 1, 2, \ldots)$; for example, accelerated failure time or proportional hazards models could be used.

The model (4.20) and other models of the form (4.19) can be related to random effects models. One limitation of many types of random effect and copula models is that only one or two parameters are used to model association among $W_1, \ldots, W_J$. This can be inadequate when association structures are complex or changing over time. A second limitation is that these approaches do not readily deal with negative associations. Problem 4.5 looks at the family of models (4.20).

Less restrictive models can be constructed. The simplest and most useful of such models are multivariate accelerated failure time models (e.g. He and Lawless, 2005), for which the variables $Y_{ij} = \log W_{ij}$ follow multivariate location-scale distributions. Specifically, for a specified J the vector of log gap times $(Y_{i1}, \ldots, Y_{iJ})'$ follows a model with

$$Y_{ij} = x'_{ij}\beta_j + e_{ij}\,, \tag{4.21}$$

where $e_{i1}, \ldots, e_{iJ}$ have a joint distribution that does not depend on the x_{ij}. Multivariate log-normal models, where the vectors $e_i = (e_{i1}, \ldots, e_{iJ})'$ are

multivariate normal with mean 0 and covariance matrix Σ_J, are especially attractive. Then $E\{Y_{ij}|x_{ij}\} = x'_{ij}\beta$ and Σ_J can accommodate general types of association among the log gap times.

The normal random effects model (4.17) is a special case of the multivariate normal model (4.21). The latter also connects directly with conditional models of the form (4.14). In particular, the distribution of Y_{ij} given $y_i^{(j-1)} = (y_{i1}, \ldots, y_{i,j-1})'$ from (4.21) is normal with mean

$$E(Y_{ij}|x_i, y_i^{(j-1)}) = x'_{ij}\beta_j + \Sigma_{j,j-1}\Sigma_{j-1}^{-1}(y_i^{(j-1)} - \mu_i^{(j-1)}) \tag{4.22}$$

and variance $\sigma_j^2 - \Sigma_{j,j-1}\Sigma_{j-1}^{-1}\Sigma'_{j,j-1}$, where $\mu_i^{(j-1)} = (x'_{i1}\beta_1, \ldots, x'_{i,j-1}\beta_{j-1})'$ and Σ_j is partitioned as

$$\Sigma_j = \begin{pmatrix} \Sigma_{j-1} & \Sigma_{j-1,j} \\ \Sigma_{j,j-1} & \sigma_j^2 \end{pmatrix}.$$

For example, Y_{i2} given x_i and y_{i1} is normal with mean and variance

$$E(Y_{i2}|x_i, y_{i1}) = x'_{i2}\beta_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(y_{i1} - x'_{i1}\beta_1)$$

$$\text{var}(Y_{i2}|x_i, y_{i1}) = (1 - \rho_{12}^2)\sigma_2^2,$$

where $\sigma_1^2 = \text{var}(Y_{i1}|x_i)$, $\sigma_2^2 = \text{var}(Y_{i2}|x_i)$, and $\rho_{12} = \text{corr}(Y_{i1}, Y_{i2}|x_i)$.

The models in this section are chosen so that the marginal distributions $S_{ij}(w_j|x_{ij})$ are of easily interpretable forms. Aside from the bivariate normal example above, the conditional distributions are usually less easy to interpret.

Maximum likelihood estimation can be based on the likelihood function (4.10), noting that the jth term requires the distribution of W_{ij} given x_i and $w_i^{(j-1)}$. Note that the ith individual's contribution to the likelihood (4.10) can alternatively be expressed in the form

$$\begin{aligned} &\Pr(W_{i1} = w_{i1}, \ldots, W_{in_i} = w_{in_i}, W_{i,n_i+1} > w_{i,n_i+1}) \\ &\quad = \frac{-\partial^{n_i} S_i(w_{i1}, \ldots, w_{in_i}, w_{i,n_i+1})}{\partial w_{i1}, \ldots, \partial w_{in_i}}, \end{aligned} \tag{4.23}$$

where $S_i(w_1, \ldots, w_j)$ is the joint survivor function for $W_{i1}, \ldots, W_{ij}$ given x_i. This form is convenient for models (4.19) in which the marginal survivor functions $S_{ij}(w_{ij})$ have simple algebraic forms.

With random effects models, integration over the distribution of the random effects gives a joint distribution for gap times. It has been noted, however, that for multiplicative hazards models (4.16), the marginal distribution for W_{ij} given x_{ij} may not have easily interpreted covariate effects once the random effect has been integrated out. This is also the case for some other types of random effects models, although not for accelerated failure time (or "log-location-scale") models such as (4.18).

4.2.4 Modulated Renewal Processes

In situations that involve time-varying covariates or more complex relationships between event occurrence and prior event history, analysis based on specifications of the event intensity function is usually preferable. When there is an intrinsic interest in gap times, modulated renewal processes are useful. As described in Section 2.3, they are multiplicative models in which the event intensity function takes the form

$$\lambda(t|H(t)) = h_0(B(t)) \exp(z'(t)\beta)\,, \tag{4.24}$$

where t represents chronological time and $z(t)$ is a vector of time-varying covariates that may be based on external covariates $\{x(s), 0 \le s \le t\}$ as well as prior event history. The function $h_0(w)$ in (4.24) is the hazard function for a gap time when $z(t)$ is identically zero. With this model the hazard function for $W_k = T_k - T_{k-1}$, given the time t_{k-1} of the $(k-1)$st event and covariates, is

$$h_k(w) = h_0(w) \exp(z'(t_{k-1} + w)\beta)\,.$$

In many cases a model where $z(t_{k-1} + w)$ is a fixed vector z_k suffices; that is, $z(t)$ is constant over intervals $(t_{k-1}, t_k]$ between events.

Semiparametric models in which $h_0(w)$ is an arbitrary hazard function can be handled in the standard fashion for the Cox survival model with time-varying covariates. That is, β in (4.24) can be estimated by maximizing the partial likelihood function

$$L(\beta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left\{ \frac{\exp(z_i'(t_{ij})\beta)}{\sum_{l=1}^{m} \sum_{k=1}^{n_\ell+1} I(w_{lk} \ge w_{ij}) \exp(z_\ell'(t_{l,k-1} + w_{ij})\beta)} \right\}, \tag{4.25}$$

where $t_{ij} = w_{i1} + \cdots + w_{ij}$ is the chronological time of the jth event for individual i. Note that the covariate value of an individual at risk of a kth event at gap time w is evaluated at the time $t = t_{l,k-1} + w$. The baseline cumulative hazard function $H_0(w) = \int_0^w h_0(u)du$ is estimated by

$$\widehat{H}_0(w) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \frac{I(w_{ij} \le w)}{\sum_{l=1}^{m} \sum_{k=1}^{n_\ell+1} I(w_{lk} \ge w_{ij}) \exp(z_l'(t_{l,k-1} + w_{ij})\widehat{\beta})} \right\}. \tag{4.26}$$

Furthermore, the usual asymptotic variance estimates and inference techniques from survival analysis can be used, thus allowing analysis to be carried out with standard software. Stratified versions of (4.24), in which $h_0(B(t))$ is replaced with $h_{0j}(B(t))$ for jth gaps, can also be handled. The regression coefficients β may also be allowed to vary across $j = 1, 2, \ldots$.

As remarked earlier in this section, the justification of (4.25) and (4.26) is rather different than typically presented in the standard survival setting. Problem 4.3 describes how these estimation procedures arise from parametric maximum likelihood methods for models in which the baseline hazard functions are piecewise-constant.

4.3 Examples

We consider two illustrations of gap time analysis. For the second, some code for S-PLUS or R is given in Appendix C.

4.3.1 Bowel Motility Cycles

In Example 4.1 renewal models were fitted to the data on the lengths W_j $(j = 1, 2, \ldots)$ of successive digestive cycles in a study on 19 human subjects. We consider here extended models which allow the assumptions for the renewal models to be assessed through tests of hypotheses. The models are log-normal, which is suggested by preliminary plots of the data, and are as follows, with $Y_{ij} = \log W_{ij}$.

Model A: Conditional Model

$$Y_{i1} \sim N(\mu_1, \sigma_1^2)$$

$$Y_{ij}|y_i^{(j-1)} \sim N(\beta_0 + \beta_1 y_{i,j-1}, \sigma_2^2) \text{ for } j = 2, 3, \ldots.$$

Model B: Random Effects Model

This is of the form (4.17) with $\beta_{0j} = \mu$ and $\beta_j = 0$; that is, for given J, $(Y_{i1}, \ldots, Y_{iJ})'$ is multivariate normal, with $E(Y_{ij}) = \mu$, $\text{var}(Y_{ij}) = \sigma^2 + \sigma_u^2$, and $\text{cov}(Y_{ij}, Y_{ik}) = \sigma_u^2$ for $j \neq k$. It is readily shown that $Y_{ij}|y_i^{(j-1)} \sim N(m_{ij}, v_{ij})$, where

$$m_{ij} = \frac{\sigma^2}{\sigma^2 + (j-1)\sigma_u^2}\mu + \frac{\sigma_u^2}{\sigma^2 + (j-1)\sigma_u^2}\sum_{l=1}^{j-1} y_{i\ell}$$

$$v_{ij} = \frac{\sigma^2(\sigma^2 + j\sigma_u^2)}{\sigma^2 + (j-1)\sigma_u^2}\,.$$

Model A is easily fitted with standard survival analysis software but it is necessary to maximize the likelihood (4.10) for Model B using general optimization software (see Appendix B). By fitting submodels of A and B we can assess whether the gap times for an individual are possibly independent. Estimates and maximum log-likelihood values for the models are provided in Table 4.1.

Table 4.1. Results of fitting various models to bowel mobility data.

Model	$\ell(\widehat{\theta})$	Parameter Estimates
A	-68.59	$\widehat{\beta}_0 = 3.89$, $\widehat{\beta}_1 = 0.13$, $\widehat{\mu}_1 = 4.75$, $\widehat{\sigma}_1 = 0.40$, $\widehat{\sigma}_2 = 0.56$
A ($\beta_1 = 0$)	-69.12	$\widehat{\beta}_0 = 4.45$, $\beta_1 = 0$, $\widehat{\mu}_1 = 4.75$, $\widehat{\sigma}_1 = 0.40$, $\widehat{\sigma}_2 = 0.56$
B	-72.17	$\widehat{\mu} = 4.54$, $\widehat{\sigma} = 0.52$, $\widehat{\sigma}_u = 0.17$
B ($\sigma_u = 0$)	-73.12	$\widehat{\mu} = 4.51$, $\widehat{\sigma} = 0.55$

Model A is best supported by the data, and a likelihood ratio (LR) test of the hypothesis that $\beta_1 = 0$ gives an observed value for the LR statistic of $2(-68.59 - (-69.12)) = 1.06$. Treating the LR statistic as approximately χ^2_1 if $\beta_1 = 0$ is true, we obtain the $p-$value as $\Pr(\chi^2_1 \geq 1.06) = 0.303$. There is therefore little evidence of first-order dependence among the gap times. Similarly, a likelihood ratio test of the hypothesis that $\sigma_u = 0$ in Model B gives an observed LR statistic of 1.90 and an approximate χ^2_1 $p-$value of 0.168, again providing little evidence of association.

The log-normal model that was fitted in Example 4.1 indicated that W_{i1} appears to have a slightly different distribution than subsequent cycle lengths; this agrees with the superiority of Model A with $\beta_1 = 0$ over Model B. Other distributions can also be fitted to the gap times, as discussed in Section 4.2.3, and give similar results; the log-normal model, however, fits as well as any. Semiparametric models based on (4.11) could also be considered.

In summary, our conclusions from this data analysis are the same as those in Example 4.1. Specifically we conclude that cycle lengths are roughly independent, and that first cycles tend to be slightly longer than subsequent ones. The estimation of marginal cycle length distributions as described in Section 4.1 and Example 4.1 is justified by the independence, and Figure 2.2 provides a good summary of the evidence. A conclusion of scientific interest is that cycle lengths are highly variable within individuals.

4.3.2 Pulmonary Exacerbations and rhDNase Treatment

Data were introduced in Section 1.2.3 on the occurrence of pulmonary exacerbations (bouts of infection) in a clinical trial of persons with cystic fibrosis. Subjects in the study were randomly assigned to receive either a daily dose of the experimental treatment rhDNase or a daily dose of a placebo. The study was double blind, and most subjects were followed for approximately 169 days. Table 1.2 shows the numbers of exacerbations per subject by treatment group, and indicates that persons in the rhDNase group experienced fewer exacerbations during the study. Here we report on some analyses of gap times between successive infections. A complication is that when an exacerbation occurs the subject is treated with antibiotics, and a subsequent exacerbation cannot occur before the end of the treatment period. The treatment periods

are highly variable, although a majority last from 10–15 days. To begin, we ignore them, except for defining the jth gap for a subject as extending from the termination of treatment for the $(j-1)$st exacerbation to the start of the jth exacerbation. Later, we consider possible association between times to exacerbations and previous treatment times.

Because relatively few persons experienced two or more exacerbations, we consider only the first two gap times, that is, the times W_{i1} to the first exacerbation and the times W_{i2} between the first and second exacerbations. In addition to the treatment indicator covariate $x_{i1} = I(\text{subject } i \text{ received rhDNase})$, a covariate x_{i2} is included which is a subject's forced expiratory volume (FEV), a measure of a person's lung function, measured at the time of randomization. We fitted two types of models:

(i) Cox proportional hazards models (4.11) for which

$$h_{ij}(w) = h_{0j}(w)\exp(\beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \beta_{j3}w_{i1}I(j=2)) \qquad j = 1, 2,$$

(ii) Log-normal accelerated failure time models (4.14), where

$$Y_{ij} = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \beta_{j3}\log(w_{i1})I(j=2) + \sigma_j\epsilon_{ij} \qquad j = 1, 2,$$

where $\epsilon_{ij} \sim N(0,1)$. The log-normal model was chosen as providing a good fit to the data.

In the analyses described below, x_{i2} is a centered FEV variable, obtained by subtracting the mean FEV value across all subjects from each FEV value.

Both sets of models are easily fitted using standard survival analysis software. We used the S-PLUS function `coxph` for (i) and `survReg` for (ii); commands are shown in Appendix C.

Results are shown in Table 4.2. The results for W_1 indicate a strong positive treatment effect after adjustment for FEV; as one might expect, FEV is also highly significant. The first gap times w_{i1} are highly significant in connection with W_2. This indicates a strong positive association between W_{i1} and W_{i2} even after adjustment for treatment and FEV, suggesting that unobserved individual factors also influence the occurrence of exacerbations. Because of the limited amount of data on second gaps and the strong association between W_{i1} and the covariates seen in the analysis of first gap times, it is impossible to separate clearly the effects of w_1, treatment, and FEV on W_2. We see in Table 4.2 that neither treatment nor FEV is in fact significant for W_2, following adjustment for W_1. A further complication is that the duration of antibiotic treatment for a first exacerbation has an effect on time to a second exacerbation, with longer treatment times increasing the risk of a second exacerbation. This can be seen by including the duration of treatment for the first exacerbation as an additional covariate in models for W_2. The effects of treatment (rhDNase or placebo) and FEV remain insignificant however.

Checks for treatment by FEV interactions did not provide evidence of an interaction. Diagnostic checks indicate that both models provide good fits to

Table 4.2. Fitted models for W_1 and for W_2 given W_1.

Gap Time	Parameter	Cox PH EST.	Cox PH S.E.	Log-normal AFT EST.	Log-normal AFT S.E.
W_1	β_{10} (intercept)	-	-	5.40	0.11
	β_{11} (trt)	-0.38	0.13	0.43	0.14
	β_{12} (FEV)	-0.021	0.003	0.022	0.003
	σ_1			1.45	0.07
W_2	β_{20} (intercept)	-	-	3.21	0.49
	β_{21} (trt)	0.36	0.23	-0.23	0.21
	β_{22} (FEV)	0.001	0.005	-0.005	0.005
	$\beta_{23}(w_1$ or $\log w_1)^\dagger$	-0.014	0.004	0.42	0.13
	σ_2	-	-	1.23	0.11

$^\dagger w_1$ for PH model and $\log w_1$ for AFT model.

the data. Although the PH and AFT models differ in important respects, the relatively few first exacerbations, and many fewer second exacerbations, do not provide sufficient information to favor one model over the other. It is noted that the two models give very similar $p-$values in tests for covariate effects.

The analysis here illustrates a general problem in assessing the effects of fixed baseline covariates on gap times with a conditional approach. If the gap times are not independent (after conditioning on the baseline covariates), then the effects of covariates on second and subsequent gap times are confounded with the effects of prior gap times. It is possible to average over W_{i1} in the models for W_{i2}, in order to obtain the marginal distribution of W_{i2}. This is unwieldy for the PH model but is not difficult for the normal model, and calculation shows that in the marginal distribution for W_{i2} given x_{i1} and x_{i2}, neither treatment or FEV is significant.

Another way to approach this is by considering joint distributions for $W_1, W_2, \ldots$ which involve convenient parameterizations for the effects of covariates on the marginal distributions. This can be done according to models (4.17) and (4.18) or the methods in Section 4.2.3. Taking the approach in Section 4.2.3, and considering only the first two gap times, we would fit a bivariate model for (W_{i1}, W_{i2}) given the covariates x_{i1} (trt) and x_{i2} (FEV). Let us consider a bivariate normal model for $(Y_{i1}, Y_{i2}) = (\log W_{i1}, \log W_{i2})$. As discussed in Section 4.2.3 this is equivalent to the model (ii) above, and takes (Y_{i1}, Y_{i2}) bivariate normal with

$$E(Y_{i1}|x_{i1}, x_{i2}) = \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2}$$
$$E(Y_{i2}|x_{i1}, x_{i2}) = \gamma_{20} + \gamma_{21}x_{i1} + \gamma_{22}x_{i2}$$

and $\text{var}(Y_{i1}|x_{i1}, x_{i2}) = \sigma_1^2$, $\text{var}(Y_{i2}|x_{i1}, x_{i2}) = \sigma_{2m}^2$, $\text{cov}(Y_{i1}, Y_{i2}|x_{i1}, x_{i2}) = \rho\sigma_1\sigma_{2m}$. We have retained here the same parameter labels for Y_{i1} as in model (ii), and by the results (4.22) in Section 4.2.3 the parameters for Y_{i2} are related to those of model (ii) as follows:

$$\gamma_{2j} = \beta_{2j} + \beta_{1j}\beta_{23} \qquad j = 0, 1, 2,$$

$$\sigma_{2m}^2 = \beta_{23}^2\sigma_1^2 + \sigma_2^2 = \sigma_2^2/(1 - \rho^2)\,.$$

Maximum likelihood estimates (with standard errors in brackets) for the parameters in the marginal distribution for Y_{i2} are $\widehat{\gamma}_{20} = 5.46$ (0.30), $\widehat{\gamma}_{21} = -0.048$ (0.215), $\widehat{\gamma}_{22} = 0.0045$ (0.0053) and $\widehat{\sigma}_{2m} = 1.37$ (0.15). As remarked above, neither treatment nor FEV is significant and the indicated effects of these covariates on W_2 are much different than their effects on W_1. A qualification of this and similar analyses is that under half of the subjects in the study experienced even a first exacerbation, and it may be misleading to compute the marginal distribution of W_2, which implicitly assumes that everyone eventually experiences a first exacerbation. Furthermore, it is assumed that the log-normal model (or some other model) provides an adequate description of the upper half of the distribution, which is effectively unobserved in the study.

We can similarly fit proportional hazards joint frailty models such as (4.16). Doing this for W_1 and W_2 by using the S-PLUS or R function `coxph` with a gamma random effect, and allowing the effects of treatment and FEV to differ for W_1 and W_2, we obtain results very similar to those for the bivariate normal models. In particular, the effects of treatment and FEV on W_2 are insignificant, and there is strong evidence of association between W_1 and W_2, captured here by the variance ϕ of the random effect.

The association between the gap times for an individual makes it more difficult to address clinically important questions concerning persistence or time trends in the treatment effects, in a gap time analysis. We look further at this point in Section 4.4.3 and also in Section 5.5.1, where more general intensity-based models are considered.

4.4 Estimation of Marginal Gap Time Probabilities

As discussed in Sections 4.2.2 and 4.2.3, we may wish to consider the marginal distributions of specific gap times, even though the successive gap times for an individual are not independent. For example, in some settings there may be a hypothesis that, even after conditioning on fixed observable covariates, the gaps W_j $(j = 1, 2, \ldots)$ between successive events tend to decrease as j increases. An example for the case of episodes of affective disorder in psychiatric patients is given by Kessing et al. (1998, 1999). As stressed in Sections 2.3.2 and 4.1, it is in most settings improper simply to fit marginal models

for jth gap times W_j $(j \geq 2)$ by ignoring the possibility of association within individuals. This is because the effective censoring time for W_j depends on $W_1 + \cdots + W_{j-1}$, and hence is not independent of W_j when gap times are non-independent. An exception is when a specified number of events are observed for every individual, but this type of observation scheme is rather rare.

Figure 4.1 provides an illustration of this effect of dependent censoring. It shows results from a simulated random sample of 200 individuals, where the log gap times, $Y_{ij} = \log W_{ij}$ $(j = 1, 2, \ldots)$, were generated from a multivariate normal distribution in which each Y_{ij} was identically distributed as $N(2, 2)$, and with equal correlations $\text{corr}(Y_{ij}, Y_{ik}) = 0.5$ for $j \neq k$. A common censoring time $C_i = 52$ was imposed on each process. This simulation mimics studies in which individuals are followed for 52 weeks, and where the number of events per individual is small. Figure 4.1 shows Kaplan–Meier estimates $\widehat{S}_j(w)$ for the first four gap times W_j $(j = 1, 2, 3, 4)$, based on the observed or censored gap times, with no accounting for association. As expected, the survivor functions for the second and subsequent gap times are severely underestimated, with the problem worsening for higher gap times.

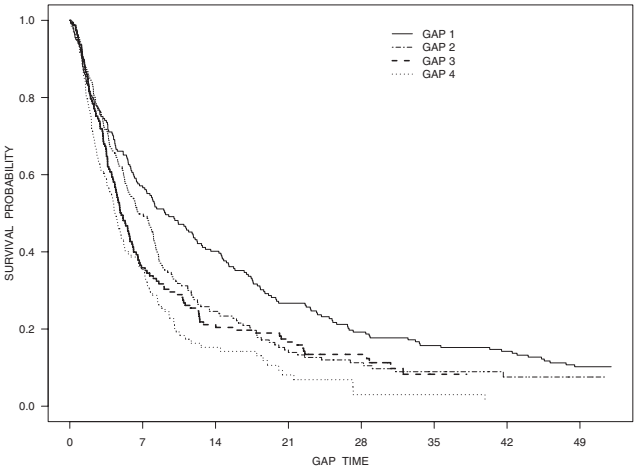


Fig. 4.1. Kaplan–Meier estimates for the first four gap times, improperly ignoring dependent censoring.

One way to facilitate examination of specific gap time distributions is to fit random effects models described in Section 4.2.2. A second approach is to specify a family of multivariate distributions for gap times, as in Section 4.2.3. A third way is to fit conditional models as discussed in Section 4.2.1, and then to obtain marginal distributions from them. A difficulty in the latter

case is that the effects of covariates in the marginal distributions will generally (except for normal models) be complex. In addition, nonparametric or semiparametric estimation is rather difficult with all approaches, as far as obtaining the marginal distributions is concerned. A fourth approach, which we consider below, is to use the idea of inverse probability of censoring (IPC) weights, as first introduced by Robins and Rotnitzky (1992). This allows estimators of a simple form to be used; we consider this in the next section.

4.4.1 Nonparametric Estimation of Marginal Gap Time Distributions

Nonparametric estimates of the marginal or joint distributions of successive gap times $W_1, W_2, \ldots$ can be useful when there is no conditioning on covariates. We begin by considering the first two gap times W_1, W_2; as discussed later this also allows the consideration of the marginal distributions of higher gaps. Assume that the pairs (W_{i1}, W_{i2}) are independent and identically distributed for $i = 1, \ldots, m$, with joint distribution function

$$F(w_1, s_2) = \Pr(W_{i1} \leq w_1, W_{i2} \leq w_2)\,. \tag{4.27}$$

Let C_i represent the censoring time for individual i, so that what we may observe is $\tilde{T}_{i1} = \min(W_{i1}, C_i)$, $\tilde{T}_{i2} = \min(W_{i1} + W_{i2}, C_i)$, $\delta_{i1} = I(W_{i1} \leq C_i)$, and $\delta_{i2} = I(W_{i1} + W_{i2} \leq C_i)$, with $(\tilde{t}_{i1}, \tilde{t}_{i2}, \delta_{i1}, \delta_{i2})$, $i = 1, \ldots, m$ used to refer to the actual observed data. The likelihood function is as in (4.10),

$$L = \prod_{i=1}^{m} f_1(w_{i1})^{\delta_{i1}} S_1(C_i)^{1-\delta_{i1}} f_{2|1}(w_{i2}|w_{i1})^{\delta_{i1}\delta_{i2}} S_{2|1}(C_i - w_{i1}|w_{i1})^{\delta_{i1}(1-\delta_{i2})}\,,$$

where $f_{2|1}$ and $S_{2|1}$ are the conditional density and survivor functions for W_2 given W_1. The Kaplan–Meier estimate for $S_1(w)$ based on the data $(\tilde{t}_{i1}, \delta_{i1})$, $i = 1, \ldots, m$ is a nonparametric maximum likelihood estimate. However, maximization of L with respect to $S_{2|1}$ is not feasible unless the model is discrete and there are sufficiently many events at each value w_1. In some settings it may be sensible to discretize the time axis and maximize with respect to $S_{2|1}(\cdot|w_1)$ for each value of w_1. This can be combined with the maximum likelihood estimate of $S_1(\cdot)$ to obtain an estimate of the joint distribution function (4.27); from this we can obtain an estimate of the marginal probabilities for W_2. Here we consider an alternative approach, based on the idea of inverse probability of censoring weights (IPCW), used by Lin et al. (1999) and developed further by van der Laan et al. (2002).

As discussed by Lin et al. (1999), it is convenient to consider instead of $F(w_1, w_2)$ the function

$$H(w_1, w_2) = \Pr(W_{i1} \leq w_1, W_{i2} > w_2)\,, \tag{4.28}$$

from which $F(w_1, w_2)$ and other quantities of interest can be obtained. A crucial observation is that $H(w_1, w_2)$ is estimable from a given dataset only for

values (w_1, w_2) for which $w_1 + w_2 \leq C_{\max}$, where $C_{\max} = \max(C_1, \ldots, C_m)$. This has repercussions, discussed below, for the estimation of the marginal distribution of W_2 or any higher gap times. For now we assume that censoring times C_i are completely independent of the event processes $\{N_i(t), 0 \leq t\}$ and let $G(c) = \Pr(C_i > c)$ be the survivor function for C_i; ways to relax this independence assumption are discussed later.

The estimate for $H(w_1, w_2)$ of Lin et al. (1999) is based on the observation that

$$E\left\{\frac{I(W_{i1} \leq w_1, W_{i2} > w_2) I(C_i > W_{i1} + w_2)}{G(W_{i1} + w_2)}\right\} = H(w_1, w_2)\,, \tag{4.29}$$

where the expectation is taken over the distributions of the full event process, and C_i. To see that (4.29) holds, note that because C_i is independent of $(W_{i1}, W_{i2}, \ldots)$, $E\{I(C_i > W_{i1} + w_2 | N_i^{(\infty)})\} = G(W_{i1} + w_2)$ and then that $E\{I(W_{i1} \leq w_1, W_{i2} > w_2)\} = H(w_1, w_2)$. If we use an estimate $\widehat{G}(c)$ of $G(c)$, then (4.29) motivates the estimate

$$\widehat{H}(w_1, w_2) = \frac{1}{m} \sum_{i=1}^{m} \frac{I(w_{i1} \leq w_1, w_{i2} > w_2, C_i > w_{i1} + w_2)}{\widehat{G}(w_{i1} + w_2)}\,. \tag{4.30}$$

This can be expressed in the equivalent form

$$\widehat{H}(w_1, w_2) = \frac{1}{m} \sum_{i=1}^{m} \frac{I(\tilde{t}_{i1} \leq w_1, \tilde{t}_{i2} - \tilde{t}_{i1} > w_2)}{\widehat{G}(\tilde{t}_{i1} + w_2)}\,,$$

where we note that for $w_1 > 0, w_2 > 0$, the condition $\tilde{t}_{i2} - \tilde{t}_{i1} > w_2$ implies that $\tilde{t}_{i1} = w_{i1}$ and $\delta_{i1} = 1$. Assuming that $\widehat{G}(c)$ is a consistent estimator of $G(c)$, (4.30) is a consistent estimator of $H(w_1, w_2)$.

The estimate $\widehat{G}(c)$ can be the empirical survivor function, if all of $C_1, \ldots, C_n$ are observed. If not, a Kaplan–Meier estimate based on the data $(\tilde{t}_{i2}, 1 - \delta_{i2})$ can be used for settings where C_i is censored by the time T_{i2} of the second event, and other estimates can be obtained for other situations. The estimate of the marginal distribution $F_1(w) = H(w_1, 0)$ that comes from (4.30) is

$$\widehat{F}_1(w_1) = \frac{1}{m} \sum_{i=1}^{m} \delta_{i1} \frac{I(\tilde{t}_{i1} \leq w_1)}{\widehat{G}(\tilde{t}_{i1})}\,.$$

If $\widehat{G}(c)$ is the Kaplan–Meier estimate based on the data $(\tilde{t}_{i1}, 1 - \delta_{i1})$ then $1 - \widehat{F}_1(w)$ can be shown to equal the Kaplan–Meier estimate for $S_1(w)$ based on the data $(\tilde{t}_{i1}, \delta_{i1})$.

Estimation of the marginal probabilities for W_2 requires careful examination when censoring is present. As noted above, $H(w_1, w_2)$ can be estimated only for values (w_1, w_2) for which $w_1 + w_2 \leq C_{\max}$. A consequence of this is that, unless W_1 has finite support $(0, \tau_1)$ with $\tau_1 < C_{\max}$, we cannot estimate

marginal probabilities $S_2(w_2)$ for any values $w_2 > 0$. Superficially, it might appear that because $S_2(w_2) = H(\infty, w_2)$, we can estimate $S_2(w_2)$ as $\widehat{H}(\infty, w_2)$. However, when $w_1 = \infty$ is put into (4.30), the estimate exists only for w_2 such that $w_{i1} + w_2 \leq C_{\max}$ for all $i = 1, \ldots, m$, and what is actually being estimated is not $S_2(w_2)$, but $H(C_{\max} - w_2, w_2)$, or $\Pr(W_1 \leq C_{\max} - w_2, W_2 > w_2)$. Unless $\Pr(W_1 > C_{\max} - w_2, W_2 > w_2)$ equals zero, this does not equal $S_2(w_2)$.

In general, for W_2 we can consider estimated probabilities

$$\widehat{\Pr}(W_2 > w_2 | W_1 \leq w_1) = \frac{\widehat{H}(w_1, w_2)}{\widehat{H}(w_1, 0)}, \tag{4.31}$$

for (w_1, w_2), where $w_1 + w_2 < C_{\max}$.

Lin et al. (1999) provide variance estimates for $\widehat{H}(w_1, w_2)$ and for (4.31), which can be used to obtain pointwise confidence limits. The estimate for (4.31) is rather complicated with $\widehat{\text{var}}\{\widehat{\Pr}(W_2 > w_2 | W_1 \leq w_1)\}$ given by

$$\frac{1}{m^2 \widehat{H}(w_1, 0)^2} \sum_{i=1}^{m} \left\{ \delta_{i1} I(\tilde{t}_i \leq w_1) \left[\frac{\widehat{H}(w_2|w_1)}{\widehat{G}(\tilde{t}_{i1})} - \frac{I(\tilde{t}_{i2} - \tilde{t}_{i1} > w_2)}{\widehat{G}(\tilde{t}_{i1} + w_2)} \right]^2 \right.$$
$$\left. - \frac{m^2 (1 - \delta_{i2}) \widehat{B}(w_1, w_2; \tilde{t}_{i2})^2}{\left[1 + \sum_{j=1}^{m} I(\tilde{t}_{j2} > \tilde{t}_{i2})\right] \sum_{j=1}^{m} I(\tilde{t}_{j2} > \tilde{t}_{i2})} \right\},$$

where $\widehat{H}(w_2|w_1) = \widehat{H}(w_1, w_2)/\widehat{H}(w_1, 0)$ and

$$\widehat{B}(w_1, w_2; u) = \widehat{H}(w_2|w_1)\{\widehat{H}(w_1, 0) - \widehat{H}(u, 0)\}^+ - \{\widehat{H}(w_1, w_2) - \widehat{H}(u - w_2, w_1)\}^+$$

where $a^+ = \max(a, 0)$. An alternative for getting variance estimates or confidence limits is to use a naive bootstrap, in which samples of size m are drawn with replacement from the m observations $(\tilde{t}_{i1}, \tilde{t}_{i2}, \delta_{i1}, \delta_{i2})$.

It should be noted that the estimate (4.31) is not necessarily strictly monotonic in w_2, for given w_1, although it tends to be as the sample size increases. A monotonic estimate is given by

$$\widehat{\Pr}(W_2 > w_2 | W_1 \leq w_1) = \frac{\min_{u \leq w_2} \widehat{H}(w_1, u)}{\widehat{H}(w_1, 0)}. \tag{4.32}$$

This has the same asymptotic properties as (4.31).

Van der Laan et al. (2002) describe more general estimators and show how to relax the assumption that the C_i are completely independent of the event processes. Their approach allows censoring to depend on prior event history or on previously observed covariates. Put into the present framework, the idea is to consider estimation of a parameter θ that is defined as $\theta = E(B_i)$, where $B_i = g(W_{i1}, \ldots, W_{ik})$ is a function of some number of gap times. For example, if $B_i = I(W_{i1} \leq w_1, W_{i2} > w_2)$ for given values $w_1 > 0, w_2 > 0$ then

$\theta = H(w_1, w_2)$. Let us also define $\Delta_i = I(B_i \text{ is observed})$ and let V_i represent the earliest time in the process $\{N_i(t), 0 \le t\}$ at which B_i can be observed. For simplicity we suppress any dependence on covariates in the notation. The key idea is based on having a model for

$$E(\Delta_i | N_i^{(\infty)}) = \Pr(C_i > V_i | N_i^{(\infty)}), \tag{4.33}$$

where $N_i^{(\infty)} = \{N_i(t), 0 \le t\}$, and noting that

$$E\left\{ \frac{\Delta_i B_i}{E(\Delta_i | N_i^{(\infty)})} \right\} = E(B_i) = \theta .$$

This motivates the estimator

$$\widehat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{\Delta_i B_i}{\widehat{E}(\Delta_i | N_i^{(\infty)})} . \tag{4.34}$$

For this to be usable, we need assumptions about $E(\Delta_i | N_i^{(\infty)})$ that allow it to be estimated from the observed data. It is certainly possible to do this if we assume that censoring times C_i are completely independent of the full event history $N_i^{(\infty)}$, but we can also do it if we assume that the hazard function for the censoring time depends only on past observations. That is, we can allow the hazard for C_i at process time t to be

$$\lambda_c(t | N^{(\infty)}) = \lambda_c(t | N^{(t^-)}) ,$$

where $N^{(t^-)}$ is the process history up to time t^-. Covariates can also be added, and a convenient strategy for modeling the censoring time hazard would often be to use a Cox model with $\lambda_c(t | N^{(\infty)}) = \lambda_0(t) \exp(z'(t)\beta)$, where $z(t)$ is a vector of observable variables that can include information on $N^{(t^-)}$ and covariates.

With $B_i = I(W_{i1} \le w_1, W_{i2} > w_2)$ for fixed (w_1, w_2), the indicator Δ_i is equal to $I(W_{i1} + w_2 < C_i)$ and $V_i = W_{i1} + w_2$. If C_i is assumed completely independent of $N_i^{(\infty)}$, then (4.34) gives the estimator (4.30). If, however, it was thought that the risk of censoring might change with the occurrence of the first event, then we could consider a model for C_i with hazard function $\lambda_c(t) = \lambda_0(t) \exp\{\beta I(N_i(t^-) > 0)\}$. Then,

$$\begin{aligned} \Pr(C_i > V_i | N_i^{(\infty)}) &= \Pr(C_i > w_{i1} + w_2 | N_i^{(\infty)}) \\ &= \exp\{-\Lambda_0(w_{i1}) + e^{\beta}[\Lambda_0(w_{i1} + w_2) - \Lambda_0(w_{i1})]\} , \end{aligned}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$. This can be estimated by fitting a Cox model to the observed data, with $z_i(t) = I(N_i(t^-) > 0)$ as a covariate. Other regression models for C_i can similarly be considered.

The methodology above can be used to consider the distribution of any subsequent gap time W_j, by treating W_j as W_2 and $T_{j-1} = W_1 + \cdots + W_{j-1}$ as W_1 in the estimators above. Although one might like to consider marginal probabilities $S_j(w) = \Pr(W_j > w)$, the best that can be done is to estimate $H_j(t, w) = \Pr(T_{j-1} \leq t, W_j > w)$ for pairs (t, w) that satisfy $t + w < C_{\max}$. We can also consider joint distributions such as $F(w_1, w_2, w_3)$ by using the approach of van der Laan et al. (2002) described above. Variance estimation or confidence intervals are most conveniently approached with the bootstrap.

We have not considered the estimation of marginal gap time distributions under the assumptions that they are the same. This can be satisfactorily approached using methods analogous to those above, but only if the support of the W_j is finite and if $C_{\max}$ is sufficiently large. It is preferable, and simpler, to use the methods in this section to suggest parametric forms, and then to consider the possibility of identical marginal distributions within a parametric family of models. An example of this approach, in a setting where there was little association among gap times, was given in Section 4.3.1. An illustration involving the methods in this section is given in Section 4.4.3 below.

4.4.2 Estimation for Marginal Regression Models

In many ways, the most satisfactory approach to estimation of the marginal distribution of individual gaps, conditional on baseline covariates x, is through parametric distributions as in Section 4.2.3; log-normal models are especially convenient, and rather flexible. Model checking can be undertaken through methods discussed in Section 4.4.3. Semiparametric random effects models, as in (4.16) of Section 4.3.2., also allow a straightforward assessment of the effects of covariates but the estimation of marginal probabilities such as $\Pr(W_2 > w|x)$ is awkward. Other possible approaches include the semiparametric specification of regression models for the distribution of W_j, given that $T_{1-j} \leq t$, by analogy with (4.31); Schaubel and Cai (2004b) describe an approach based on Cox models and IPCW methods. Weighted estimators based on ranks have also been proposed for multivariate location-scale models for the log gap times, $\log W_{ij}$ (e.g. Chang, 2004). We consider a more straightforward approach, based on conditional models, in the following section.

4.4.3 Pulmonary Exacerbations in Cystic Fibrosis

We consider some further analysis of the data discussed in Section 4.3.2, where conditional regression models for W_1 and for W_2 given W_1 were fitted. The analysis there showed a strongly significant effect of treatment on the time W_1 to a first exacerbation, but in the models for W_2 given W_1, treatment was no longer significant. It is also interesting to consider the effect of treatment on W_2 without conditioning so heavily on W_1. We considered estimation of the marginal distribution of W_2 in Section 4.3.2, and found no effect of treatment.

However, we noted that because of the heavy censoring of both W_1 and W_2, an undesirable degree of extrapolation was involved. Here, we consider the distribution of W_2, given that $W_1 \leq 100$ days, which is more reasonable. We do this in two ways.

First, we use the methodology in Section 4.4.1 to examine the distribution of W_2, given that a first exacerbation occurs before 100 days; that is, given that $W_1 \leq 100$. We chose 100 days because we wished to allow sufficient followup time that a substantial number of second exacerbations could be observed. We consider the distributions for W_2 and also for W_1 separately for each treatment group, without adjusting for the covariate FEV. This provides a valid comparison of the treatment groups, because treatment is randomly assigned to subjects and is therefore independent of FEV.

The top panel of Figure 4.2 shows the Kaplan–Meier estimates for the survivor functions $S_1(w_1) = \Pr(W_1 > w_1)$, for the two treatment groups. As expected from the analysis of Section 4.3.2, there is a clear indication that time to a first exacerbation tends to be longer for persons receiving rhDNase. The bottom panel of Figure 4.2 shows estimates of $\Pr(W_2 > w_2 | W_1 \leq 100)$, given by (4.31). Unlike in the case of first exacerbations, there is now no indication of a difference between the rhDNase and placebo groups. We remark that $\Pr(W_2 > w_2 | W_1 \leq 100)$ is estimable only for $W_2 \leq C_{\max} - 100$, as discussed in Section 4.4.1. In this study almost all subjects were followed for close to 169 days, but a very few were followed longer, so that for the placebo group, for example, $C_{\max} = 196$ days. The estimates beyond $w_2 = 69$ days in the bottom panel are consequently very imprecise and the divergence of the two estimates in that region does not indicate a significant difference.

We can take another look at W_2 by using the conditional regression models fitted in Section 4.3.2 to estimate marginal probabilities, $\Pr(W_2 > w_2 | W_1 \leq L, x)$ where L is a specified value and x contains specified values for the treatment and FEV covariates. In principle, we can use the log-normal model shown in Table 4.2 to estimate the marginal distribution for W_2, given x, as discussed in Section 4.3.2. However, we cannot assess the validity of the model for values (w_1, w_2) with $w_1 + w_2$ larger than about 170 days, so we once again consider instead the conditional distribution of W_2, given that $W_1 \leq L$, with L chosen to be 100 days.

Letting $y_1 = \log w_1$, $y_2 = \log w_2$, and $y_L = \log L$, we have via the log-normal model in Table 4.2 that

$$\Pr(W_2 > w_2 | W_1 \leq L, x) = \frac{\int_{-\infty}^{y_L} \bar{F}_N\left(\frac{y_2 - x'\beta_2 - \beta_{23} y_1}{\sigma_2}\right) \sigma_1^{-1} f_N\left(\frac{y_1 - x'\beta_1}{\sigma_1}\right) dy_1}{F_N\left(\frac{y_L - x'\beta_1}{\sigma_1}\right)}, \tag{4.35}$$

where $x = (1, x_1, x_2)'$, and $f_N(\cdot)$, $F_N(\cdot)$, $\bar{F}_N(\cdot)$ are, respectively, the density, distribution, and survivor functions for the standard normal distribution. The parameter values are taken from Table 4.2, as follows: $\widehat{\beta}_1 = (5.40, 0.43, 0.022)'$,

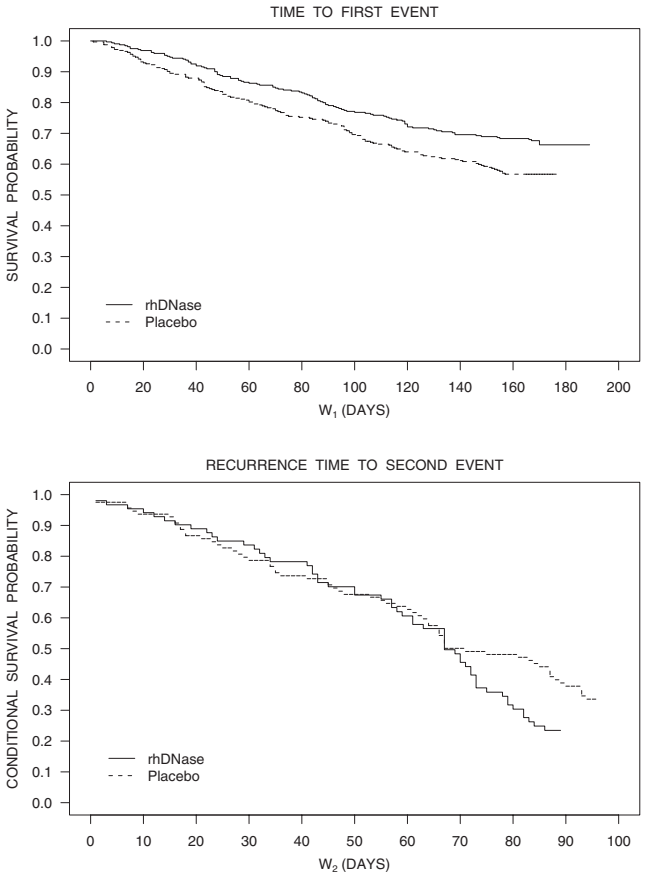


Fig. 4.2. Estimates of survivor functions for times to pulmonary exacerbations W_1 and for W_2, given $W_1 \leq 100$, for two treatment groups.

$\widehat{\sigma}_1 = 1.45$, $\widehat{\beta}_2 = (3.21, -0.23, -0.005)'$, and $\widehat{\beta}_{23} = 0.42, \widehat{\sigma}_2 = 1.23$. Estimated conditional probabilities are shown in Table 4.3 for $L = 100$ and selected values of w_2, for persons with $x_2 = 0$ (the mean FEV value across all subjects), $x_1 = 0$ (placebo), and $x_1 = 1$ (rhDNase), respectively.

We also show in Table 4.3 estimates of $\Pr(W_2 > w_2 | W_1 \leq 100, x)$ that are based on the Cox proportional hazards model given in Table 4.2. For this model we have

$$\Pr(W_2 > w_2 | W_1 \leq L, x) = \tag{4.36}$$
$$\frac{\int_0^L \exp\{-\Lambda_{20}(w_2) e^{x'\beta_2 + \beta_{23} w_1}\} \exp\{-\Lambda_{10}(w_1) e^{x'\beta_1}\} e^{x'\beta_1} d\Lambda_{10}(w_1)}{1 - \exp\{-\Lambda_{10}(L) e^{x'\beta_1}\}},$$

where $x = (x_1, x_2)'$. To estimate (4.36) we plug in the parameter estimates given in Table 4.2: $\widehat{\beta}_1 = (-0.38, -0.021)'$, $\widehat{\beta}_2 = (0.36, 0.001)'$, $\widehat{\beta}_{23} = -0.014$, and the generalized Nelson–Aalen estimates for the baseline hazard functions $\Lambda_{10}(w)$ and $\Lambda_{20}(w)$, which are given in the output from the S-PLUS or R function `coxph`. The resulting estimate of the numerator in (4.36) then becomes

$$\sum_{i=1}^{m} \left[\frac{I(w_{i1} \le L, \delta_{i1} = 1) \exp\{-\widehat{\Lambda}_{20}(w_2) e^{x'\widehat{\beta}_2 + \widehat{\beta}_{23} w_{i1}}\} e^{x'\widehat{\beta}_1}}{\sum_{\ell=1}^{m} I(w_{\ell 1} \ge w_{i1}) e^{x'_\ell \widehat{\beta}_1}} \right],$$

where m is the total number of subjects in the study.

Table 4.3. Estimated probabilities $\Pr(W_2 > w_2 | W_1 \le 100)$ for persons with average FEV, in the rhDNase and placebo treatment groups.

	rhDNase Group		Placebo Group	
w_2	Log-Normal	PH	Log-Normal	PH
20	.894	.876	.917	.909
40	.758	.741	.800	.807
60	.647	.634	.699	.721
80	.559	.537	.616	.639

We observe that the estimated conditional survival probabilities under the two models are similar. The survival probabilities are slightly higher for the placebo group, but the difference is not significant. We observe as well that the probabilities shown, for a person with average FEV, are similar to those shown in Figure 4.2, where survival is averaged over the FEV covariate.

4.5 Left Truncation of First Gap Times and Initial Conditions

In the notation of Section 1.4.3 and subsequently, we think of $t = 0$ as the starting time for an individual's event process, and assume that in a study the individual is observed over a time interval $[\tau_0, \tau]$ with $\tau_0 \ge 0$. In some settings the process of interest may actually have been going on for some period of time before the individual enters a study. For example, Sections 1.2.3, 1.5.2, and 3.8.2 all describe settings in which individuals have chronic conditions that make them susceptible to recurrent infections or outbreaks of symptoms. In this type of situation we frequently do not know the precise time at which the process started. In this case it is convenient for a prospective study to define $\tau_0 = 0$ as the start of followup, which will provide information

on $\{N(t), 0 \leq t\}$. The event process may, however, have started earlier and we may have information about it over some time period prior to τ_0. Such information is part of the initial conditions $H(0)$ for the process. In fact, the selection of an individual for a study sometimes depends on prior event history, which then should be included in $H(0)$.

Similar situations occur in many areas and indeed, for observational studies it is the norm that the event process for an individual has started prior to her selection for a study. In this section we consider some of the implications for analysis when the event process has been under way prior to the start of followup. We consider the special case of a simple renewal process first, and then more general processes where the focus is on gap times.

4.5.1 Initial Conditions and Renewal Processes

We assume that an individual is followed over the time interval $[0, \tau]$, so that the times of events $t_1 < t_2 < \cdots$ are potentially observed. In addition, we assume the process has been under way before $t = 0$, and we let $t_0 \leq 0$ denote the time of the last event in $(-\infty, 0]$. If there is an intervention (e.g. the assignment of a treatment) at $t = 0$ that affects the event process, then we may decide to ignore t_0 in defining the first gap time as $W_1 = T_1$. However, if there is no such intervention then we may prefer to define the first gap time as $W_1 = T_1 - t_0$. Then, W_1 is left-truncated, because it must satisfy the condition $W_1 \geq -t_0$. The benefit of this definition of W_1 is that when the event process (before and after $t = 0$) is a renewal process, it can be related to the distribution of $W_2 = T_2 - T_1$, $W_3 = T_3 - T_2$, and so on. We consider this situation here.

If the process is a renewal process as described, the essential information in $H(0^+)$ is just t_0, and

$$\Pr\{T_1 > t | H(0^+)\} = \exp\left\{-\int_0^t h(u - t_0)du\right\} = \frac{S(t - t_0)}{S(-t_0)}, \quad (4.37)$$

where $S(w)$ is the common survivor function for gap times. The distribution of $W_1 = T_1 - t_0$ is left-truncated at $-t_0$; given $H(0)$, W_1 satisfies the condition $W_1 \geq -t_0$. The term *delayed entry* is also sometimes used in this context: the entry of the individual to the study is "delayed" for a time $-t_0$ after the event at t_0 occurs. Of course, if $t_0 = 0$ then the start of followup is an event time and there is no left truncation. We assume from now on in this section that $t_0 < 0$.

If the time $t_0 < 0$ of the last event before time $t = 0$ is missing, there are two options. One is to ignore the incomplete first gap and to treat the individual's followup as starting from the time t_1 of the first event after $t = 0$, as far as estimation of $S(w)$ is concerned. This does not create any bias, because we are effectively redefining τ_0 to equal t_1, and the choice $\tau_0 = t_1$ is a stopping time. Thus, the likelihood function (4.2) starting with $w_{i1} = t_{i2} - t_{i1}$

is valid. This approach may involve a significant loss of information when followup of individuals is short enough that few events are observed. The other option is to impose a distributional assumption on the times of events occurring before $t = 0$. Let $f_0(t)$ denote the event rate at time $t < 0$. We can think of $f_0(t)dt$ as the probability an individual has an event in $[t, t + dt)$. Then, the marginal density of $T_1 > 0$ is

$$f_1(t_1) = \frac{\int_{-\infty}^{0} f_0(t) f(t_1 - t) dt}{\int_{-\infty}^{0} f_0(t) S(-t) dt}, \tag{4.38}$$

where $f(w)$ and $S(w)$ are the density and survivor functions for the gap times. The difficulty with this approach is the need to specify $f_0(t)$; it is necessary to have reliable information on which to base this. In the case where $f_0(t) = c$ for $t < 0$ and where W has finite mean μ, (4.38) reduces to

$$f_1(t_1) = \frac{1}{\mu} S(t_1). \tag{4.39}$$

This is known as the *equilibrium forward recurrence time* density, and is the density of the time to the first event after $t = 0$ in a renewal process that started an arbitrarily long time prior to $t = 0$; see Problems 2.8 and 4.10.

Sometimes those conducting a study may ascertain the times $t_0 > t_{-1} > t_{-2} \ldots$ of one or more events prior to τ_0 and wish to treat these as responses. This is termed a retrospective study, and care must be taken in the analysis; we discuss such studies in Section 7.3. The following example considers the points discussed above, including the information in t_0.

Example 4.5.1

Suppose the gap time distribution is taken to be parametric, with survivor function $S(w; \theta)$ and density $f(w; \theta)$. Consider a random sample of n individuals who are selected at chronological time $\tau_o = 0$, with the times $t_{i0} \leq 0$ of their most recent events being available. If the individuals are then followed prospectively until chronological times τ_i, the time t_{i1} of the next event is observed if and only if $t_{i1} \leq \tau_i$. Let $t_i = \min(t_{i1}, \tau_i)$ and $\delta_{i1} = I(t_{i1} \leq \tau_i)$; the portion of the likelihood function that is based on these first events after time τ_0 is then, from (4.37),

$$L(\theta) = \prod_{i=1}^{m} \frac{f(t_i - t_{i0}; \theta)^{\delta_i} S(t_i - t_{i0}; \theta)^{1-\delta_i}}{S(-t_{i0}; \theta)}. \tag{4.40}$$

There will also be additional terms for any subsequent gap times observed over the intervals $(0, \tau_i]$, as per (4.2).

The likelihood (4.40) is unavailable if the times t_{i0} are unknown. As discussed, one option is then to ignore t_{i1}, $i = 1, \ldots, m$ and (4.40), and to base

the likelihood on events occurring after these times. In settings where there are very few events over $(0, \tau_i]$ for the individuals, this may severely limit the amount of information and it may be desirable to adopt a rate function $f_0(t_0)$ for events prior to $t = 0$. In this case information about $f(w;\theta)$ can be recovered from t_{i1}, $i = 1, \ldots, m$ via (4.38). However, this information depends on the assumed $f_0(t_0)$, whose validity cannot be checked with the data on hand. Therefore, this approach should be used only when there is reasonable external information on which to base $f_0(t_0)$, and the sensitivity of results to changes in $f_0(t_0)$ should be examined.

A special case that is often considered is where the event processes have been under way a long time prior to time $t = 0$, and where the selection of individuals is independent of their prior event history. This "equilibrium" setting (see Problem 2.8) is consistent with $f_0(t_0) = c$ $(t_0 < 0)$, and we find as in (4.39) that the marginal distribution of T_{i1} is then

$$f_1(t_{i1};\theta) = \frac{1}{\mu(\theta)} S(t_{i1};\theta) \qquad t_{i1} > 0\,,$$

where $\mu(\theta) = E(W_{i1}) = \int_0^\infty S(w;\theta)dw$ is the expected time between successive events (i.e. the average gap time). The likelihood function based on the observations (t_i, δ_i), $i = 1, \ldots, m$ is then

$$L_1(\theta) = \prod_{i=1}^{m} \frac{1}{\mu(\theta)} S(t_i;\theta)^{\delta_i} G(t_i;\theta)^{1-\delta_i}\,, \tag{4.41}$$

where $G(t;\theta) = \int_t^\infty S(w;\theta)dw$.

A third possibility is that $f_0(t)$ is assumed known, and t_{i0}, $i = 1, \ldots, m$ are observed. In that case one might wish to treat both t_{i0} and t_{i1} as responses. Their joint density gives the terms in the following likelihood,

$$\prod_{i=1}^{m} \frac{f_0(t_{i0}) f(t_i - t_{i0};\theta)^{\delta_i} S(t_i - t_{i0};\theta)^{1-\delta_i}}{\int_{-\infty}^{0} f_0(t) S(-t;\theta)dt}$$

and in the special case where $f_0(t) = c$ $(t < 0)$ this becomes

$$L_2(\theta) = \prod_{i=1}^{m} \frac{1}{\mu(\theta)} f(t_i - t_{i0};\theta)^{\delta_i} S(t_i - t_{i0};\theta)^{1-\delta_i}\,. \tag{4.42}$$

This is an example of the use of retrospective information (i.e. the t_{i0}) as responses. The likelihood $L_2(\theta)$ contains more information than $L(\theta)$ or $L_1(\theta)$, if $f_0(t) = c$ is valid. In the case of exponential gap times, for example, where $S(w;\theta) = \exp(-w/\theta)$ and $\mu(\theta) = \theta$, we find that

$$L(\theta) = L_1(\theta) = \prod_{i=1}^{m} \frac{1}{\theta^{\delta_i}} \exp(-t_i/\theta)$$
$$L_2(\theta) = \prod_{i=1}^{m} \frac{1}{\theta^{1+\delta_i}} \exp(-(t_i - t_{i0})/\theta)\,.$$

These give asymptotic variances for $\sqrt{m}(\widehat{\theta} - \theta)$ as the inverse of the Fisher information $I(\theta) = m^{-1}E\{-\partial^2 \log L(\theta)/\partial\theta^2\}$, which are (see Problem 4.9)

$$I(\theta) = I_1(\theta) = \frac{E(\delta_i)}{\theta^2}, \qquad \text{and} \qquad I_2(\theta) = \frac{1 + E(\delta_i)}{\theta^2}\,. \tag{4.43}$$

In the case where there is no censoring of T_{i1}, the information in $L_2(\theta)$ is twice that in $L(\theta)$ or $L_1(\theta)$. This arises from the fact that the backward recurrence time $-T_{i0}$ has an exponential distribution with mean θ, just as T_{i1} does. See Problem 4.9 for some additional discussion.

When covariates are present, the case where values t_{i0} are missing is more difficult to handle if the covariates affect the distribution of T_{i0}. We turn to the question of initial conditions in such more complex situations next.

4.5.2 Initial Conditions and General Gap Time Analysis

As seen earlier in this chapter, gap time analysis is reasonably straightforward when the time τ_{i0} at which the ith individual is selected and followup starts is the beginning of their event process. However, as noted above, there are many instances where this is not the case. For example, in observational followup studies on persons with psychiatric disorders or chronic disease, each person typically has some history of events prior to τ_{i0}. Furthermore, the selection of an individual for followup at time τ_{i0} may depend on her covariate and process history $H_i(\tau_{i0})$. It is important to consider how the initial conditions $H_i(\tau_{i0})$ influence the event process over $t \geq \tau_{i0}$, and to what extent the information in $H_i(\tau_{i0})$ is available. This can affect what choices for modeling and analysis are feasible; for example, we have seen in Section 4.5.1 that if the time t_{i0} of the last event prior to τ_{i0} is unavailable then additional assumptions or discarding of information are necessary even if the process is a renewal process.

It is difficult to give a completely general discussion, in view of the varying degrees of information that may be available about individuals, and different types of study objectives. We consider essentially the same setup as in the preceding section: the ith individual is followed over the time period $\tau_{i0} \leq t \leq \tau_i$, where with no essential loss of generality we may take $\tau_{i0} = 0$. Information available in $H_i(\tau_{i0}^+)$ often includes the time t_{i0} of the last event prior to τ_{i0}^+, and for simplicity we assume that any other previous event history is contained in a vector z_{i1} of fixed covariates that is observed at $\tau_{i0} = 0$. Let $W_{ij} = t_{ij} - t_{i,j-1}$ $(j = 1, 2, \ldots)$ denote the gap times starting with the event at t_{i0}; sometimes we may choose to define W_{i1} alternatively as t_{i1}, but for the discussion here we define it as stated. We assume that the distribution of $W_{i1}, W_{i2}, \ldots$, is the focus of analysis.

Assuming that t_{i0} and z_{i1} are available, we can consider a likelihood contribution of the form (4.40), based on $\Pr(W_{i1}|z_{i1}, t_{i0}, W_{i1} > -t_{i0})$. If t_{i1} is observed (i.e. if $t_{i1} \leq \tau_i$) then an additional term based on W_{i2} is available and so on, giving a likelihood function of the form (4.10), except with the

terms for W_{i1} reflecting the conditions $W_{i1} > -t_{i0}$. A key requirement is that the model for each W_{ij} condition on enough history $w_i^{(j-1)}$, in addition to z_{i1}, to provide a satisfactory representation of the event process.

In many settings we wish to relate the models for W_{i1} and for later gaps $W_{i2}, W_{i3}, \ldots$. We have seen in Section 4.5.1 how missing information can create problems. When baseline covariates x_i are present and t_{i0} is missing, the approach described there is more complicated, because one should allow the distribution of t_{i0} to depend on x_i. There may, however, be little empirical basis for doing this, except possibly when x_i is categorical and takes only a few values.

Models incorporating random effects pose even greater problems. Random effects are commonly introduced in models as a way to represent heterogeneity among individuals, caused by covariates or other factors which influence individual event processes but are unobserved. A key assumption in most models is that the random effects are independent of observable covariates and other factors such as t_{i0} and $H_i(\tau_{i0}^+)$, given the covariates. This cannot be maintained under the general types of initial conditions discussed here. For example, suppose that W_{i1} and t_{i0} are not thought to be independent, but that there is a random effect u_i, with density function $g(u)$ in the study population, which may make such independence reasonable, as follows: given z_{i1} and u_i, W_{i1} and t_{i0} are independent. For simplicity, we suppress z_{i1} notationally and let $f_1(w|u_i)$ and $S_1(w|u_i)$ denote the density and survivor functions for W_{i1}, given u_i and z_{i1}. Then, the conditional density for W_{i1} given z_{i1}, t_{i0} and that $W_{i1} > -t_{i0}$, is

$$\Pr\left(w_{i1}|z_{i1}, t_{i0}, W_{i1} > -t_{i0}\right) = \frac{\int_{-\infty}^{\infty} f_1\left(w_{i1}|u\right) f_0\left(t_{i0}|u\right) g(u)du}{\int_{-\infty}^{\infty} S_1\left(-t_{i0}|u\right) f_0(t_{i0}|u)g(u)du}\,, \qquad (4.44)$$

because t_{i0} can in general be expected to depend on u_i. Note that (4.44) can be rewritten as

$$\int_{-\infty}^{\infty} \frac{f_1(w_{i1}|u)}{S_1(-t_{i0}|u)} g_1(u|H_i(0^+))du\,, \qquad (4.45)$$

where

$$g_1\left(u|H_i(0^+)\right) = \frac{f_0\left(t_{i0}|u\right) S_1\left(-t_{i0}|u\right) g(u)}{\int_{-\infty}^{\infty} f_0\left(t_{i0}|u'\right) S_1\left(-t_{i0}|u'\right) g(u')du'}$$

is the density of u_i given the initial conditions $H_i(0^+)$, which include the information on t_{i0} and that $W_{i1} > -t_{i0}$. Therefore, random effects modeling should not ignore information in $H_i(0^+)$.

In circumstances like those described, it is perhaps best to avoid random effects, and to handle issues such as possible dependence of W_{i1} on t_{i0}, by including t_{i0} as a covariate in the density for W_{i1}. When t_{i0} is missing, one option is to disregard W_{i1} in the analysis, and begin with the next gap time W_{i2}. As discussed earlier, this amounts to redefining $\tau_{i0} = t_{i1}$. In doing so, one should allow for the distinct possibility that W_{i2} may not be independent

of t_{i1} and so include t_{i1} as a covariate in the model for W_{i2}. A second option is to consider analyses of $W_{i1}, W_{i2}, \ldots$ as in Section 4.2.1 or Section 4.4, with W_{i1} defined as t_{i1}. A negative aspect of this approach is that the distributions of W_{i1} and later gap times are not linked, whereas we might wish to do so in some contexts. An example involving these issues is given in Section 6.7.2.

4.6 Bibliographic Notes

Methods for gap time analysis in renewal processes are essentially just the methods of survival analysis, which are discussed at length by Kalbfleisch and Prentice (2002), Lawless (2003a), and others. Certain technical issues arise in the justification for the semiparametic analysis of the Cox model (Dabrowska et al., 1994; Lawless et al., 2001), although it was used for some time before rigorous justification was provided (e.g. Gail et al., 1980; Prentice et al., 1981). Similarly, technical issues arise in a rigorous justification of the Kaplan–Meier and Nelson–Aalen estimators for a common gap time distribution (Gill, 1980; Andersen et al., 1993, Sections 10.1 and 10.2).

The assumption of independent gap times in a renewal model is a strong one, and methods based on independence are markedly nonrobust to departures from independence. Cox and Lewis (1966) discussed methods for checking independence when no covariates are present. More recently, a wide variety of gap time models that include independence models as special cases has been proposed. Conditional methods described in Section 4.2 are once again based on survival analysis. Early examples based on the Cox model were given by Gail et al. (1980) and Prentice et al. (1981), with rigorous justification for such methodology given by Dabrowska et al. (1994). Chang and Wang (1999) and Therneau and Grambsch (2000) provided other examples of gap time analyses using Cox models. Frailty or random effects models were discussed by Aalen and Husebye (1991) and Follman and Goldberg (1988). Clayton (1994) and Pickles and Crouchley (1994) surveyed random effects in the general context of event history analysis; Xue and Brookmeyer (1996), Cook et al. (1999), Lawless and Fong (1999), and Fong et al. (2001) considered random effects models in slightly more complex repeated gap time settings. Chang (2004) considered rank-based methods for accelerated failure time frailty models. Multivariate survival distributions have received much recent attention (e.g. Joe, 1997; Hougaard, 2000; Lawless, 2003a, Ch. 11). Examples with gap time analysis were given by Pena et al. (2001) and He and Lawless (2003).

Nonparametric estimation of marginal survivor functions for different gap times has been discussed by Visser (1996), Wang and Wells (1998), Wang and Chang (1999), Schaubel and Cai (2004a), and Lin et al. (1999), whose approach was described in Section 4.4.1. Van der Laan et al. (2002) considered more general problems of marginal estimation using inverse-probability-of-censoring weights, and also discussed how to increase efficiency. The IPCW idea itself originated with Robins and Rotnitzky (1992) and Robins (1993).

Gomez et al. (2006) applied this approach to the estimation of a second gap time, given the first, and Chang and Tzeng (2006) considered left truncation of first gap times. Huang (2002), Chang (2004), and Strawderman (2005) considered marginal accelerated failure time models for gap times; Huang and Chen (2003) considered proportional hazards. Schaubel and Cai (2004b) considered estimation via semiparametric Cox regression models for separate gap times, without making assumptions about association among an individual's gap times.

Left truncation and initial conditions issues for renewal processes and gap time models have been discussed by Guo (1993), Aalen and Husebye (1991), Lawless and Fong (1999), Wang (1999), and others. There is considerable discussion of these topics and general issues of modeling in the econometrics and social sciences literatures, where the analysis of times between events such as repeated pregnancies or job losses is of interest. See, for example, Chapter 2 of Heckman and Singer (1985), Heckman and Singer (1986), and Hamerle (1991). Nonparametric estimation of survivor functions under left truncation was reviewed by Lawless (2003a, Section 3.5.1 and p. 138). Nonparametric estimation of the gap distribution in a renewal process when the start time for the first gap is unknown has been extensively studied, starting with Cox (1969), Laslett (1982), and Vardi (1982a,b). Asgharian et al. (2002) provided a review and references.

4.7 Problems and Supplements

4.1. Weibull lifetime distributions (Lawless, 2003a, Ch. 1) have hazard functions

$$h(t;\alpha,\gamma) = \frac{\gamma}{\alpha}\left(\frac{t}{\alpha}\right)^{\gamma-1} \qquad t \geq 0\,, \tag{4.46}$$

where $\gamma > 0$ and $\alpha > 0$ are parameters. If T has hazard function (4.46) then $Y = \log T$ has an extreme value location-scale distribution with survivor function

$$S(y;\mu,\sigma) = \Pr(Y \geq y;\mu,\sigma) = \exp\{-\exp(\tfrac{y-\mu}{\sigma})\} \qquad -\infty < y < \infty\,,$$

where $\mu = \log\alpha$ and $\sigma = \gamma^{-1}$.

a. For the bowel motility data discussed in Example 4.1.1, use survival analysis software to fit separate Weibull models to (i) first cycles (gaps), and (ii) second and later cycles. Assume that gap times are mutually independent within and between subjects. Compare the survivor functions for the fitted models with the Kaplan–Meier estimates for first and for subsequent gaps, and with the log-normal survivor functions from model (ii) in Example 4.1.1.
b. Fit a Weibull model analogous to the log-normal Model A in Section 4.3.1; in this case $Y_{ij} = \log W_{ij}$ has an extreme-value distribution, given $y_i^{(j-1)}$.

[Sections 4.1, 4.2]

4.2. Parametric multiplicative intensity models for gap times are sometimes useful, and can be fitted using general optimization software (Appendix B). In particular, consider the model (4.11), with the baseline hazard functions specified parametrically as $h_{0j}(w;\alpha_j)$. The log-likelihood function from (4.10) is then

$$\ell(\alpha,\beta)=\sum_{i=1}^{m}\left\{\sum_{j=1}^{n_i}\log h_{ij}(w_{ij};\alpha_j,\beta_j)-\sum_{j=1}^{n_i+1}H_{ij}(w_{ij};\alpha_j,\beta_j)\right\}, \qquad (4.47)$$

where

$$\begin{aligned} h_{ij}(w;\alpha_j,\beta_j)&=h_0(w;\alpha_j)\exp(z_{ij}'\beta) \qquad (4.48)\\ H_{ij}(w;\alpha_j,\beta_j)&=\int_0^w h_{ij}(u;\alpha_j,\beta_j)du\,. \end{aligned}$$

Fit Weibull models for first and second gap times for the pulmonary exacerbation data of Section 4.3.2, where $z_{ij}=(x_{i1},x_{i2},w_{i1}I(j=2))$ and

$$h_{0j}(w;\alpha_j)=\left(\frac{\gamma_j}{\alpha_j}\right)\left(\frac{w}{\alpha_j}\right)^{\gamma_j-1} \qquad j=1,2\,.$$

Compare the results with those for the Cox proportional hazards model in Table 4.2.

Note: You can use general optimization software to obtain parameter estimates. Alternatively, you can use survival analysis software for accelerated failure time (AFT) models, because the Weibull model is both an AFT and a proportional hazards model (Lawless, 2003a, Section 6.1). In AFT form, we consider the equivalent location-scale models (4.14) for $Y_{ij}=\log W_{ij}$, with extreme value errors. Then $\beta_{0j}=\log\alpha_j$, $\sigma_j=\gamma_j^{-1}$ and the regression parameters $\beta_j=\beta_j^{\rm AFT}$ in (4.14) are related to the regression parameters $\beta_j=\beta_j^{\rm PH}$ in (4.11) by $\beta_j^{\rm AFT}=-\beta_j^{\rm PH}/\gamma_j$.

[Section 4.2]

4.3. Parametric multiplicative intensity models of the form (4.11) with $h_{0j}(w;\alpha_j)$ taken to be piecewise-constant are sometimes useful, and they provide insight into methodology for the semiparametric models of Section 4.2. They have

$$h_{0j}(w;\alpha_j)=\alpha_{jr} \qquad a_{j,r-1}<w\le a_{jr} \qquad (4.49)$$

$(r=1,\ldots,R_j)$, where $0=a_{j0}<a_{j1}<\cdots<a_{jR_j}$ are specified cutpoints, with a_{jR_j} greater than or equal to the largest observed jth gap time. Often, we might take the cutpoints to be the same across $j=1,2,\ldots$. Such models can, with appropriately chosen cutpoints, approximate underlying baseline intensities rather well. They are also easy to handle and, with large R_j, provide

inferences about the regression parameter vectors β_j in (4.11) that are close to those given by the Cox likelihood functions (4.12). In addition, they produce estimates of the cumulative baseline hazard functions $H_{0j}(w;\alpha_j)$ that are close to the generalized Nelson–Aalen estimates (4.13). In fact, letting the R_j become arbitrarily large, and the distances $a_{jr} - a_{j,r-1}$ between successive cutpoints become arbitrarily small, we obtain the semiparametric results (4.12) and (4.13).

a. Working from (4.11), (4.47), and (4.48) with $h_{0j}(w;\alpha_j)$ given by (4.49), show that the likelihood equations $\partial\ell(\alpha,\beta)/\partial\alpha_{jr} = 0$ give

$$\tilde{\alpha}_{jr}(\beta_j) = \frac{\sum_{i=1}^{m} I(j \le n_i) I(a_{j,r-1} < w_{ij} \le a_{jr})}{\sum_{i=1}^{m} I(j \le n_i + 1) e^{z'_{ij}\beta_j} \Delta_{ijr}}, \tag{4.50}$$

where

$$\Delta_{ijr} = \int_{a_{j,r-1}}^{a_{jr}} I(w_{ij} \ge u) du$$

is the length of the intersection of $(a_{j,r-1}, a_{jr}]$ and $(0, w_{ij}]$. Note that the numerator of (4.50) is the number of jth gap times falling into the interval $(a_{j,r-1}, a_{jr}]$ and if $\beta = 0$ the denominator is the total time at risk for a jth event occurring in the interval $(a_{j,r-1}, a_{jr}]$.

b. By inserting (4.50) into (4.47) we obtain a sum of profile log-likelihood functions for each β_j, which can be maximized to give estimates $\widehat{\beta}_j$ following which $\widehat{\alpha}_{jr}$ can be obtained from (4.50). Show that if the $R_j \to \infty$ and $\max |a_{jr} - a_{j,r-1}| \to 0$, then the profile likelihood function for β_j approaches (4.12).

c. Show that the estimates of the cumulative baseline hazard functions

$$H_{0j}(w) = \int_0^w h_{0j}(u;\alpha_j) du$$

are

$$\widehat{H}_{0j}(w) = \sum_{r=1}^{R_j} \left\{ \frac{e_{jr}}{\sum_{\ell=1}^{m} \Delta_{jr}(w_{\ell j}) \exp(z'_{\ell j}\widehat{\beta}_j)} \right\}, \tag{4.51}$$

where e_{jr} is the numerator of (4.50), and

$$\Delta_{jr}(w) = \int_{a_{j,r-1}}^{a_{jr}} I(w \ge u) du\,.$$

Show that (4.51) approaches (4.13) under the limit described above.

[Section 4.2; Lawless et al., 2001; Lawless, 2003a, Section 7.4]

4.4. Consider the random effects model (4.16) in the case where the u_i have gamma distributions with mean 1 and variance ϕ, with density function (2.28). Show that the likelihood contribution for an individual with n_i observed gap times w_{ij} $(j = 1, \ldots, n_i)$ and a final censored gap time w_{i,n_i+1} is

$$L_i = \left\{ \prod_{j=1}^{n_i} h_{ij}(w_{ij}) \right\} \frac{\Gamma(n_i + \phi^{-1})}{\Gamma(\phi^{-1})} \frac{\phi^{n_i}}{\left\{ 1 + \phi \sum_{j=1}^{n_i+1} H_{ij}(w_{ij}) \right\}^{n_i + \phi^{-1}}} ,$$

where

$$h_{ij}(w) = h_{0j}(w) \exp(x'_{ij}\beta_j) \quad \text{and} \quad H_{ij}(w) = H_{0j}(w) \exp(x'_{ij}\beta_j) .$$

Models where the $h_{0j}(w)$ are specified parametrically can be fitted using general optimization software.

[Section 4.2.2]

4.5. As noted following (4.20), certain copula models are closely related to random effects models, and (4.20) is in fact related to a model with gamma random effects. In particular, consider the multiplicative model where, given u_i, the gap times w_{ij} $(j = 1, 2, \ldots, J)$ are independent with hazard functions $u_i h_{ij}(w_j)$.

a. If u_i has a gamma distribution with density (2.28), show that the joint distribution of $W_{i1}, \ldots, W_{iJ}$ has a survivor function of the form (4.20), where
$$S_{ij}(w_j) = \{1 + \phi H_{ij}(w_j)\}^{-\phi^{-1}} . \tag{4.52}$$

b. Show that if $h_{ij}(w)$ has the multiplicative form in Problem 4.4, then the marginal distribution (4.52) does not have a hazard function of multiplicative form. Moreover, the marginal distributions involve ϕ. This shows that one must be clear whether covariate effects are being considered conditional on a random effect u_i, or unconditionally.
A model (4.20) in which the marginal hazard functions for w_{ij} are multiplicative is an alternative to the random effects model above. This would use $S_{ij}(w_j) = \exp\{-H_{ij}(w_j)\}$, where

$$H_{ij}(w_j) = H_{0j}(w_j) e^{x'_{ij}\beta_j} .$$

[Section 4.2; Lawless, 2003a, Section 11.2]

4.6. Use the general results for multivariate normal models (4.21), given in (4.22) and the line following it, to obtain the conditional means and variances m_{ij} and v_{ij} for Model B in Section 4.3.1. Give an alternative derivation by considering the distributions of u_i given $w_{i1}, \ldots, w_{ij}$ for successive $j = 1, 2, \ldots$.

[Sections 4.2.2, 4.3.1]

4.7. McGilchrist and Aisbett (1991) discussed data on recurrent times to infection around the catheter for kidney dialysis patients. They gave data for only the first two waiting (gap) times to infection, as shown in Table 4.4. Either of the gap times may be censored, because catheters were sometimes removed for reasons other than infection. Covariates of interest were also given: age (years), sex (1 = male, 2 = female) and types of kidney disease (labeled 0, 1, 2, 3).

Assess the effects of covariates and the shape of the distribution of time to infection; note that times to infection for the same patient might display association even after conditioning on covariates.

[Section 4.2]

4.8. Testing for trends in gap times is sometimes of interest, and can be carried out in ways analogous to approaches in Section 3.7.1. For example, suppose that the null hypothesis is that events follow a renewal process with gap time hazard function $h_0(w)$ for each of m independent individuals. Consider now the expanded model where the intensity function is given by (4.24), which we rewrite as the hazard function for W_{ij},

$$h_{ij}(w) = h_0(w) \exp\left(z_i'(t_{N_i(t^-)} + w)\beta\right) . \tag{4.53}$$

To develop trend tests we typically choose a scalar covariate $z_i(t)$ that represents trend, and then test that $\beta = 0$ in (4.53). For simplicity, let us assume that $z_i(t_{j-1} + w) = z_{ij}$; that is, $z_i(t_{N_i(t^-)} + w)$ has a fixed value z_{ij} during the jth gap $(t_{i,j-1}, t_{ij}]$. This is very useful; for example, we might use $z_{ij} = t_{i,j-1}$ or $z_{ij} = j - 1$ or other scores $z_{i1} < z_{i2} < \cdots$ designed to reflect trend.

a. Using (4.53) with
$$h_{ij}(w) = h_0(w) \exp(z_{ij}\beta) ,$$
develop a score test of $H_0 : \beta = 0$ for the Cox semiparametric model, working from the likelihood (4.25). Show that this gives the test statistic
$$U(0) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ z_{ij} - \frac{\sum_{\ell=1}^{m} \sum_{k=1}^{n_\ell+1} I(w_{\ell k} \geq w_{ij}) z_{\ell k}}{\sum_{\ell=1}^{m} \sum_{k=1}^{n_\ell+1} I(w_{\ell k} \geq w_{ij})} \right\} ,$$
and obtain a variance estimate for $U(0)$.

b. The preceding test is not valid when there is heterogeneity across individuals. Consider the model where (4.53) is generalized to give
$$h_{ij}(w) = h_{0i}(w) \exp(z_{ij}\beta) \qquad i = 1, \ldots, m ,$$

where z_{ij} is defined as in part (a). In this case each individual forms a stratum. Consider the stratified partial likelihood $L(\beta) = \prod_{i=1}^{m} L_i(\beta)$, where

$$L_i(\beta) = \prod_{j=1}^{n_i} \left\{ \frac{e^{\beta z_{ij}}}{\sum_{k=1}^{n_i+1} I\left(w_{ik} \geq w_{ij}\right) e^{\beta z_{ik}}} \right\} .$$

Use $L(\beta)$ to develop a score test for $\beta = 0$, showing that the test statistic obtained is

$$U(0) = \sum_{i=1}^{m} I(n_i > 0) \sum_{j=1}^{n_i} \left\{ z_{ij} - \frac{\sum_{k=1}^{n_i+1} I(w_{ik} \geq w_{ij}) z_{ik}}{\sum_{k=1}^{n_i+1} I(w_{ik} \geq w_{ij})} \right\} .$$

Obtain a variance estimate for $U(0)$.
For this test to be effective, we need n_i that are not too small.

[Section 4.2]

4.9. Consider the case of left-truncated initial gap times when the gap time distribution is exponential, as discussed at the end of Example 4.5.1. Use the condition $E\{\partial \log L(\theta)/\partial\theta\} = 0$ with each of $L_1(\theta)$ and $L_2(\theta)$ to show that $E(\delta_i) = E(t_i)/\theta$ in each case. Use this to obtain the Fisher information values in (4.43.)

[Section 4.4]

4.10. Let $f_0(t_0)$ be the rate function for $t_0 < 0$, as defined for (4.38). Show that the probability density function $f_0^*(t_0)$ of t_0, given both that $t_0 < 0$ and that $t_1 > 0$, is

$$f_0^*(t_0) = \frac{f_0(t_0)S(-t_0)}{\int_{-\infty}^{0} f_0(u)S(-u)du} .$$

Show also that when $f_0(t_0) = c$, the form of $f_0^*(t_0)$ is the same as $f_1(t_1)$ given by (4.39). Explain why this makes sense intuitively.

[Section 4.5]

Table 4.4. Times to two successive infections for patients on dialysis.

Patient	Times	Censoring†	Age	Sex	Disease Type
1	8, 16	1, 1	28	1	3
2	23, 13	1, 0	48	2	0
3	22, 28	1, 1	32	1	3
4	447, 318	1, 1	31-32	2	3
5	30, 12	1, 1	10	1	3
6	24, 245	1, 1	16-17	2	3
7	7, 9	1, 1	51	1	0
8	511, 30	1, 1	55-56	2	0
9	53, 196	1, 1	69	2	1
10	15, 154	1, 1	51-52	1	0
11	7, 333	1, 1	44	2	1
12	141, 8	1, 0	34	2	3
13	96, 38	1, 1	35	2	1
14	149, 70	0, 0	42	2	1
15	536, 25	1, 0	17	2	3
16	17, 4	1, 0	60	1	1
17	185, 177	1, 1	60	2	3
18	292, 114	1, 1	43-44	2	3
19	22, 159	0, 0	53	2	0
20	15, 108	1, 0	44	2	3
21	152, 562	1, 1	46-47	1	2
22	402, 24	1, 0	30	2	3
23	13, 66	1, 1	62-63	2	1
24	39, 46	1, 0	42-43	2	1
25	12, 40	1, 1	43	1	1
26	113, 201	0, 1	57-58	2	1
27	132, 156	1, 1	10	2	0
28	34, 30	1, 1	52	2	1
29	2, 25	1, 1	53	1	0
30	130, 26	1, 1	54	2	0
31	27, 58	1, 1	56	2	1
32	5, 43	0, 1	50-51	2	1
33	152, 30	1, 1	57	2	2
34	190, 5	1, 0	44-45	2	0
35	119, 8	1, 1	22	2	3
36	54, 16	0, 0	42	2	3
37	6, 78	0, 1	52	2	2
38	63, 8	1, 0	60	1	2

†1-uncensored, 0-censored.

5
General Intensity-Based Models

5.1 Time Scales and Intensity Modeling

Previous chapters have discussed models that emphasize either event counts or gaps between successive event times for individual processes. In this chapter we consider more general intensity-based modeling in which the intensity function can depend on arbitrary features of previous event history. Such models may incorporate dependence on previous gap times or event counts.

We assume that a process time scale t with a well-defined origin has been specified, as discussed in Section 1.4.1, and let $H(t) = \{N(s) : 0 \leq s < t\}$. Models that incorporate various aspects of previous event history are readily written down. For example, the intensity function

$$\lambda(t|H(t)) = \exp\{\alpha + \beta g_1(t) + \gamma I(N(t^-) > 0) g_2(B(t))\}, \qquad (5.1)$$

where $B(t) = t - T_{N(t^-)}$ is the time since the most recent event, and g_1 and g_2 are specified functions, allows the intensity at t to depend on both calendar time and the gap since the last event. It is of neither Markov nor renewal (semi-Markov) form, but includes such models as special cases with $\gamma = 0$ and $\beta = 0$, respectively.

As illustrations of where "hybrid" models involving both calendar time and gap times are useful, consider the following.

(i) In studies of recurrent infections over an extended period of time, it is desirable to allow for both calendar time trends reflecting exposure and trends in a person's susceptibility, and gap time factors that reflect adjustments to an individual's susceptibility when they experience an infection.
(ii) In models for equipment or software failures, we may wish to include calendar time trends that reflect aging or deterioration, and gap time factors that reflect repairs or adjustments to deal with previous failures.

For effective model formulation, specific frameworks that posit some primary structure are essential, especially when covariates are present. Two gen-

eral approaches are through modulated Markov or renewal processes, as discussed in Sections 2.2.2 and 2.3.1. For the former, we specify a time-varying covariate vector $z(t)$ that may include both external and internal covariates; the latter may represent features of previous event history. The intensity function is then taken to be of the form

$$\lambda(t|H(t)) = \lambda_0(t)\, g(z(t))\,. \tag{5.2}$$

Modulated renewal models are of the analogous form

$$\lambda(t|H(t)) = h_0(B(t))\, g(z(t))\,. \tag{5.3}$$

For either (5.2) or (5.3), log-linear specifications where $g(z(t)) = \exp(z'(t)\beta)$ are often useful, provided covariates are satisfactorily defined. It may be noted that (5.1) is a model in both the families (5.2) and (5.3).

In choosing one of (5.2) and (5.3) we decide to take either chronological time (t) or gap time (w, or $B(t)$) as the primary time scale in the sense that the reference or "baseline" process, against which the effects of external covariates or event history are measured, has an intensity of the form $\lambda_0(t)$ or $h_0(B(t))$, respectively. Which time scale we choose to emphasize depends on their relevance for the context at hand and, generally, on a desire for parsimonious models. Another important consideration is the representation of treatment or fixed covariate effects; their interpretation and estimated values depend heavily on the type of model adopted. These points are discussed further in subsequent sections.

The likelihood function from m independent event processes whose intensities depend on parameters θ is, from (2.7) or (2.55),

$$L(\theta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \lambda_i(t_{ij}|H_i(t_{ij})) \cdot \exp\left\{ -\int_0^\infty Y_i(u)\lambda_i(u|H_i(u))du \right\}, \tag{5.4}$$

where $\lambda_i(t|H_i(t))$ is the intensity for the ith process, which has $n_i \geq 0$ observed events, at times t_{ij} ($j = 1, \ldots, n_i$). As in (2.55), $Y_i(u)$ indicates whether process i is under observation at calendar time u (1 = yes, 0 = no) and we assume that the observation of the event processes is conditionally independent of the process in the sense described in Section 2.6.

For fully parametric models such as (5.1), maximum likelihood estimation and associated inference procedures can be based on (5.4), as we discuss in Section 5.2. However, semiparametric models of various types can also be formulated, and provide relief from certain types of parametric assumptions. Convenient models of this type are given by (5.2) and (5.3), with $\lambda_0(t)$ and $h_0(w)$ allowed to be arbitrary nonnegative functions. Semiparametric analysis is discussed in Sections 5.3 and 5.4.

Random effects may also be incorporated in models such as (5.2) or (5.3). Random effects induce additional dependence on previous event history in the intensity function. For example, the process given by (2.26), which is

conditionally Poisson given a random effect, has an intensity function (2.33) that depends on the number of previous events. A criticism of commonly used random effects models, though, is that the random effects are time-invariant and are too simplistic for complex processes. The main approach in this chapter is to model the intensity directly on observable quantities, rather than through unobservable random variables. Nevertheless, random effects are useful for reflecting heterogeneity across individuals due to unmeasured fixed covariates, and in the examples of Section 5.5, we consider whether there is excess variability that could be described by individual-specific random effects.

5.2 Parametric Analysis for Two Useful Models

Analysis for models that are fully specified parametrically can be based on the likelihood function (5.4). In this section we describe two rather flexible families of models and consider inference procedures for them. Software for semiparametric models of the form (5.2) or (5.3) place restrictions on the intensity and on the form of time-varying covariates, and parametric models provide additional flexibility.

We remark that in this section, and later in this chapter, multiplicative models are emphasized. Multiplicative models can be made very flexible, and the availability of straightforward maximum likelihood methodology and software makes them a convenient and natural choice in many settings. Additive models, as discussed in Section 3.4.4, are sometimes useful and could also be considered. For fully parametric additive models, maximum likelihood estimation is in principal straightforward although restrictions on parameter values that are needed to keep the intensity function positive can introduce practical complications. Semiparametric models that extend the basic model of Section 3.4.4 can also be used; Martinussen and Scheike (2006) provide a detailed development of methodology.

5.2.1 Log-Linear Intensity Models

We consider models where the intensity function can be written in the form

$$\lambda(t|H(t)) = \exp\{z'(t)\theta\}\,, \tag{5.5}$$

where $z(t) = (z_1(t), \ldots, z_p(t))'$ is a vector of observable functions that may depend on t and on features of $H(t)$, and θ is a vector of parameters. Many common models are special cases of (5.5); for example, $z(t) = (1, \log t, x'(t))'$ and $\theta = (\log\alpha, \delta, \beta')'$ give a Poisson model with intensity $\alpha t^{\delta}\exp(x'(t)\beta)$ and $z(t) = (1, \log B(t), x'(t))'$ with the same θ gives a renewal model with intensity function $\alpha B(t)^{\delta}\exp(x'(t)\beta)$. Poisson models of this general form were introduced in Section 3.2.2.

From (5.4), the log-likelihood function for θ in (5.5) is

$$\ell(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} z_i'(t_{ij})\theta - \int_0^\infty Y_i(t) \exp(z_i'(t)\theta) dt \right\} . \tag{5.6}$$

A nice property of (5.5) is that the log-likelihood (5.6) is convex under mild conditions (see Problem 5.1) and so there is a unique maximizer $\widehat{\theta}$. It is often simplest to maximize $\ell(\theta)$ using general-purpose optimization software that does not require derivatives of $\ell(\theta)$; see Appendix B. For many models numerical integration is needed to evaluate (5.6). This can be handled with standard software but for most practical purposes the following approach is satisfactory. First, noting there are often discontinuities in the functions $z_j(t)$ at the event times t_{ij}, we write

$$\int_0^\infty Y_i(t) \exp(z_i'(t)\theta) dt = \sum_{j=1}^{n_i+1} \int_{t_{i,j-1}}^{t_{ij}} \exp(z_i'(t)\theta) dt \,, \tag{5.7}$$

where t_{i0} is the start-of-observation time τ_{i0} for individual i, t_{i,n_i+1} is the end-of-observation time τ_i, and where we have assumed that individual i is observed continuously from t_{i0} to t_{i,n_i+1}. More generally, the integrals in (5.7) should be split at any discontinuity points for each $z_i(t)$ but for simplicity we assume here that jumps in $z_i(t)$ occur only at the event times t_{ij}.

Numerical integration software can be applied to the separate integrals in (5.7), or a simple but robust procedure such as the trapezoidal rule or Simpson's rule (Press et al., 1986, Ch. 4) can be used to approximate $\ell(\theta)$ to a desired degree of accuracy. Using the trapezoidal rule, we select positive integers r_{ij} and define $\Delta_{ij} = (t_{ij} - t_{i,j-1})/r_{ij}$ and constants a_{ijk} by

$$a_{ijk} = t_{i,j-1} + (k-1)\Delta_{ij} \qquad k = 1, \ldots, r_{ij} + 1 \,.$$

Then $\int_{t_{i,j-1}}^{t_{ij}} \exp(z_i'(t)\theta) dt$ is approximated as

$$\Delta_{ij} \{ .5 \exp(z_i'(t_{i,j-1})\theta) + .5 \exp(z_i'(t_{ij})\theta) + \sum_{k=2}^{r_{ij}} \exp(z_i'(a_{ijk})\theta) \}$$

and $\ell(\theta)$ is approximated by

$$\ell^A(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} z_i'(t_{ij})\theta - \sum_{j=1}^{n_i+1} \Delta_{ij} \sum_{k=1}^{r_{ij}+1} w_{ijk} \exp(z_i'(a_{ijk})\theta) \right\} , \tag{5.8}$$

where $w_{ij1} = w_{ij,r_{ij}+1} = 1/2$ and $w_{ijk} = 1$ $(k = 2, \ldots, r_{ij})$. A simple and effective way to choose Δ_{ij} is to select a small incremental value Δ and then to define $r_{ij} = \text{top } \{(t_{ij} - t_{i,j-1})/\Delta\}$, the smallest integer greater than or equal to $(t_{ij} - t_{i,j-1})/\Delta$. The value of Δ may be reduced in a sequence of calculations of $\ell^A(\theta)$ until a desired degree of accuracy is reached.

Some optimization software requires, or optionally accepts, first and second derivatives of $\ell(\theta)$. The approximation $\ell^A(\theta)$ gives easily computed approximations to $U(\theta) = \partial\ell(\theta)/\partial\theta$ as

$$U^A(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} z_i(t_{ij}) - \sum_{j=1}^{n_i+1} \Delta_{ij} \sum_{k=1}^{r_{ij}+1} w_{ijk} z_i(a_{ijk}) \exp(z_i'(a_{ijk})\theta) \right\} \quad (5.9)$$

and to $I(\theta) = -\partial^2\ell(\theta)/\partial\theta\partial\theta'$ as

$$I^A(\theta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i+1} \Delta_{ij} \sum_{k=1}^{r_{ij}+1} w_{ijk} z_i(a_{ijk}) z_i'(a_{ijk}) \exp(z_i'(a_{ijk})\theta) \right\} . \quad (5.10)$$

The main requirement for (5.8)–(5.10) is that the values of the $z_i(a_{ijk})$ be available. When $z_i(t)$ is a function of the process history there is no difficulty. As is the case in general, when $z(t)$ includes time-varying covariates that are measured only intermittently it is necessary to impute some values.

An illustration of this methodology is given in Section 5.2.4.

5.2.2 A Trend-Renewal Model

Linqvist et al. (2003) introduced a model which they term the *trend renewal process* (TRP). It is defined as follows in the absence of covariates. Let $A(t)$ be an increasing function for $t \geq 0$, and assume that for $T_0 = 0$ and successive event times T_j $(j = 1, 2, \ldots)$ the values $V_j = A(T_j) - A(T_{j-1})$ are independent and identically distributed with distribution function $F(v)$, on $v \geq 0$. This model includes Poisson processes as the special case where F is an exponential distribution with mean 1, whereupon $A(t)$ is the mean or cumulative rate function. Renewal processes are also included as the special case where $A(t) = t$. Scale factors in $A(t)$ and $F(t)$ will be confounded so we usually assume a fixed scale parameter for F, designed for example to make the mean or median of F equal to one.

This model nicely combines a trend in calendar time with a renewal after each event. The intensity function is, from the definition of the process,

$$\lambda(t|H(t)) = h[A(t) - A(t_{N(t^-)})]a(t)\,, \quad (5.11)$$

where $h(v)$ is the hazard function $F'(v)/(1 - F(v))$ for V_j and $a(t) = A'(t)$ is the derivative of $A(t)$. This model cannot be expressed in the form (5.5), and combines renewal and trend features in a different way than (5.5) does. A model of the form (5.5) incorporates the effect of trend into the intensity as a multiplicative factor. The model (5.11), on the other hand, is a time transform, or accelerated time, model; the effect of trend is expressed in the time scale $A(t)$. The TRP model might be plausible when an individual's propensity for events changes over time, but the process triggering events is

stationary, whereas the models of the form (5.5) might be used when the process triggering events has a time trend, but an individual's propensity for an event is renewed or adjusted each time an event occurs.

Fixed covariates x can be incorporated into the TRP framework by generalizing $A(t)$ to

$$A(t|x) = A_0(t)g(x)\,,$$

where $g(x)$ is a positive-valued function. Random effects can also be introduced in the same multiplicative fashion, but we do not pursue this here. Alternative specifications where the distributions $F(v) = Pr(V_i \leq v)$ depend on covariates or random effects could also be given. Dealing with time-varying covariates is more difficult: the appropriate extension to the model above is to define

$$a(t|x) = a_0(t)g(x(t)) \quad \text{and} \quad A(t|x) = \int_0^t a_0(u)g(x(u))du \tag{5.12}$$

giving models analogous to accelerated failure time models used in survival analysis (Lawless, 2003a, Section 8.2.2). In fact the model (5.11) with $t_{N(t^-)} = 0$, corresponding to the intensity for the time to the first event, is the hazard function for an accelerated failure time regression model when $A(t|x)$ is of the form (5.12).

Likelihood inference is reasonably straightforward for the TRP model. Assuming that individual i ($i = 1, \ldots, m$) is observed from $t = 0$ to the stopping time τ_i, the likelihood function can be given using (5.4) and (5.11), or in the more convenient form

$$L(\theta) = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} f[A_i(t_{ij}) - A_i(t_{i,j-1})]a_i(t_{ij}) \right\} S[A_i(\tau_i) - A_i(t_{i,n_i})]\,, \tag{5.13}$$

where $A_i(t) = A_0(t)g(x_i)$, $a_i(t) = a_0(t)g(x_i)$ and where $f(v) = F'(v)$ and $S(v) = 1 - F(v)$ are the density and survivor functions for the random variables $V_{ij} = A_i(T_{ij}) - A_i(T_{i,j-1})$.

The log of the likelihood (5.13) can be maximized using general-purpose optimization software. For example, consider a model where $A_i(t)$ takes the Weibull accelerated failure time form

$$A_i(t) = \{t/\exp(x_i'\beta)\}^{\delta}\,,$$

where we include an intercept in $x_i'\beta$ by defining $x_{i1} = 1$, and where the V_{ij} have a Weibull distribution with scale parameter one. The density and survivor function for $V_{ij} = A_i(T_{ij}) - A_i(T_{i,j-1})$ are then

$$f(v;\alpha) = \alpha v^{\alpha-1}\exp(-v^{\alpha})\,, \quad \text{and} \quad S(v;\alpha) = \exp(-v^{\alpha})\,,$$

respectively. With $\theta = (\alpha, \delta, \beta)$, the log-likelihood is readily computed for any given θ with $\alpha > 0$, $\delta > 0$.

An illustration involving TRP models is given in Section 5.2.4.

5.2.3 Model Checking

The most flexible and powerful way to assess model assumptions is by fitting expanded models within which a base model may be tested. However, informal graphical methods can provide insight, and in this regard generalized residuals analogous to those considered for Poisson processes in Section 3.7.3 are useful.

If a process has intensity function $\lambda(t|H(t))$, then under the conditions of Section 2.1 the quantities

$$E_j = \int_{T_{j-1}}^{T_j} \lambda(t|H(t))dt \qquad j = 1, 2, \dots ,$$

are independent exponential random variables with mean 1; see (2.46) and the following discussion. The E_j are determined by the event times $T_1 < T_2 < \cdots$ in an individual process, with T_0 for convenience defined as the start of observation. If we denote T_{n+1} as the end of observation, then E_{n+1} is a right-censored exponential variable. As usual it is assumed that T_0 and T_{n+1} satisfy conditions laid down in Section 2.6. The properties of the E_j imply that the transformed times $R_j = E_1 + \cdots + E_j$ are equivalent to event times in a homogeneous Poisson process. Thus, procedures described in Section 3.7.3 can be applied to the more general processes considered in this chapter. With independent processes $i = 1, \dots, m$, we can define generalized residuals $\widehat{E}_{ij}$ $(j = 1, \dots, n_i + 1)$ and $\widehat{R}_{ij}$ $(j = 1, \dots, n_i + 1)$ and use plots described in Section 3.7.3.

Martingale residuals can also be defined as in Section 3.7.3. In particular, if $\bar{N}_i(t) = \int_0^t Y_i(u)dN_i(u)$ as before, the processes

$$\widehat{M}_i(t) = \bar{N}_i(t) - \int_0^t Y_i(u)\widehat{\lambda}_i(u|H(u))du \qquad t > 0\,,$$

should be roughly mean zero processes with uncorrelated increments, for sufficiently large m and assuming the process intensities are correctly specified. Plots of residuals such as $\widehat{M}_i(\infty)$ against covariates, and test statistics similar to (3.49), can indicate lack of fit, as in Section 3.7.3.

5.2.4 An Illustration: Air-Conditioning System Failures

To illustrate the use of the models and diagnostics in the preceding sections, we consider some much-discussed data on airplane air-conditioning system failures (Proschan, 1963). The times w_j $(j = 1, \dots, 30)$ between successive failures for a single plane, denoted plane 6 by Cox and Lewis (1966, p. 6), are considered; times are in hours of operation. The w_j $(j = 1, \dots, 30)$ are

23	261	87	7	120	14	62	47	225	71
246	21	42	20	5	12	120	11	3	14
71	11	14	11	16	90	1	16	52	95

and the event times are $t_j = w_1 + \cdots + w_j \quad (j = 1, \ldots, 30)$.

In this setting, because we are considering a single plane, there is a single event process. Figure 5.1 shows a plot of the cumulative number of events $N(t)$ versus t, where we have left out the usual staircase feature. Figure 5.1 also contains a curve which is the estimate of the mean function $\mu(t) = E\{N(t)\}$ for a Poisson process model, described below.

We first fit models of the form (5.5). Figure 5.1 suggests that the rate of failures is increasing over time, and also that failures may tend to cluster somewhat. We consider two models, motivated by the plot, with intensity functions as follows.

Model 1: $\lambda(t|H(t)) = \exp\{\alpha + \beta t + \gamma z(t)\} \qquad t \geq 0\,,$
where $z(t) = I(N(t^-) > 0)(t - t_{N(t^-)})$ is the time since the last event.

Model 2: $\lambda(t|H(t)) = \exp\{\alpha + \beta t + \gamma z(t)\} \qquad t \geq 0\,,$
where $z(t) = I(N(t^-) > 0)I(t - t_{N(t^-)} \leq 20)$ is a binary covariate equalling 1 for up to 20 hours after a failure.

Maximum likelihood estimates are obtained by maximizing the log-likelihood (5.6), which in the case of both Models 1 and 2 has a closed form. The estimates, with standard errors in parentheses obtained from the inverse of the observed information matrix, are as follows.

Model 1: $\widehat{\alpha} = -4.87(0.57) \quad \widehat{\beta} = 0.000865(0.000400) \quad \widehat{\gamma} = -0.00148(0.00326)$
$\ell(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) = -149.37.$

Model 2: $\widehat{\alpha} = -5.17(0.48) \quad \widehat{\beta} = 0.000801(0.000392) \quad \widehat{\gamma} = 0.746(0.371)$
$\ell(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}) = -147.52.$

In Model 1, the renewal term parameter γ is not statistically significant but in Model 2 it is, at the 5% level of significance. This agrees with the impression conveyed by Figure 5.1, which shows instances of rather closely clustered failures, but because the specific form of $z(t)$ in Model 2 was guided by an examination of Figure 5.1, the results should be considered provisional. Nevertheless, there is a suggestion that the probability of a new failure is increased temporarily after a failure; this phenomenon is not uncommon in repairable systems.

As an additional check on the need for a renewal term in the intensity, we fit a nonhomogeneous Poisson process,

Model 3: $\lambda(t|H(t)) = \exp(\alpha + \beta t) \qquad t \geq 0\,,$
$\widehat{\alpha} = -5.02(0.47) \qquad \widehat{\beta} = 0.000917(0.000386)$
$\ell(\widehat{\alpha}, \widehat{\beta}) = -149.47.$

The estimated mean function $\mu(t) = \exp(\widehat{\alpha})(\exp(\widehat{\beta}t) - 1)/\widehat{\beta}$ is shown in Figure 5.1, and mimics the data well. A likelihood ratio test of Model 3 versus Model 1 does not indicate evidence against Model 3, but a test of Model 3 versus Model 2 gives a likelihood ratio statistic of $2(-147.52+149.47) = 3.902$. Using χ^2_1, this gives a $p-$value just under .05, in close agreement with a Wald test based on $\widehat{\gamma}$ in Model 2. Finally, another check on the Poisson model can be made using the score test statistics (3.47) described in Section 3.7.3. With the parameter estimates for Model 3, and defining $z(t)$ as in Model 2, we find the observed value of the statistic (3.47) to be $S(0) = 5.222$, and the corresponding variance term is $V(0) = 6.045$; the scaled test statistic is then $S(0)/V(0)^{1/2} = 2.12$. Here there is only one process, but fairly many events, so we compute a $p-$value for the test of the Poisson model using the standard normal distribution, as described in Section 3.7.3 and as used here for the Wald test. This gives a $p-$value of 0.034, in agreement with the Wald and likelihood ratio tests.

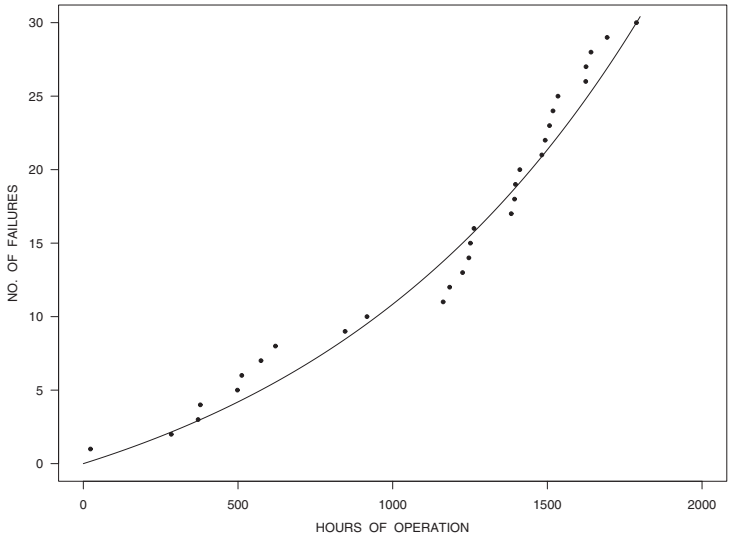


Fig. 5.1. Cumulative air-conditioning failures and estimated mean function.

Let us also consider the trend renewal process model of Section 5.2.2. We fit a model with $a(t) = \exp(\alpha + \beta t)$ and $A(t) = \exp(\alpha)(\exp(\beta t) - 1)/\beta$ and $F(t)$ chosen to be a Weibull distribution with unit scale, so that $F(t) = 1 - \exp(-t^{\delta})$, where $\delta > 0$ is an unknown parameter. Using the likelihood function (5.13), we find maximum likelihood estimates for this model, termed

Model 4 as follows.

Model 4: TRP with $a(t) = \exp(\alpha + \beta t)$ and $S(t) = \exp(-t^{\delta})$
$\widehat{\alpha} = -5.01(0.52)$ $\widehat{\beta} = 0.000947(0.000426)$ $\widehat{\delta} = 0.918(0.130)$
$\ell(\widehat{\alpha}, \widehat{\beta}, \widehat{\delta}) = -149.25.$

The special case $\delta = 1$ gives an exponential distribution for $S(t)$, in which case Model 4 becomes the nonhomogeneous Poisson process, Model 3. A Wald test of the hypothesis $\delta = 1$ based on $\widehat{\delta}$ (giving $Z = (\widehat{\delta} - 1)/\text{s.e.}(\widehat{\delta}) = 0.63$ and $Z^2 = 0.40$) and a likelihood ratio test of Model 3 versus Model 4 (giving the likelihood ratio statistic $\Lambda = 2(-149.25 + 149.47) = 0.44$) are in close agreement, and provide no evidence against the Poisson process.

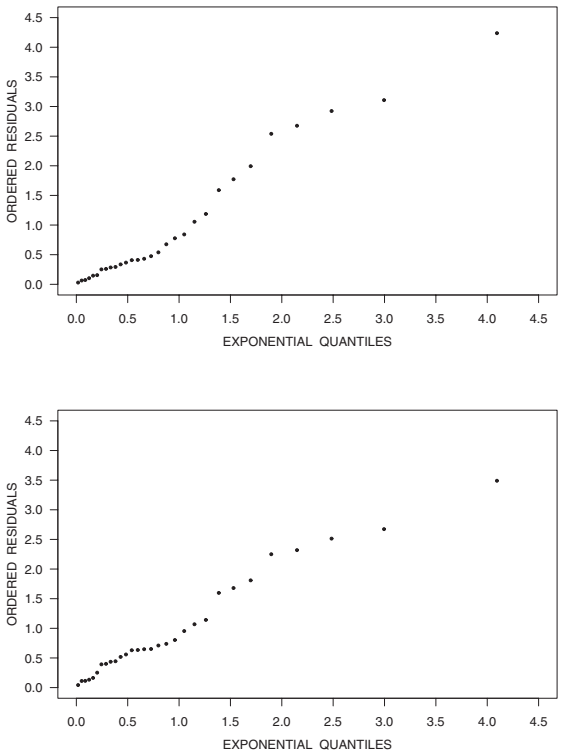


Fig. 5.2. Exponential residual plot for Model 3 (top panel) and Model 2 (bottom panel).

In this example we see that neither Model 1 nor Model 4 provides evidence against the Poisson process, which is Model 3. However, Model 2 confirms a

visual impression from Figure 5.1, and suggests that there may be a temporary elevation of the failure intensity following a failure (and its subsequent repair). This indicates the need for a close examination of event patterns, flexibility in modeling, and a careful assessment of model fit. We can also assess the fit of Model 4 by plotting residuals discussed in Section 5.2.3. Figure 5.2 shows an exponential probability plot of the residuals $\widehat{E}_j$ described in Section 5.2.3, for Models 2 and 3. The $\widehat{E}_j$ are defined as follows for the two cases, with the estimates being the maximum likelihood estimates for each model.

$$\text{Model 3: } \widehat{E}_j = \{\exp(\widehat{\alpha} + \widehat{\beta} t_j) - \exp(\widehat{\alpha} + \widehat{\beta} t_{j-1})\}\widehat{\beta}^{-1}$$

$$\begin{aligned}\text{Model 2: } \widehat{E}_1 &= \{\exp(\widehat{\alpha} + \widehat{\beta} t_1) - 1\}\widehat{\beta}^{-1} \\ \widehat{E}_j &= \{\exp(\widehat{\alpha} + \widehat{\gamma} + \widehat{\beta} t_j) - \exp(\widehat{\alpha} + \widehat{\gamma} + \widehat{\beta} t_{j-1})\}\widehat{\beta}^{-1} \\ &\qquad \text{if } j \geq 2 \text{ and } t_j - t_{j-1} \leq 20 \\ \widehat{E}_j &= \{\exp(\widehat{\alpha} + \widehat{\beta} t_j) - [1 + e^{\widehat{\gamma} - 20\widehat{\beta}} - e^{\widehat{\gamma}}] \exp[\widehat{\alpha} + \widehat{\beta}(t_{j-1} + 20)]\}\widehat{\beta}^{-1} \\ &\qquad \text{if } j \geq 2 \text{ and } t_j - t_{j-1} > 20.\end{aligned}$$

Figure 5.2 gives plots of the ordered $\widehat{E}_{(j)}$ against the standard exponential quantiles, defined as $-\log\{1 - (j - 0.5)/30\}$, for $j = 1, \ldots, 30$. The top panel, for Model 3, has a departure from linearity that might reflect a mixture of two distributions for the E_j instead of a single standard exponential. The bottom panel, for Model 2, is more nearly linear and does not provide strong evidence against Model 2. Similar residual plots for Models 1 and 4 show patterns like that for Model 1; this is in line with our finding that Models 1 and 4 do not improve significantly on the Poisson process, Model 3.

5.3 Semiparametric Markov Analysis

5.3.1 Models with Dependence on Prior Counts

Modulated Markov models of the general form (5.2) often provide parsimonious descriptions of event processes when a calendar time such as age, time on study, or time from some other specified origin is a natural scale. Frequently it is desirable to consider dependence of the intensity on the number of previous events. One way to do this is by letting $z(t) = N(t^-)$ in (5.2) and specifying a parametric form such as $g(z(t)) = \exp(\beta N(t^-))$. Models of this type, where $z(t)$ includes information on previous event history, can be fitted and analyzed exactly as for the multiplicative models in Section 3.4.2, as we discuss in Section 5.3.3.

A more general approach is obtained by considering the multistate Markov model depicted in Figure 5.3. In this model, states represent the cumulative number of events experienced, and only transitions from state k to $k + 1$ are possible. Thus the counting process $\{N_i(t), 0 \leq t\}$ also indicates the state occupied over time. If $H_i(t) = \{N(s) : 0 \leq s < t\}$ then

$$\alpha_k(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta N_i(t) = 1 | N_i(t^-) = k, H_i(t))}{\Delta t} \tag{5.14}$$

is the intensity function for transitions from state k to $k+1$ under a Markov model. If $Y_{ik}(t) = I(N_i(t^-) = k)$, then the intensity function for the event process can be written as

$$\lambda_i(t|H_i(t)) = \sum_{k=0}^{\infty} Y_{ik}(t)\alpha_k(t)\,.$$

Under this Markov model, the intensities $\alpha_k(t)$, $k = 0, 1, \ldots$, are assumed to be functionally independent. We also let $A_k(t) = \int_0^t \alpha_k(u)du$ denote the cumulative transition intensity out of state k. If $\alpha_k(t) = \alpha_{k-1}(t)\exp(\beta)$ then we obtain the simpler model mentioned in the first paragraph of this section, for which

$$\lambda_i(t|H_i(t)) = \lambda_0(t)\exp(\beta N_i(t^-))\,, \tag{5.15}$$

with $\lambda_0(t) = \alpha_0(t)$. Note also that if $\alpha_k(t) = \alpha_0(t)$, for $t > 0$ and $k = 1, 2, \ldots$, then a Poisson model is obtained.

Fig. 5.3. A multistate representation for a recurrent event process.

The transition probability functions under the Markov model above are defined as

$$p_{k\ell}(s,t) = \Pr(N(t) = \ell | N(s) = k) \qquad 0 \le s \le t\,. \tag{5.16}$$

For any given setting there is typically some upper limit on the number of events that could be seen for an individual over the time period $0 \le t \le \tau$. For convenience, we let K denote a value greater than or equal to such a limit, and assume that $\alpha_K(t) = 0$ for $0 \le t \le \tau$. Then we write the transition probabilities as a $(K+1)\times(K+1)$ matrix $\mathcal{P}(s,t) = [p_{k\ell}(s,t)]$, where $0 \le k \le \ell \le K$ and $0 \le s \le t \le \tau$. The transition probability functions and transition intensity functions are linked via the product integral (Andersen et al., 1993, p. 93),

$$\mathcal{P}(s,t) = \prod_{(s,t]} (\mathcal{I} + \mathcal{A}(u)du)\,, \tag{5.17}$$

where $\mathcal{I}$ is the identity matrix and $\mathcal{A}(t)$ is the $(K+1)\times(K+1)$ matrix of transition rates given by

$$\mathcal{A}(t) = \begin{bmatrix} -\alpha_0(t) & \alpha_0(t) & 0 & 0 & \dots & 0 & 0 \\ 0 & -\alpha_1(t) & \alpha_1(t) & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \dots & -\alpha_{K-1}(t) & \alpha_{K-1}(t) \\ 0 & 0 & \dots & \dots & \dots & 0 & 0 \end{bmatrix}.$$

The state occupancy probabilities at time t are the probabilities for $N(t)$ and are given by the first row of $\mathcal{P}(0,t)$:

$$\Pr(N_i(t) = k | N_i(0) = 0) = p_{0k}(t) \qquad k = 0, 1, \dots, K.$$

There is considerable appeal to a multistate formulation of recurrent events because it can be used to address many natural questions that arise. Features of common interest include the probabilities above and the cumulative mean function $\mu(t)$, which can be expressed here as

$$\mu(t) = E\{N_i(t)\} = \sum_{k=1}^{\infty} k\, p_{0k}(t)\,. \tag{5.18}$$

We now discuss estimation of these quantities.

5.3.2 Markov Nonparametric Estimation

Consider a dataset of m subjects in which subject i is followed over $[0, \tau_i]$ where τ_i is a censoring time independent of $\{N_i(t), 0 \le t\}$, the event process for subject i. Let $Y_{ik}(t) = I(N_i(t^-) = k)$ indicate, as before, whether subject i is at risk of a transition at time t out of state k (i.e. is at risk of a $(k+1)$st event), $k = 0, 1, \dots$. Let $dN_{ik}(t) = 1$ if subject i makes a transition from state k to $k+1$ at time t, and $dN_{ik}(t) = 0$ otherwise. Finally, let $N_{ik}(t) = \int_0^t dN_{ik}(u)$ indicate whether a transition from k to $k+1$ occurred for subject i over $[0,t]$. As before, we denote the corresponding observable quantities so that $\bar{Y}_{ik}(t) = Y_i(t)Y_{ik}(t)$ indicates whether individual i is under observation and at risk of a $k \to k+1$ transition, $d\bar{N}_{ik}(t) = Y_i(t)dN_{ik}(t)$ indicates whether a $k \to k+1$ transition is observed at t, and $\bar{N}_{ik}(t) = \int_0^t Y_i(u)dN_{ik}(u)$ indicates whether a $k \to k+1$ transition was observed over the interval $[0,t]$.

The Nelson–Aalen estimate of the $k \to k+1$ cumulative intensity function is

$$\widehat{A}_k(t) = \int_0^t \frac{I(\bar{Y}_{\cdot k}(u) > 0) d\bar{N}_{\cdot k}(u)}{\bar{Y}_{\cdot k}(u)}\,, \tag{5.19}$$

where $d\bar{N}_{\cdot k}(u) = \sum_{i=1}^m d\bar{N}_{ik}(u)$ and $\bar{Y}_{\cdot k}(u) = \sum_{i=1}^m \bar{Y}_{ik}(u)$, $k = 0, 1, \dots$, and where $0/0$ is defined as 0. This is the same form as the Nelson–Aalen estimate (3.17) for the cumulative rate (mean) function for a simple recurrent event, applied here to the $(k+1)$st event, but $A_k(t)$ is not the expected number of $(k+1)$st events. As discussed in Section 3.4.1, $\widehat{A}_k(t)$ estimates

$A_k(t)$ only when $\bar{Y}_{\cdot k}(s) > 0$ for $0 \le s \le t$. If $\bar{Y}_{\cdot k}(s) > 0$ only for $\tau_0^{(k)} \le s \le t$, then (5.19) estimates $A_k(t) - A_k(\tau_0^{(k)}) = A_k(t; \tau_0^{(k)})$ and more generally it estimates only increments in $A_k(t)$. This point is especially important here, because individuals are not at risk of a $(k+1)$st event until they experience a kth event.

Here, the expected number of $(k+1)$st events by time t is the same as the probability of at least $k+1$ events occurring by time t. The *Aalen–Johansen estimate* (Aalen and Johansen, 1978) of the transition probability matrix provides estimates of these probabilities, and is obtained from (5.17) as

$$\widehat{\mathcal{P}}(s,t) = \prod_{(s,t]} (I + \widehat{\mathcal{A}}(u)du)\,, \tag{5.20}$$

where $\widehat{\mathcal{A}}(u)$ is the estimated matrix of transition rates, with $d\widehat{A}_k(u) = I(\bar{Y}_{\cdot k}(u) > 0)d\bar{N}_{\cdot k}(u)/\bar{Y}_{\cdot k}(u)$. The estimate $\widehat{\mathcal{P}}(s,t)$ can be easily computed because the product integral in (5.20) becomes a product over a finite number of unique time points where at least one transition was observed. The prevalence functions are then estimated from $\widehat{\mathcal{P}}(0,t)$ as $\widehat{p}_{0k}(t)$. Andersen et al. (1993; Section IV.4.1.3) provide variance estimates for (5.20) under the Markov model; see also Præstgaard (1991).

If $\bar{Y}_{\cdot k}(s) > 0$ only for $\tau_0^{(k)} \le s \le t$ then the observed data do not give any information on $A_k(t)$ for $t < \tau_0^{(k)}$. Asymptotic properties of (5.20) depend on there being information about $A_k(u)$ for $s \le u \le t$ (Andersen et al., 1993, Section IV.4.1.2) but with real data we should note any limits on the information. Our use of (5.20) here is solely for estimation of the $p_{0k}(t)$ and for this purpose the estimates, with the convention that $0/0 = 0$ in (5.19), seem reasonable. In the case where there is no censoring of event histories before time t, (5.20) gives the empirical estimates $\widehat{p}_{0k}(t) = \bar{Y}_{\cdot k}(t)/m = Y_{\cdot k}(t)/m$ (see Problem 5.7).

Aalen et al. (2001) and Datta and Satten (2001) have pointed out that the Aalen–Johansen estimator of the prevalence functions, although formally justified under a Markov assumption, provides a consistent estimate of the state occupancy probabilities (prevalence functions) for non-Markov multistate processes provided the censoring is independent of the processes. That is, the censoring process cannot depend on features of the event history for non-Markov models; for Markov models the censoring process can be adaptive. To see this, note that the $\alpha_k(t)$, defined in (5.14) for a Markov model, can also apply to general models, if interpreted as rates

$$\alpha_k(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr\left(\Delta N_i(t) = 1 | N_i(t^-) = k\right)}{\Delta t}\,.$$

For the Nelson–Aalen estimates $d\widehat{A}_k(s) = \widehat{\alpha}_k(s)ds$ to be valid more generally for $dA_k(s) = \alpha_k(s)ds$, we need

$$E\left\{\bar{Y}_{ik}(s)[dN_{ik}(s) - dA_k(s)]\right\} = E\left\{Y_i(s)\, Y_{ik}(s)(dN_{ik}(s) - dA_k(s))\right\}$$

to equal zero. This holds if $\{Y_i(t), 0 \leq t\}$ is independent of $\{N_i(t), 0 \leq t\}$ or if $E\{dN_i(s)|Y_i(s) = 1, N_i(s^-) = k\} = E\{dN_{ik}(s)|N_i(s^-) = k\}$.

Glidden (2002) discusses robust variance estimation of the prevalence functions obtained from the estimated transition probability matrix and describes how to construct simultaneous confidence bands for the prevalence functions via simulation. In our context these give confidence bands for $p_{0k}(t) = \Pr(N_i(t) = k)$, for $t > 0$.

Another estimate of $p_{0k}(t)$ can be based on estimates of the survivor functions for T_{ik}, $k = 1, 2, \ldots$. Specifically, although $\Pr(N_i(s) = k) = \Pr(T_{ik} \leq s) - \Pr(T_{i,k+1} \leq s)$, if $\widehat{F}_k(s)$ is the Kaplan–Meier estimate of the cumulative distribution function for T_{ik}, $(k = 1, 2, \ldots)$, then $\widehat{\Pr}(N_i(s) = k) = \widehat{F}_k(s) - \widehat{F}_{k+1}(s)$. Such an estimate is obviously robust to distributional assumptions for the event process under independent censoring. Pepe (1991) provides robust variance estimates under completely independent censoring. Because this is a strictly marginal approach (i.e. there is no conditioning on the process history for estimation of $F_k(s)$, $k = 1, 2, \ldots$), inconsistent estimates of $\Pr(N_i(s) = k)$ are obtained under adaptive (e.g. state-dependent) censoring and so this approach should be used with caution. The effects of adaptive censoring are discussed in Section 7.2.

The mean function $\mu(t)$ can be estimated with the Nelson–Aalen estimate (3.17), which is valid for Poisson processes but also more generally, provided censoring is completely independent of the event processes. A second approach is to use estimates of $p_{0k}(t)$ in (5.18); if the Aalen–Johansen estimates from (5.20) are used, this has the added advantage of being valid under event-dependent censoring, provided the Markov process of Figure 5.3 is satisfactory.

5.3.3 Models with Covariates

Modulated Markov models are of the multiplicative form (5.2). In the present section we allow the baseline intensity in the multiplicative model to vary with k. If a $p \times 1$ covariate vector x is available, we may specify common covariate effects for the intensity functions $\alpha_{ik}(t)$, $k = 0, 1, 2, \ldots$ for $(k+1)st$ events as

$$\alpha_{ik}(t) = \alpha_{k0}(t) \exp(x_i'(t)\beta) \qquad k = 0, 1, 2, \ldots .$$

This is sometimes referred to as the *stratified Andersen–Gill model* when $\alpha_{k0}(t)$ has an unspecified form, because subjects advance to different "strata" as they experience new events. The model is not Poisson like the ordinary Andersen–Gill model, but the methods of analysis of Section 3.4.3 may be adapted.

Different covariate effects may also be allowed as

$$\alpha_{ik}(t) = \alpha_{k0}(t) \exp(x_i'(t)\beta_k) \qquad k = 0, 1, 2, \ldots .$$

The former model is fit in S-PLUS or R using the `coxph` function with the `strata(·)` option, whereas the latter arises from specifying a "strata by covariate" interaction or, equivalently, by fitting separate models for each k,

$k = 0, 1, \ldots$. In terms of the intensity function we can write the more general expression as

$$\lambda_i(t|H_i(t)) = \sum_{k=0}^{\infty} Y_{ik}(t)\alpha_{k0}(t)\exp(x_i'(t)\beta_k)\,. \tag{5.21}$$

There is of course an upper limit for the number of strata with any particular dataset and we indicate this by K in what follows. We remark that the covariates in (5.21) may include functions of previous event history, but this should be done in accordance with the assumption of different baseline intensity functions for each k.

Assume that individual i is observed over $[0, \tau_i]$. The product integral representation of the likelihood function (2.7) reveals that for the stratified model with common regression coefficients, profiling arguments can be applied as in Sections 3.4.2 and 3.4.3 to give a partial or profile likelihood for the model with common β as

$$L(\beta) = \prod_{i=1}^{m}\prod_{k=0}^{K-1}\left\{\frac{\exp(x_i'(V_{i,k+1})\beta)}{\sum_{l=1}^{m}\bar{Y}_{\ell k}(V_{i,k+1})\exp(x_\ell'(V_{i,k+1})\beta)}\right\}^{\delta_{i,k+1}}, \tag{5.22}$$

where $V_{ik} = \min(T_{ik}, \tau_i)$ and $\delta_{ik} = I(V_{ik} = T_{ik})$, $k = 1, 2, \ldots, K$; $i = 1, \ldots, m$. Technically K can be taken equal to the largest number of events observed across all subjects, but if very few individuals reach that maximum number, we may choose a smaller K. The data from subjects observed to have more than K events will then be censored at T_{iK}, which is a stopping time, so likelihood-based inferences remain valid. An alternative that is sometimes used is to assume all events after the kth have the same intensity as the kth events, and include them in the kth stratum. Note that because $Y_{ik}(u)$ is zero unless $N_i(u^-) = k$, (5.22) can only be maximized using packages for survival analysis that allow for stratification and delayed entry in the Cox model; `coxph` in S-PLUS and R does this.

With different regression coefficients β_k for each stratum, we can factor the overall likelihood into functionally independent parts and maximize these separately. The kth part gives the partial or profile likelihood function

$$L_k(\beta_k) = \prod_{i=1}^{m}\left\{\frac{\exp(x_i'(V_{i,k+1})\beta_k)}{\sum_{\ell=1}^{m}\bar{Y}_{\ell k}(V_{i,k+1})\exp(x_\ell'(V_{i,k+1})\beta_k)}\right\}^{\delta_{i,k+1}}. \tag{5.23}$$

Likelihood ratio tests can be carried out for $H_0 : \beta_k = \beta$ by assessing $2[\sum_{k=0}^{K-1}\ell_k(\widehat{\beta}_k) - \ell(\widehat{\beta})]$ against a $\chi^2_{(K-1)p}$ distribution, where $p = \dim(\beta_k)$, $\ell_k(\widehat{\beta}_k) = \log L_k(\widehat{\beta}_k)$, and $\ell(\widehat{\beta}) = \log L(\widehat{\beta})$.

Estimates of the baseline cumulative intensity functions $A_{k0}(t) = \int_0^t a_{k0}(u)du$ are given by the generalized Nelson–Aalen estimates (3.24):

$$\widehat{A}_{k0}(t) = \sum_{i=1}^{m} \left\{ \frac{I(V_{i,k+1} \le t, \delta_{i,k+1} = 1)}{\sum_{\ell=1}^{m} \bar{Y}_{\ell k}(V_{i,k+1}) \exp(x'_\ell(V_{i,k+1})\widehat{\beta}_k)} \right\} \qquad k = 0, 1, \ldots, \quad (5.24)$$

where $\widehat{\beta}_k$ is either the common $\widehat{\beta}$ obtained from (5.22) or the distinct $\widehat{\beta}_k$ obtained from (5.23), depending on the model.

5.3.4 Analysis of Outbreaks Due to Herpes Simplex Virus

Section 3.8.2 reported on analyses of data from a two-period crossover trial of patients with herpes simplex virus infection, based on Poisson process models. Here we consider data from patients receiving episodic treatment with valacyclovir during period I of this trial to illustrate the use of Markov models described in this section. This provides insight into the recurrent infections process. Figure 5.4 gives the empirical distribution function for the time of crossover to period II in this trial, which varies across subjects from 161 to 196 days (median 170); because this analysis is restricted to period one data, this represents the duration of followup. First we consider some descriptive analysis, ignoring covariates.

The cumulative marginal rates for transitions from state k to state $k+1$ ($k = 0, 1, 2, 3$) are estimated based on (5.19) and plotted in Figure 5.5. They feature some separation but we must remember the effect on $\widehat{A}_k(t)$ of times when $\bar{Y}_{\cdot k}(t) = 0$. The functions for the third and fourth events cross those for the first and second, but this is essentially because most persons are not at risk for the third or fourth events for small values of t. In effect, only $A_k(t) - A_k(t_k^{\min})$ for $t > t_k^{\min}$ is estimated for $k = 1, 2, 3$, where $t_k^{\min}$ is the smallest kth event time in the data. In comparing Nelson–Aalen estimates we should examine their "slopes", which estimate the event rate functions. Doing this in Figure 5.5, we see that the predominant message is that a person with one event at time t is at slightly higher risk of a new event at time t than a person with zero events, and that persons with two or three prior events are at still higher risk of a new event.

Figure 5.6 shows the estimated prevalence functions $\widehat{p}_{0k}(t)$ based on (5.20) as well as based on the difference in Kaplan–Meier estimates, $\widehat{F}_k(t) - \widehat{F}_{k+1}(t)$, for successive states, which for $k = 1, 2, 3, 4$ estimate the proportion of patients with exactly 1, 2, 3, and 4 events over time. These two estimates agree closely over most of period I but differ slightly towards the end of observation where the censoring times begin to occur. In fact, if there is no censoring before time t, both estimates of $p_{0k}(t)$ can be shown to equal $\sum_{i=1}^{m} Y_{ik}(t)/m$. The proportion of subjects with one outbreak rises fairly quickly over the first month and then begins to decline as patients go on to have additional outbreaks.

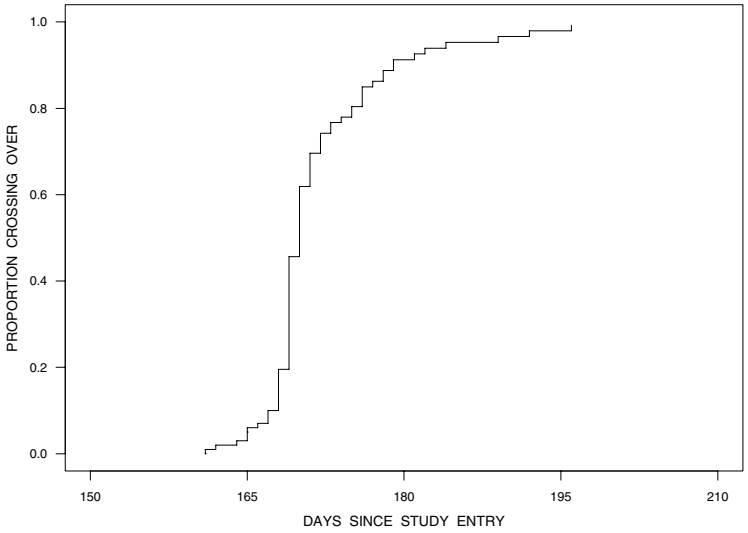


Fig. 5.4. Empirical distribution of the crossover time in herpes trial for the episodic-suppressive sequence group.

Figure 5.7 gives the Nelson–Aalen estimates (3.17) of the mean functions for patients taking episodic treatment with valacyclovir for the outbreak of symptoms, as well as the corresponding estimate for patients taking valacyclovir daily to suppress outbreaks; these estimates are truncated versions of Figure 3.4 in Section 3.8.2. Superimposed on these are the respective mean function estimates obtained from (5.18) by replacing $p_{0k}(t)$ with its estimate obtained from (5.20). These two methods can be shown to give identical estimates in the absence of censoring (see Problem 5.7). The mean functions diverge rapidly reflecting a large effect of suppressive therapy on reducing the occurrence of outbreaks.

The suggestion that the risk of a new infection at time t increases with the number of prior infections is consistent with the analysis of Section 3.8.2, where there was evidence of extra-Poisson variation. However, some fixed covariates were also considered there, and these may account for some or all of the apparent increased risk for later events seen here.

Table 5.1 contains estimates obtained by fitting various multiplicative Markov regression models to the data on outbreaks from patients on episodic therapy for period I. As in Section 3.8.2, the baseline covariates include age (years), sex (males versus females), type of herpes simplex virus (HSV type II versus type I), and the number of recurrences in the year before entry to the study. The Poisson, or Andersen–Gill, model (Model 1) suggests that males

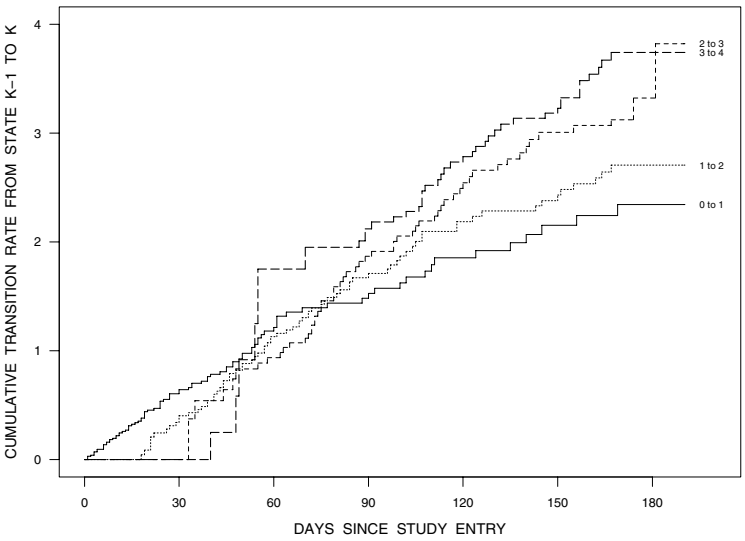


Fig. 5.5. Estimated cumulative transition rates from herpes trial (Period I - episodic arm).

have a lower rate of outbreaks than females ($p = 0.053$), and that those with a higher number of recurrences in the previous year have a higher risk of recurrence on study ($p = 0.017$); the other covariate effects are not significant. When controlling for the cumulative number of prior outbreaks since study entry (Model 2) as in (5.15), the findings are broadly similar but sex and the number of recurrences in the previous year are no longer significantly associated with event occurrence; the coefficient for $N_i(t^-)$ is significant, suggesting that the greater the cumulative number of events since study entry the greater the risk of subsequent events. This is in line with the evidence from the plots of the cumulative transition rates in Figure 5.5 and also with the analysis of the full dataset in Section 3.8.2, where there was an indication of some extra-Poisson variation when $N_i(t^-)$ was not included. Specifically, from this model there is an estimated 21.1% relative increase in the risk of future events associated with each additional event ($RR = 1.211$, 95% CI (1.112, 1.319), p <0.001). Model 3 is the negative binomial analogue of Model 2 but here the variance ϕ of the gamma random effect is estimated to be 0.678, indicating some evidence of overdispersion relative to Model 2. A likelihood ratio test of $\phi = 0$ gives an observed value of 4.96, based on a comparison of the maximum log-likelihoods for Models 2 and 3. The $p-$value based on $0.5 \cdot \Pr(\chi^2_1 > 4.96) = 0.013$.

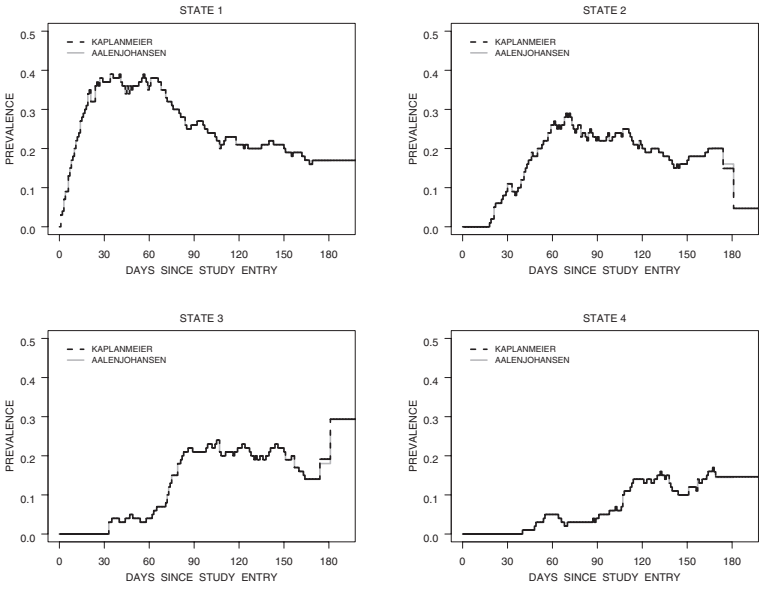


Fig. 5.6. Estimated state occupancy probabilities based on (5.20) and by taking the difference of Kaplan–Meier estimates (Period I - episodic arm).

An alternative is to generalize Model 2 with event-dependent stratification, as in Section 5.3.3. In the stratified Andersen–Gill analysis represented in (5.21) (Model 4) the simple effect of event occurrence on future events (see Model 2) is lost, but one can examine plots analogous to Figure 5.5. In Model 4 the regression coefficients and standard errors for age, sex, virus type, and previous recurrences are very similar to those in Model 2. There is no evidence of strata by covariate interactions for the covariates in this model. In Model 5 we consider possible additional effects of prior event history on the intensity function. When controlling for the duration of the previous three gaps the estimates suggest that the longer the most recent and second most recent gaps, the lower the risk of a subsequent outbreak, but conclusions concerning covariate effects are unchanged.

Tests of the multiplicative intensities assumption (Section 3.7) for the covariates in Model 5, carried out using the S-PLUS or R function `cox.zph`, give significant evidence of nonproportionality in the effect of the most recent gap time ($p = 0.016$). The corresponding diagnostic plot of Schoenfeld residuals in Figure 5.8 suggests fitting a model with an interaction between the most recent gap and $I(t < 100)$, thereby allowing the effect to be constant over the first 100 days and then fixed at a different level for the remaining time. This more general model (Model 6) is reported in the last column of Table 5.1. The tests for this model do not suggest any serious problems.

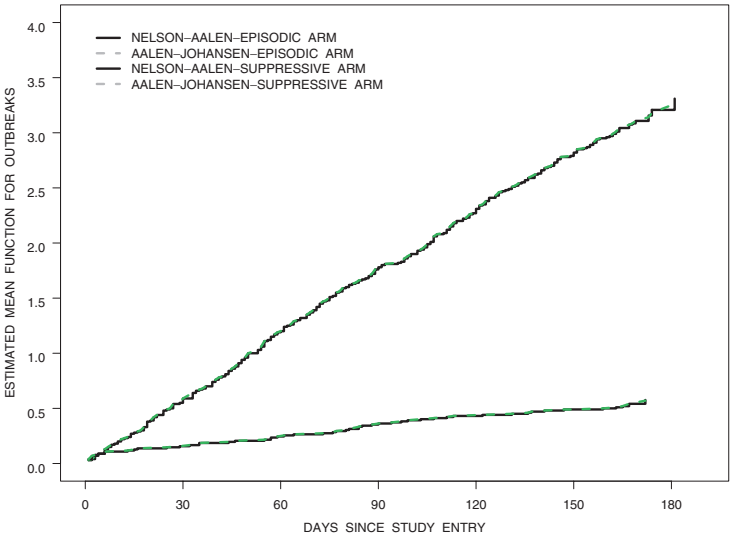


Fig. 5.7. Nelson–Aalen and Aalen–Johansen estimates of the mean functions (Period I - episodic arm and suppressive arm).

Model 6 mildly suggests a lower risk of outbreaks for males, and a higher risk of outbreaks among individuals with more recurrences in the previous year. Subjects who have had a long delay between their most recent outbreaks tend to be at lower risk of a new outbreak over the first three months following this outbreak; long gaps between the previous two outbreaks are also associated with lower risk. In the previous analysis of Section 3.8.2, which focused on treatment effects in the full study, males showed a moderate but nonsignificant decrease in risk in mixed Poisson models. The present analysis gives a more detailed examination of outbreak patterns across individuals, and gives a similar result. Such an analysis can also be applied to the full dataset.

Stratification offers a method of conditioning on the event history which is more general than including a time-dependent covariate $N_i(t^-)$, because a multiplicative effect of event occurrence is not assumed. As shown in Section 3.5.2, the introduction of gamma distributed random effects also produces intensity functions which are conditional on cumulative event counts, but the event counts modulate the intensity in a very specific way; different random effect distributions lead to different forms of intensities. Stratification allows more flexibility, but the models are still of modulated Markov form.

It must be stressed that although the types of models fitted here can shed light on the dynamics of event occurrence within individuals, the regression

Table 5.1. Regression estimates from several models for herpes outbreak for the episodic group during period I.

		Nonstratified Models					
		Model 1		Model 2		Model 3	
Covariate		EST.	S.E.	EST.	S.E.	EST.	S.E.
Age	Years	-0.001	0.006	-0.0005	0.006	-0.002	0.009
Sex	Male	-0.249	0.129	-0.205	0.129	-0.326	0.223
Virus type	II	0.155	0.115	0.130	0.116	0.205	0.212
Previous recurrences		0.059	0.025	0.038	0.025	0.084	0.046
Event count	$N_i(t^-)$	–	–	0.240	0.065	0.356	0.094
Variance ϕ		–		–		0.678	
Log-likelihood		-1408.699		-1401.651		-1398.719	
		Stratified Models					
		Model 4		Model 5		Model 6	
Covariate		EST.	S.E.	EST.	S.E.	EST.	S.E.
Age	Years	0.0001	0.006	0.001	0.006	0.002	0.006
Sex	Male	-0.201	0.129	-0.206	0.130	-0.219	0.130
Virus type	II	0.124	0.118	0.151	0.118	0.183	0.120
Previous recurrences		0.038	0.026	0.044	0.026	0.046	0.026
Previous gaps‡							
	gap1	–	–	-0.008	0.004	–	–
	gap1; t $<$ 100	–	–	–	–	-0.022	0.007
	gap1; t $\geq$ 100	–	–	–	–	-0.003	0.004
	gap2	–	–	-0.016	0.006	-0.017	0.006
	gap3	–	–	-0.013	0.008	-0.013	0.008
Log-likelihood		-1077.739		-1070.995		-1068.427	

‡ gap1 is the most recent gap, gap2 the second most recent gap, and gap3 the third most recent gap.

coefficients β_j represent different effects in different models, so their estimates may not be directly comparable. In particular, models may involve different types of conditioning on prior event history, and regression coefficients must be interpreted with other covariates and event history in the intensity held fixed.

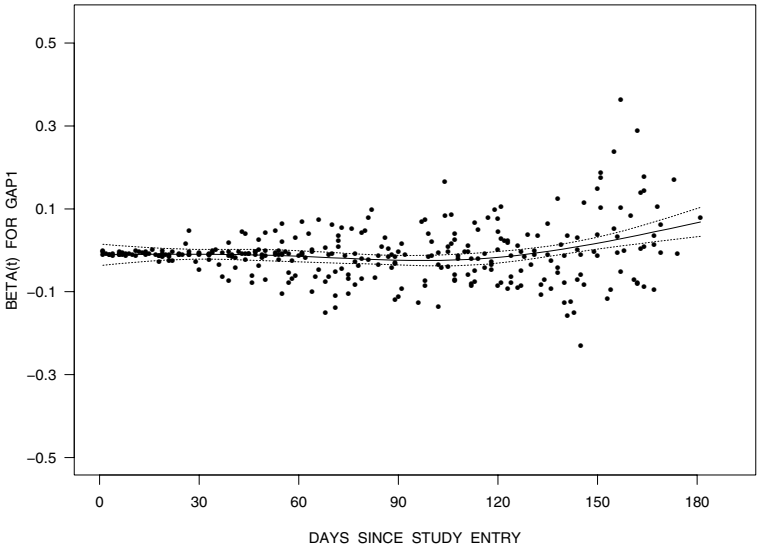


Fig. 5.8. Plot of Schoenfeld residuals with loess smooth for Model 5.

5.4 Semiparametric Modulated Renewal Analysis

5.4.1 Analysis Based on the Cox Model

The preceding section has described how analysis can be based on Markov (Andersen–Gill) models, modulated or supplemented with covariates that may depend on previous event history. This methodology treats baseline intensity functions nonparametrically, and can be implemented using software for the Cox model. In this section we discuss analogous methods for modulated renewal processes (5.3) with the intensity function of the form

$$\lambda(t|H(t)) = h_0(B(t)) \exp(z'(t)\beta)\,, \tag{5.25}$$

where $B(t)$ is the backward recurrence time, or time since the last event. This is in fact model (4.24), and models of this type were discussed in Section 4.2. They can be handled using software for the Cox model, as described there.

Let $w_{ij} = t_{ij} - t_{i,j-1}$ represent the jth gap time ($j = 1, 2, \ldots$) for individual i, where $t_{i0} = \tau_{i0}$, the start-of-observation time for individual i. We assume for convenience that $\tau_{i0} = 0$, unless it is specified otherwise. Let τ_i represent the end-of-followup time for individual i so that if n_i events are observed, we have the values w_{ij} ($j = 1, \ldots, n_i$) and, if $\tau_i > t_{in_i}$, a final censored gap time, $w^*_{i,n_i+1} = \tau_i - t_{in_i}$. For the model (5.25), the Cox partial likelihood function

for β is then given by (4.25), and the cumulative baseline hazard function $H_0(w)$ is estimated by (4.26).

The covariates $z_i(t)$ in (5.25) can include terms such as $N_i(t^-)$, allowing dependence on the previous number of events. As in the case of the modulated Markov models in Section 5.3.3, we can also use a stratified Cox model, where for $j = 0, 1, 2, \ldots$,

$$\lambda(t|H(t), N(t^-) = j) = h_{0j}(B(t)) \exp(z'(t)\beta) . \tag{5.26}$$

Different β could be used in (5.26), in which case analysis amounts to a separate treatment for each of gap times $W_1, W_2, W_3, \ldots$. A dependence on calendar time can also be included, for example, using t_{j-1} as a covariate when considering the jth gap times W_j.

It has been emphasized in Chapter 4 that the validity of analyses based on partial likelihoods such as (4.8) or (4.25) depends on the conditional independence of a gap time W_{ij} and its censoring time, given any observed covariates. Thus, if (5.26) or another model used for analysis is misspecified in the sense that W_{ij} is not conditionally independent of $C_{ij} = \tau_i - t_{i,j-1}$, given the observed values of covariates in the model, then estimates of regression parameters and baseline cumulative hazard or survivor functions may be biased. An illustration involving the estimation of survivor functions in the absence of covariates was given in Section 4.4.1. As discussed in Section 4.4, we can protect against this by incorporating terms in the covariates for W_{ij} that account for any dependence on previous event history, including gap times. If there is a desire to examine the individual gap times dependent only on baseline covariates, however, then we may need to either (i) use random effects or a joint distribution for gap times in order to model association among an individual's gap times, or (ii) use inverse probability of censoring (IPC) weights in conjunction with marginal models. These approaches were discussed in Section 4.4.

Although the emphasis in this section is on semiparametric proportional intensity models, recall that because W_{ij} $(j = 1, 2, \ldots)$ can be considered as a sequence of failure times, parametric survival analysis methods and software can also be used, as described in Chapter 4.

It should also be noted that in many contexts, especially with observational studies, a time origin for the process under study may be unknown or uncertain. In addition, as discussed in Section 4.5, the start of observation for an individual may fall between successive events. For example, in studying episodes of hospitalization or other events for persons with psychiatric disorders (e.g. Kessing et al., 1999), the process might be considered to start with the diagnosis of the disorder, but this could be made after the person has already been hospitalized. In carrying out gap time analyses, the first gap is often defined as $W_{i1} = T_{i1} - \tau_{i0}$, where τ_{i0} and T_{i1} are the calendar times of the start-of-followup and the first event after followup starts, respectively, for the ith individual. Because it is not an ordinary gap time unless an event occurs at τ_{i0}, we may wish to treat W_{i1} separately from subsequent gap times.

In some settings, the start-of-followup for an individual is taken for analysis purposes to be the time at which his first observed event occurs. As discussed in Section 4.5, doing this does not lead to any bias, assuming that the model used is appropriate.

5.4.2 Illustration: Cerebrospinal Fluid Shunts

Tuli et al. (2000) discussed an observational study on children who had internal shunts inserted surgically to deal with hydrocephalus. Such shunts drain excess cerebrospinal fluid away from the head, typically to the abdominal area, and have led to a major decline in neurological deficit and death. Shunts are designed to stay in patients indefinitely, but "failures" occur due to blockages, infections, and other conditions. In the case of a failure, the existing shunt is typically partially replaced. Tuli et al. (2000) discussed data for 839 children who had initial shunts inserted during the years 1987–1996 at one Canadian hospital. The data include the dates of the initial shunt insertion and any subsequent failures up to the end of 1997. Three primary modes of failure are indicated: Obstruction, Infection, and Other (other causes). About 70% of the failures are due to Obstruction, with the Infection and Other causes accounting for about 15% each. Information on deaths is also available; 121 patients died over the study period.

We consider analysis of the data with a view to identifying risk factors associated with failure. There are four primary types of factors,

Age: the age of the child at the time of shunt insertion. This is categorized here as age < 0, $0 \leq$ age ≤ 1 year, and age > 1 year. The age < 0 category is due to the fact that some children were born prematurely and had shunts inserted before their full-term birth date.

Etiology: cause of the hydrocephalus which necessitates the shunt. This is represented by eight categories: IVH (intraventricular hemorrhage), Men (meningitis), Adsten (acqueductal stenosis), Tumor, Trauma, MMC (myelominingocele), Other (other causes), and Con (congenital).

Shunt type: categorized as vp (ventriculoperitoneal) or other.

Concurrent surgery: whether there was other surgery at the same time as the shunt insertion.

The shunt failures for an individual are recurrent events, and some patients experience a large number. Table 5.2 shows the distribution of total shunt failures across the 839 patients, but we should remember that followup times for individuals range from about one year to eleven years. Each new shunt has its own characteristics and "lifetime" and so it is natural to focus on these lifetimes, or gap times between successive failures. There are multiple

event types, but in this section we consider shunt failure as a single type of event. We consider the separate types of failures, along with death, in Section 6.2. Lawless et al. (2001) provided additional background on these data, and observed that the effects of risk factors were for the most part similar for the different failure modes.

Table 5.2. Number of shunt failures per patient.

No. of failures	0	1	2	3	4	5	6–10	>10
No. of patients	386	209	102	49	31	16	32	14

We consider analyses of first, second, and third failures based on semiparametric proportional intensity models of the form (5.26), but we allow both the regression coefficients β_j and baseline intensity functions $h_{0j}(t)$ to vary across successive failures $j = 1, 2, \ldots$. That is, the models used are of the form (4.11), with β_j estimated from the partial likelihood (4.12) and $H_{0j}(t)$ estimated by (4.13). The factor Age is represented by two binary covariates indicating $0 \leq$ age ≤ 1 and age > 1, respectively, so that age < 0 is baseline; Etiology is represented by seven binary covariates, with Con (congenital) as baseline; Shunt type is represented by a binary covariate (= 1 if type = vp, 0 otherwise), as is Concurrent Surgery (= 1 if yes, 0 if no). Etiology is fixed, but type of shunt and concurrent surgery may vary across repeated shunt insertions.

We fit models of the form (4.11) with vectors z_{ij} which include covariates for age, etiology, shunt type, and concurrent surgery, as described. The jth gap or failure time W_j for a patient equals the age of the patient at the time shunt $j+1$ was inserted minus the age when shunt j was inserted ($j = 1, 2, \ldots$); age and time are measured in days. The age at insertion of the first shunt varies widely across patients. For $j = 2, 3$ we also considered prior failure times (w_1 for $j = 2$ and w_1, w_2 for $j = 3$) as covariates, in order to assess any association between successive times after conditioning on covariates. Doing this also makes it more plausible that censoring times for w_2 and w_3, due to end-of-followup, are independent. It should be noted, however, that death is a competing risk, so the intensity models for first, second, and third failures are interpreted in the competing risks framework for failure times (e.g. Lawless 2003a, Ch. 9). We also considered the effect of the type of the previous failure on second and third failures, but this did not prove significant. Finally, we considered age at the time of the jth shunt insertion as a covariate for W_j ($j = 2, 3$). The effects of age and of prior failure times are necessarily ambiguous because the factors are confounded (e.g. age at first shunt insertion plus time to first failure equals age at second shunt insertion). Therefore we consider two types of models:

(a) Models where "age" is the age at first surgery for $j = 1, 2, 3$, and with the previous failure time w_{j-1} (in 1000-day units) also taken as a covariate

for $j = 2, 3$

(b) Models in which "age" is the age at the time of the current ($j = 1, 2, 3$) shunt insertion, with w_{j-1} also included as a covariate for $j = 2, 3$.

Estimates of regression coefficients for the various fitted models are shown in Table 5.3. There is a substantially increased risk of first failures for young patients, with estimated relative risks ($\exp(\widehat{\beta})$) of 1.88 for the 0–1-year age category, and 2.64 for the age < 0 category that refers to premature infants. We remark that these age categories were defined following preliminary analysis, and capture the age effect well. Diagnostic checks on the proportional intensity assumption using the function `cox.zph` in S-PLUS indicated mild but not conclusive evidence that the increased risk for younger patients decreases with time since surgery.

The etiology categories all have estimated relative risk over one for first failures, relative to the baseline congenital category, but for second and third failures etiology effects, aside from IVH and Men for second failures, are not significant. The effect of shunt type is significant for first shunts only, and appears to decline thereafter. Concurrent surgery shows a consistent increased risk, but is statistically significant only for first failures. For model (a), in which age at first surgery is a covariate, a strong positive association with time to first failure is indicated for second failures, but as we might expect, the age at first surgery is then insignificant. For third failures neither the baseline age nor previous failure time is significant. For model (b), the age factor is age at the current shunt insertion; it is significant for both second and third failures. We also observe that its effect is similar across first, second, and third failures. The previous failure time is significant for second but not for third failures.

The models (b) are more plausible and more easily interpreted. Once adjustment is made for age at the current shunt insertion, the association between first and second failure times is much reduced, and for second and third failures it disappears.

There is not much motivation for considering second or subsequent failure times conditional only on baseline covariates, because of the clear importance of covariates (age, concurrent surgery) that vary across failure times. However, there is clinical interest in whether time between successive failures tends to shorten, after accounting for fixed and time-varying risk factors; this would be expected if there are other unobserved risk factors that affect failure. A reasonable look at this can be based on the models (b) in Table 5.3; the effect of the previous gap time is only mildly significant for $j = 2$ and is not significant for $j = 3$ so let us consider models (b) in which the previous failure time is dropped. Figure 5.9 shows the generalized Nelson–Aalen estimates $\widehat{H}_{0j}(t)$ for these models, given by (4.13), for $j = 1, 2, 3$. Note that they are the estimated cumulative failure intensity functions for a patient with age $>$ 1, etiology = Congenital, shunt type = other, and concurrent surgery = no.

Table 5.3. Fitted Cox models for first, second, and third shunt failures.

	Age as Age at First Surgery					
	First Shunt		Second Shunt		Third Shunt	
Covariate	EST.	S.E.	EST.	S.E.	EST.	S.E.
Age at 1st surgery (yrs)						
< 0	0.97*	0.20	0.44	0.27	0.67	0.40
0–1	0.63*	0.16	0.20	0.23	0.07	0.36
> 1	–	–	–	–	–	–
Etiology						
Adsten	0.60*	0.25	-0.26	0.39	-0.01	0.50
IVH	0.61*	0.21	0.61*	0.30	0.05	0.40
Men	0.72*	0.26	0.75*	0.34	-0.06	0.46
MMC	0.53*	0.19	-0.03	0.29	-0.01	0.39
Trauma	1.05*	0.36	0.35	0.52	0.05	0.68
Tumor	0.76*	0.23	0.47	0.33	0.03	0.44
Other	0.41	0.21	0.21	0.31	-0.37	0.44
Con	–	–	–	–	–	–
VP shunt type	-0.36*	0.16	-0.19	0.23	-0.01	0.31
Concurrent surgery	0.44*	0.15	0.30	0.27	0.57	0.33
Previous failure time	–	–	-0.56*	0.17	-0.24	0.17
	Age as Age at Current Shunt Insertion					
	First Shunt		Second Shunt		Third Shunt	
Covariate	EST.	S.E.	EST.	S.E.	EST.	S.E.
Age at current surgery (yrs)						
< 0	0.97*	0.20	0.89*	0.34	1.21*	0.52
0–1	0.63*	0.16	0.24	0.20	0.43	0.23
> 1	–	–	–	–	–	–
Etiology						
Adsten	0.60*	0.25	-0.22	0.38	-0.29	0.51
IVH	0.61*	0.21	0.56	0.30	0.17	0.39
Men	0.72*	0.26	0.77*	0.34	-0.17	0.46
MMC	0.53*	0.19	0.00	0.28	-0.15	0.38
Trauma	1.05*	0.36	0.31	0.52	-0.09	0.68
Tumor	0.76*	0.23	0.46	0.32	0.02	0.40
Other	0.41*	0.21	0.22	0.31	-0.50	0.43
Con	–	–	–	–	–	–
VP shunt type	-0.36*	0.16	-0.19	0.23	-0.03	0.30
Concurrent surgery	0.44*	0.15	0.28	0.27	0.59	0.33
Previous failure time	–	–	-0.38	0.20	0.00	0.20

* $|\widehat{\beta}/\text{s.e.}(\widehat{\beta})| > 1.96$.

The plots indicate increasing risks of failure as we go from first to second to third shunts. A point of caution, however, is that for second shunts we have ignored the mild but significant positive association with the first shunt failure time. This makes the censoring times for w_2 slightly dependent on w_2, and similarly for w_3, especially for persons with shorter followup, and may result in a slight inflation of the estimates for $H_{02}(t)$ and $H_{03}(t)$. It should also be noted that only about 50% of individuals become at risk for even a second event, and additional selection effects may be present. This could be addressed with longer followup.

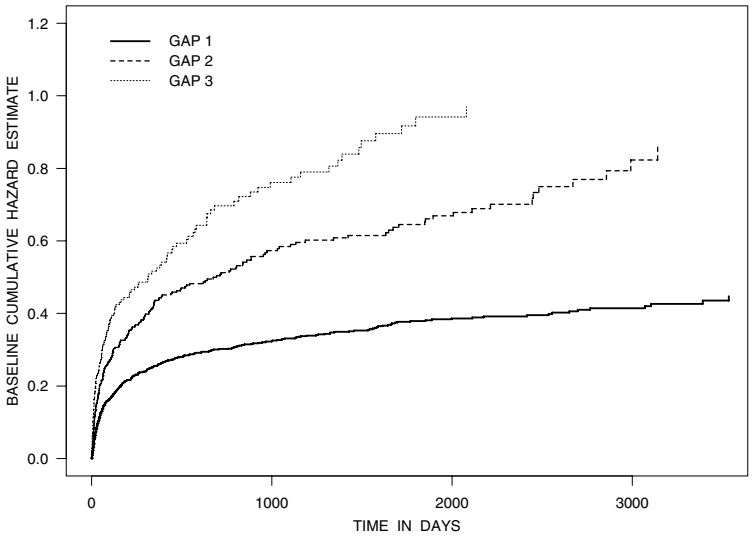


Fig. 5.9. Baseline cumulative hazard estimates for gaps 1, 2, and 3 between shunt failures.

5.5 Some Additional Illustrations

Many processes that produce recurrent events are complex, and of course the models that we employ to represent them are, at best, approximations to reality. What types of intensity specifications to consider, and whether to emphasize gap times or cumulative time on study (calendar time), are determined by the objectives of analysis and by evidence concerning the adequacy of the models in describing the event process. Different objectives may call for

quite different approaches. For example, rather detailed intensity-based models that incorporate previous event history or random effects may be useful in understanding the dynamics of individual event processes. However, such models usually do not provide easily interpreted comparisons of treatments, or assessment of fixed covariates. For that purpose simpler descriptive models, such as those for mean functions in Section 3.6, may be preferable.

When scientific background and objectives do not point clearly to one type of approach, it is worth examining the data through a variety of models. In presenting conclusions, parsimonious models are desirable, but fitting more complex models may provide additional insights concerning the processes under study. We consider here a pair of illustrations which involve different looks at the data and the processes which give them.

5.5.1 Pulmonary Exacerbations in the Study of rhDNase

In Sections 4.3.2 and 4.4.3 we considered data on the occurrence of pulmonary exacerbations in a randomized clinical trial on persons with cystic fibrosis. The distribution of number of exacerbations per subject, in each of two treatment arms, was shown in Table 1.2, and it is noted that subjects in the experimental treatment (rhDNase) group tend to experience fewer exacerbations than those in the control (placebo) group. However, only 139 of 324 placebo subjects, and 104 of 321 rhDNase subjects, experienced at least one event (exacerbation), and only 42 and 39 subjects, respectively, experienced two or more events, over the study period. Finally, when an individual experiences an exacerbation he is treated with antibiotics, and a new exacerbation cannot occur until treatment has ended and the person has recovered. Many of these time periods are in the range 10–15 days, but there is a high degree of variability.

One reason to consider a gap time analysis is that once an exacerbation has been cleared, the subject is in a sense renewed. Exacerbations promote scarring of the lungs, which tends to decrease lung function and increase susceptibility to infection, but we hope to deal with this by allowing the distributions of successive times $W_1, W_2, \ldots$ to be different. However, there is undoubtedly subject-to-subject heterogeneity in susceptibility, even after controlling for treatment and FEV, because the subjects vary in terms of age and the effects of disease. This manifests itself in the observed association between first and subsequent gap times.

In Section 4.3.2 we fitted conditional models to the gap times $W_1, W_2, \ldots$ between successive exacerbations. The results for W_1 (see Table 4.2) show that rhDNase use and higher baseline forced expiratory volume are both strongly associated with longer times W_1 to a first exacerbation. However, the first gap time W_1 is strongly positively associated with second gap time W_2, in models that include treatment and FEV as covariates, and the effects of these covariates are no longer significant. Because of the strong positive association between W_1 and the covariates, it is impossible to separate the effects of W_1, treatment, and FEV on W_2 and so this analysis does not provide insight into

whether treatment or FEV has persistent effects beyond a first exacerbation. As discussed in Sections 4.3.2 and 4.4.3, we can also obtain estimates of the marginal distribution for W_2. When this is done, and bearing in mind that under half of the subjects in the study experienced a first exacerbation and so were at risk for a second, no evidence of an effect of treatment or FEV on W_2 is found. Similar analysis of third gap times W_3 likewise shows no covariate effects.

Modulated Markov models, based on time on study t, provide another way to assess persistence of treatment effect. We show in Table 5.4 the results of fitting four models, which are as follows, with $\lambda_i(t)$ used to mean $\lambda(t|H_i(t))$.

Model 1: $\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2})$,
where $x_{i1} = I$ (subject i received rhDNase), x_{i2} = centered FEV, and $Y_i(t)$ indicates whether subject i is at risk of an exacerbation at time t. In particular $Y_i(t) = 0$ when a subject is being treated for an exacerbation.

Model 2: $\lambda_i(t|u_i) = u_i Y_i(t)\lambda_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2})$,
where the u_i are independent random effects, each with a gamma distribution with mean 1 and variance ϕ.

Model 3: $\lambda_i(t|u_i) = u_i Y_i(t)\lambda_0(t)\exp(\beta_1 x_{i1} I(t \le 80) + \beta_1' x_{i1} I(t > 80) + \beta_2 x_{i2})$.

Model 4: $\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(\beta_1 x_{i1} I(t \le 80) + \beta_1' x_{i1} I(t > 80) + \beta_2 x_{i2} + \beta_3 z_i(t))$, where $z_i(t) = I(N_i(t^-) > 0)$.

Models 1 and 2 are just semiparametric Markov (Andersen–Gill) models with and without random effects, as discussed in Sections 3.4 and 3.5. Table 5.4 indicates that ϕ in Model 2 is significantly different from zero, thus providing strong evidence against Model 1. This was to be expected, given the previous indication of dependence in the gap times. Although Model 1 is strongly contradicted by the data, it gives estimates of treatment effects that are close to those for Model 2, but the standard errors are smaller. This is in line with the discussion in Sections 3.5.3 and 3.6, where it is noted that when observation (at-risk) periods are independent of the event processes, both Models 1 and 2 produce robust estimates of the process mean function. At-risk periods are not quite independent of event history here because a person is not at risk while being treated for an exacerbation, but the effect of this is relatively small. Adding the duration of the immediately preceding antibiotic treatment for second and subsequent exacerbations takes away the marginal interpretation of treatment and FEV that we have in Models 1 and 2, but their estimates remain roughly the same.

Diagnostics for Model 2 discussed in Section 3.7 suggest that the treatment effect may vary over time. Model 3 incorporates time dependence; the discontinuous effect here is not highly realistic but allows a rough look. Table 5.4 suggests that the treatment is highly effective at reducing exacerbations over the first 80 days, but much less so thereafter. Finally, Model 4 presents

an alternative to the use of a random effect, by including a time-dependent covariate that indicates whether a subject has a previous exacerbation. An extension to Model 4 would be to use $z_i(t) = N_i(t^-)$, but in view of the rather small number of subjects with two or more events, we consider the simpler model. The intensity function in Model 3, with u_i integrated out, is of the form shown in (2.33), and is qualitatively similar to the intensity in Model 4 in the sense that the occurrence of an event increases the intensity of a new event.

The regression parameters β measure somewhat different things in terms of the intensities in Models 3 and 4 and hence the estimates are different. In particular, in Model 3 the parameters give relative risks for an individual, conditional on the random effect u_i that affects their baseline risk. In Model 4 the parameters are relative risks for individuals, adjusting for whether they have had a prior exacerbation. As discussed in Sections 3.5 and 3.6, the parameter estimates in Model 4 are expected to be smaller, but the Wald statistics and $p-$values for the regression parameters are similar.

Table 5.4. Modulated Markov models for pulmonary exacerbations.

	Model 1		Model 2		Model 3		Model 4	
	EST.	S.E.	EST.	S.E.	EST.	S.E.	EST.	S.E.
Treatment	-0.29	0.11	-0.31	0.13	–	–	–	–
Treatment ($t \leq 80$)	–	–	–	–	-0.51	0.18	-0.42	0.16
Treatment ($t > 80$)	–	–	–	–	-0.16	0.16	-0.14	0.14
FEV	-0.017	0.002	-0.019	0.003	-0.019	0.003	-0.017	0.002
$I(N(t^-) \geq 1)$	–	–	–	–	–	–	0.81	0.23
Variance (ϕ)	–	–	0.94	48.0†	0.94	48.0†	–	–

†No standard error available from software; value shown is likelihood ratio statistic for testing $\phi = 0$ (1 d.f).

The modulated Markov analyses give conclusions consistent with the gap time analysis. In each case there is an indication of variability across subjects that is not explained by treatment or FEV, and also a suggestion that the treatment effect may diminish somewhat with the time on study. In experimental studies it is often important to provide easily interpreted estimates of treatment effects which may be understood in population terms. Among the analyses conducted so far, Models 2 and 3 provide the best such summaries, because the regression coefficients can be interpreted in terms of the rate of events per subject; see Sections 2.2.3 and 3.5.1. If it were not for the antibiotic treatment periods, we would also have an interpretation through mean functions and the population rate at time t would be the conditional rate with $u_i = 1$. However, this does not hold exactly here, because $E(u_i|Y_i(t) = 1)$ is

not equal to $E(u_i)$, and so in models 2 and 3 the regression coefficients are not exactly population average effects.

A way to look at the population average effect of treatment is to consider a mean function for the number of events per subject, estimated as in Section 3.6, without model assumptions. One approach is to consider the model

$$E\{N_i(t)|x_i\} = \mu_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2})\,, \tag{5.27}$$

which represents the expected number of exacerbations up to t days on study, without any adjustment for days not at risk due to treatment for an infection. The followup time (as opposed to total at-risk time) of approximately 169 days for most subjects is independent of their event process, and so the methodology of Section 3.6.2 can be applied. Table 5.5 shows the resulting estimates of β_1 and β_2. Consistent with the discussion above concerning the fairly small effect of periods where subjects received antibiotics, this gives estimates that are similar to those for Models 1 and 2. Diagnostic checks of (5.27) show that the effect of treatment may not be constant over time, so we fit a second model that stratifies on treatment:

$$E\{N_i(t)|x_i\} = \mu_{0,x_{i1}}(t)\exp(\beta x_{i2})\,. \tag{5.28}$$

The estimated effects of FEV, shown in Table 5.5 are comparable but the more interesting point is to compare $\widehat{\mu}_0(t)$ and $\widehat{\mu}_1(t)$, shown in Figure 5.10. This indicates a roughly linear mean function, and roughly constant rate of exacerbations, for the placebo group ($x_{i1} = 0$), but suggests a very slightly increasing rate of exacerbations for the rhDNase group. This is consistent with a treatment effect that is diminishing slightly with time on study.

Table 5.5. Estimates of mean function for rhDNase study based on models (5.27) and (5.28).

	Model (5.27)		Model (5.28)	
	EST.	S.E.	EST.	S.E.
Treatment	-0.27	0.12	–	–
FEV	-0.017	0.003	–	–
Stratified model				
FEV for control patients	–	–	-0.019	0.004
FEV for rhDNase patients	–	–	-0.015	0.004

5.5.2 Analysis of Asthma Exacerbations

We consider data from a multicenter randomized trial designed to assess the effect of 400 versus 200 μg/day of fluticasone propionate for the prevention of

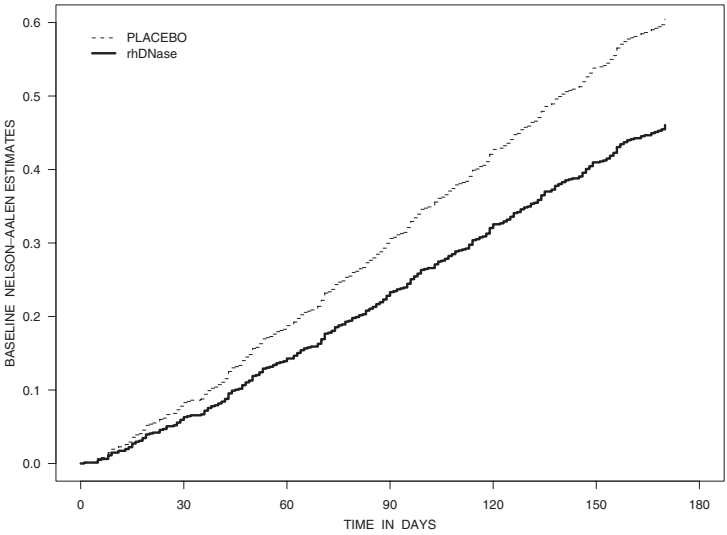


Fig. 5.10. Expected cumulative number of pulmonary exacerbations versus time on study for rhDNase study.

exacerbations among children with asthma between four and eleven years of age (Verona et al., 2003). In the study, 261 and 267 children were randomized to received 400 μg and 200 μg daily doses, respectively. The original protocol specified 3 months followup but after discussion, a protocol amendment was established to extend followup to 12 months. During followup, participants attended scheduled assessments and clinical examinations to record use of additional medications, assess quality of life, and obtain lung function and other laboratory measurements. The primary outcome in the following analyses is the occurrence of exacerbations. Table 5.6 displays the distribution of the number of exacerbations experienced across all subjects, but it should be noted that a substantial number of subjects dropped out of the study after 3 months, as described below.

Figure 5.11 gives the Nelson–Aalen estimates of the cumulative mean functions $\widehat{\mu}(t)$ for the two groups. Comparison of the estimates reveals a separation early on which increases slightly over the followup period. A robust test of a treatment effect based on (3.52) with $a(u) = 1$ gives $p = 0.0732$. The relative rate of exacerbations for individuals receiving 400 μg/day versus 200 μg/day of fluticasone propionate based on a proportional rate function model as in Section 3.6.3 is 0.793 and a 95% confidence interval based on a robust variance estimate is (0.616, 1.022). This is a robust estimate of the effect of treatment dose, relevant for primary analysis of this clinical trial. As discussed in Section

Table 5.6. Distribution of frequency of exacerbations in asthma study with daily doses of fluticasone propionate.

	Dose of Treatment				
Frequency of	200 μg/day		400 μg/day		
Exacerbations	Number	Percent	Number	Percent	Total
0	126	47.2	132	50.8	258
1	59	22.1	55	21.2	114
2	38	14.2	39	15.0	77
3	14	5.2	15	5.8	29
4	7	2.6	10	3.8	17
5	9	3.4	0	0.0	9
6	4	1.5	2	0.8	6
7	5	1.9	2	0.8	7
8	2	0.7	3	1.2	5
9	0	0.0	2	0.8	2
10	1	0.4	0	0.0	1
11	0	0.0	0	0.0	0
12	2	0.7	0	0.0	2
Total	267	100.0	261	100.0	528

3.6, robust marginal comparisons of this sort are valid provided the censoring process is completely independent of the event process.

The regression analyses that follow are more directed at understanding prognostic variables for exacerbations and process dynamics. These types of analyses are driven by questions regarding risk factors associated with event occurrence, which may help identify high-risk patients or time periods of higher risk. We include the treatment covariate because there is some evidence that it has an effect, based on direct comparisons of the two treatment arms. As discussed earlier, we caution readers regarding the different interpretations of the treatment effect in the models that follow, however. Models featuring conditioning on event occurrence after randomization (i.e. events responsive to treatment) can lead to quite different estimates than marginal comparisons; often the estimate of treatment effect is attenuated. In intensity-based models the treatment effect is interpreted as the relative intensity of events among subjects with the same covariate values and relevant event histories.

Regression Analysis

We consider here regression analyses involving the following covariates: x_{i1}, a treatment indicator such that $x_{i1} = 1$ if subject i was randomized to receive 400 μg/day and $x_{i1} = 0$ otherwise, x_{i2} such that $x_{i2} = 1$ if subject i is male and 0 otherwise, x_{i3} which is the subject's age in years (centered on the sample

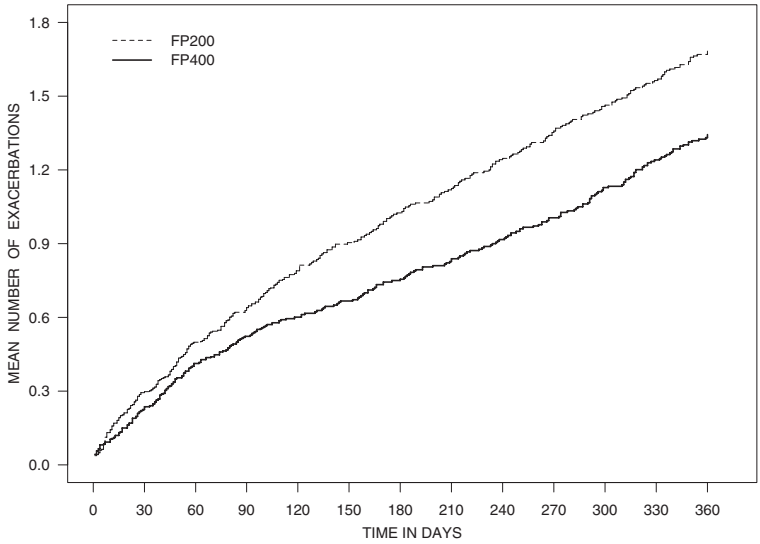


Fig. 5.11. Nelson–Aalen estimates of the cumulative mean function for the groups receiving 400 and 200 μg/day of fluticasone propionate.

mean), x_{i4} which is the subject's weight in kilograms (centered on the sample mean), x_{i5} which is the subject's centered percentage predicted expiratory flow (PEF) rate, a measure of lung function, and a categorical time-dependent covariate $x_{i6}(t)$ denoting the season, with $x_{i6}(t) = 1$ during January–March, $x_{i6}(t) = 2$ during April–June, $x_{i6}(t) = 3$ during July–September, and $x_{i6}(t) = 4$ during October–December.

Table 5.7 displays the results of fitting several multiplicative models to the data. The first model (Model 1A) is Poisson with intensity

$$\lambda_i(t|H_i(t)) = \rho_0(t)\exp(x_i'(t)\beta)\,,$$

where $x_i(t) = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}^*(t))'$, with $x_{i6}^*(t) = (I(x_{i6}(t) = 2), I(x_{i6}(t) = 3), I(x_{i6}(t) = 4))$. Model 2A is a more general model controlling for the cumulative number of exacerbations up to the fifth. The intensity is of the form

$$\lambda_i(t|H_i(t)) = \lambda_0(t)\exp(x_i'(t)\beta + \gamma N_i^*(t^-))\,,$$

where $N_i^*(t) = N_i(t)$ for $N_i(t) \leq 5$ and $N_i^*(t) = 5$ if $N_i(t) > 5$. A Wald or likelihood ratio test of Model 2A versus 1A (i.e. of $\gamma = 0$) suggests that the introduction of $\gamma N_i^*(t^-)$ significantly improves the fit and so Model 2A is preferred over 1A. The estimate of treatment effect in Model 1A represents

the log-relative rate of exacerbations among subjects with the same sex, age, weight, %PEF, and during the same season. In Model 2A, the treatment comparison is further restricted to subjects with the same cumulative number of events and it is not surprising that the estimates of effect are quite different in Models 1A and 2A. For each additional previous exacerbation, the risk of subsequent exacerbation increases by an estimated 60% ($RR = 1.60$, $p < 0.0001$).

Model 3A represents a further generalization through stratification. Based on the frequency distribution in Table 5.6 we consider six strata and assume that the baseline rate of events is the same for the fifth and subsequent events; this is equivalent to stratifying on $N_i^*(t^-)$ to give

$$\lambda_i(t|H_i(t)) = \sum_{k=0}^{5} Y_{ik}(t)\alpha_{k0}(t)\exp(x_i'(t)\beta)\,.$$

This model relaxes the assumption in Model 2A of the proportional effect of each exacerbation on the intensity. The estimates and conclusions regarding covariate effects are very similar to those of Model 2A; their interpretations are quite similar too. Event rates $\alpha_{k0}(t)$ can be compared through plots of the estimated cumulative baseline rate functions (5.24).

It should be noted that the log-likelihood for model 3A is not comparable to those for Models 1A and 2A because of the stratification. We discuss later in the example how to assess the need for stratification.

Table 5.7. Results from fitting several models to data on recurrent asthma exacerbations.

		Unstratified				Stratified	
		Model 1A		Model 2A		Model 3A	
Covariate		EST.	S.E.	EST.	S.E.	EST.	S.E.
Treatment		-0.212	0.080	-0.106	0.080	-0.101	0.081
Sex		-0.066	0.088	-0.052	0.088	-0.072	0.089
Age		-0.135	0.024	-0.069	0.024	-0.073	0.024
Weight		0.015	0.005	0.007	0.005	0.007	0.005
% PEF		0.003	0.002	0.003	0.002	0.001	0.002
Season	Jan–March	–	–	–	–	–	–
	April–June	0.001	0.347	-0.071	0.350	-0.138	0.350
	July–Sept	0.156	0.468	0.129	0.466	-0.043	0.469
	Oct–Dec	0.369	0.424	0.381	0.425	0.291	0.426
Event count		–	–	0.468	0.029	–	–
Log-likelihood		-3894.650		-3778.942		-2992.146	

Table 5.8. Results from fitting several random effect models to data on recurrent asthma exacerbations.

		Unstratified				Stratified	
		Model 1B		Model 2B		Model 3B	
Covariate		EST.	S.E.	EST.	S.E.	EST.	S.E.
Treatment		-0.211	0.127	-0.158	0.107	-0.110	0.085
Sex		-0.102	0.142	-0.083	0.118	-0.080	0.094
Age		-0.128	0.037	-0.102	0.032	-0.079	0.025
Weight		0.015	0.009	0.011	0.007	0.008	0.006
% PEF		0.002	0.003	0.003	0.002	0.001	0.002
Season	Jan–March	–	–	–	–	–	–
	April–June	-0.024	0.345	-0.047	0.348	-0.132	0.349
	July–Sept	0.141	0.465	0.140	0.464	-0.033	0.469
	Oct–Dec	0.355	0.424	0.370	0.423	0.297	0.426
Event count		–	–	0.209	0.037	–	–
Variance (ϕ)		1.187		0.569		0.066	
Log-likelihood		-3778.729		-3775.373		-2991.887	

Models 2A and 3A incorporate dependence on prior events in the intensity function. Some or all of the apparent dependence could be due to unobserved individual effects, so we also consider analogous random effects models of the form (3.28). Table 5.8 gives the estimates and standard errors obtained from models with gamma random effects using `coxph` with the `frailty(id)` option. We refer to the mixed Poisson model analogous to Model 1A as Model 1B. The maximum likelihood estimate of the random effect variance is $\widehat{\phi} = 1.19$, suggesting the presence of extra-Poisson variation. This is also indicated by a comparison of the maximum log-likelihoods under Models 1A and 1B, and reflected in the substantially larger standard errors for regression coefficient estimates in Model 1B than in 1A. Although the point estimates are similar in the two models, treatment and weight are not strongly significant in Model 1B. The gamma variance parameter estimate is quite a bit smaller in Model 2B where $N_i^*(t^-)$ is controlled for; here one obtains $\widehat{\phi} = 0.57$. The difference in maximum log-likelihoods for Models 2A and 2B is much smaller than for 1A and 1B, but still significant. The maximum likelihood estimate of ϕ in Model 3B is only 0.07 and the standard errors are much closer to those of Model 3A, indicating little need to model any subject to subject variability beyond that explained by the fixed effects in Model 3A.

Note that the models in Table 5.7 are nested within the corresponding models in Table 5.8 (e.g. 3A is nested within 3B). Thus likelihood ratio tests performed for the need for random effects are based on a 50:50 mixture of a point mass at zero and a χ_1^2 distribution under the null hypothesis that $\phi = 0$ (e.g. Moran, 1971). The likelihood ratio test of 3A versus 3B gives $p =$

0.24. Figure 5.12 gives the profile relative likelihood functions for ϕ based on Models 1B–3B, respectively. These are obtained using the `coxph` function with a specified value for the variance parameter; the command is `+frailty(id, theta=theta.fixed)`, where `theta.fixed` is the specified value of ϕ.

We consider Model 3A as a possible final model for the exacerbation process. Additional analyses did not yield any evidence of a dependence on previous gap times. For a more formal assessment of the need for stratification we also consider the model

$$\lambda_i(t|H_i(t)) = \lambda_0(t)\exp(x_i'(t)\beta + \sum_{k=1}^{4}\gamma_k I(N_i(t^-) = k) + \gamma_5 I(N_i(t^-) \geq 5))$$

which over the first few exacerbations may be viewed as a model intermediate between 2A and 3A. It is more general than 2A because it does not assume the same multiplicative increase in the rate for the second event as for the first event. If $\gamma_k = k\gamma_1$ for $k = 2, 3, 4, 5$, then these models would be comparable for up to the fifth event. This model is less general than 3A however, because it assumes that the baseline functions for the four strata are proportional, but it facilitates testing the need for stratification. The estimated coefficients (s.e.) are $\widehat{\gamma}_1 = 0.729$ (0.105), $\widehat{\gamma}_2 = 1.000$ (0.138), $\widehat{\gamma}_3 = 1.797$ (0.167), $\widehat{\gamma}_4 = 1.763$ (0.201), $\widehat{\gamma}_5 = 2.314$ (0.169), and the corresponding log-likelihood is -3772.816. The likelihood ratio test of this model versus Model 2A gives a likelihood ratio statistic 23.362 with $p = \Pr(\chi_4^2 > 12.252) = 0.016$. Because the intermediate model appears superior to Model 2A, we take this as evidence that some form of stratification is warranted. The tests for the proportionality of the covariates effects for $I(N_i(t^-) = k)$, $k = 1, 2, 3, 4$, and $I(N_i(t^-) \geq 5)$ in this intermediate model give fairly strong evidence against the assumption of multiplicative effects for $I(N_i(t^-) = 1)$ and $I(N_i(t^-) \geq 5)$ with $p-$values $p = 0.012$ and $p = 0.023$, respectively, although the global test of fit based on `cox.zph` gives $p = 0.299$. Because of the evidence against multiplicative effects for some of these variables we conclude that full stratification as given in Model 3A is a more appropriate way of addressing the dependence on the event history for these data and therefore Model 3A preferred.

The estimates in Model 3A suggest that, when controlling for all other covariates, there may be a mild reduction in the risk of exacerbations with 400 μg/day versus 200 μg/day of fluticasone ($RR = 0.90$, 95% CI (0.77, 1.06), $p = 0.212$). For each additional year of age the risk of exacerbations decreases by an estimated 7% ($RR = 0.93$, 95% CI (0.89, 0.97), $p = 0.002$). There is no significant assocation between weight or percentage predicted peak expiratory flow on the risk of exacerbations ($p = 0.193$ and $p = 0.403$, respectively). A three degree of freedom Wald test reveals that there is no significant seasonal variation in the rate of exacerbations ($p = 0.807$).

There is always the question as to whether dependence on prior event history is partly or fully due to heterogeneity across subjects. In this case we prefer the fixed effect Model 3A to the random effects Model 2B, because in

Model 2B there remains a strong effect due to prior event count, whereas the addition of a random effect to Model 3A (see Model 3B) is insignificant. Note, however, that although Models 3A and 2B have different covariate effects, the Wald statistics and $p-$values for tests of no effect are similar.

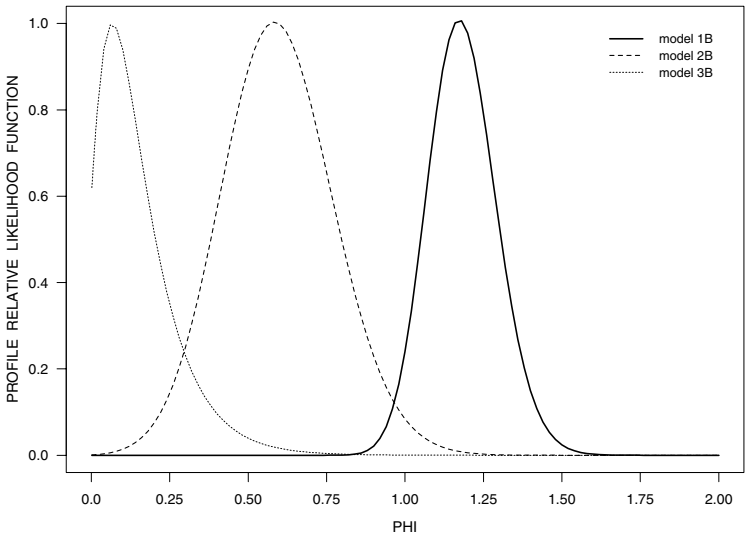


Fig. 5.12. Profile relative likelihood plots for ϕ in Models 1B–3B.

5.6 Bibliographic Notes

Intensity-based modeling was discussed in Chapters 1 and 2, and the Bibliographic Notes in Sections 1.7 and 2.7 noted that the literature on point processes (e.g. Cox and Isham, 1980; Daley and Vere-Jones, 1988) contains much material. In the counting process literature on event history analysis, Andersen et al. (1993) provide a very thorough account of intensity-based methods, emphasizing modulated Markov models. Aalen et al. (2004) provide interesting discussion on intensity-based versus random effects modeling. Fosen et al. (2006a,b) give an interesting perspective on internal covariates and the use of path analysis methods to assess internal and external covariate effects.

Cox (1972b) introduced semiparametric modulated renewal models and Andersen and Gill (1982) considered Markov models, stimulated by the Cox

model in survival analysis (Cox, 1972a). Lawless and Thiagarajah (1996) considered the hybrid parametric models of Section 5.2.1 and Lindqvist et al. (2003) introduced the trend renewal process of Section 5.2.2, extending an idea of Berman (1981). Examples of other intensity-based models abound in the literature associated with long series of events, for example, in the analysis of earthquake occurrences (Ogata, 1988) and signal processing (Snyder and Miller, 1991).

Modulated Markov processes and Cox-type statistical analysis were emphasized by Prentice et al. (1981), who also introduced stratification. Prentice et al. (1981) and Gail et al. (1980) considered modulated and stratified renewal models of the semiparametric Cox type. Different methods of estimating transition probabilities in multistate models, as used in Section 5.3, were discussed by Couper and Pepe (1997), Aalen et al. (2001), Datta and Satten (2001), Glidden (2002), Cook et al. (2003), and others. Maller et al. (2002) consider the Kaplan–Meier method (Pepe, 1991) for estimation of $\mu(t)$ via (5.18). Andersen et al. (1993) is an authoritative reference on modulated Markov models as well as on estimation for Markov multistate models. Formal justification of the methodology in Section 5.4 for modulated revewal or semi-Markov models is considered by Oakes and Cui (1994) in the case of a single process and, for multiple processes, by Dabrowska et al. (1994). Lawless et al. (2001) provide informal justification by noting the connection between maximum likelihood methods for semi-Markov models with piecewise-constant baseline intensity functions and the Cox-type semiparametric analysis. Chang and Wang (1999) consider further examples of modulated renewal or semi-Markov models with stratification. Many applications of modulated semi-Markov models are found in the economics and social sciences literature; see, for example, Heckman and Singer (1985, 1986) and Heckman and Walker (1992) for insightful discussion and examples.

A number of authors provide comparisons of calendar time and gap time analyses of specific datasets, and of different strategies for assessing the effects of event history on the intensity. For examples, see Therneau and Hamilton (1997), who consider the rhDNase dataset of Section 5.5.1, Keiding et al. (1998), Therneau and Grambsch (2000, Ch. 8), Kalbfleisch and Prentice (2002, Ch. 9), and Cai and Schaubel (2004b). Lin (1994), Clayton (1994), Gao and Zhou (1997), and Wei and Glidden (1997) consider alternative methods for settings that include multivariate failure times as well as recurrent events.

5.7 Problems and Supplements

5.1. Obtain the $p \times p$ Hessian matrix $H = \partial^2 \ell(\theta)/\partial\theta\partial\theta'$ for the log-likelihood function (5.6) for the log-linear model (5.5). Prove that $a'Ha \leq 0$ for any real vector $a = (a_1, \ldots, a_p)'$, and thus deduce that $\ell(\theta)$ is convex and has a unique maximizer $\widehat{\theta}$, provided $\sum_{i=1}^m Y_i(t) z_i(t) z_i'(t)$ is positive definite for at least one value of $t > 0$.

Prove a similar result for an additive model where $\lambda(t|H(t)) = z'(t)\theta$. What complication does the need to restrict θ so that $z'(t)\theta \geq 0$ introduce?

[Section 5.1]

5.2. Additive models as in (2.22) sometimes describe recurrent event data well. For the system failure data of Section 5.2.4, consider parametric models with intensities of the form

$$\lambda(t|H(t)) = g_1(t;\beta) + I(N(t^-) > 0)g_2(B(t);\alpha)\,,$$

where $B(t) = t - N(t^-)$. Fit such models and assess their adequacy.

[Section 5.2]

5.3. Consider the data in Problem 3.18 on unscheduled maintenance events for engine number 4 on the submarine U.S.S. Grampus.

a. Use parametric models like those in Section 5.2.4 to investigate the nature of the event intensity. Do either Poisson or renewal processes appear satisfactory?
b. There were also scheduled engine overhauls at various times (Ascher and Feingold, 1984, p. 76). Overhauls were performed at times (in thousands of hours of operation) 1.203, 3.197, 5.414, 7.723, 10.594, 14.357, and 15.574. Assess whether overhauls have any effect on the unscheduled event intensity. Does there appear to be any renewal effect following an overhaul?

[Section 5.2]

5.4. Tests for trend were discussed in Section 3.7.1. Generalize the score tests there by considering the intensity model

$$\lambda(t|H(t)) = h_0(B(t);\gamma)\exp(\beta g(t))\,, \tag{5.29}$$

where $h_0(w;\gamma)$ is a specified parametric hazard function and $g(t)$ is a specified function. That is, given data over $[0,\tau_i]$ for independent individuals $i = 1,\ldots,m$, consider partial score test statistics (A.10) in Appendix A for $H_0 : \beta = 0$, where $U(0,\tilde{\gamma}(0))$ is of the form

$$U = \sum_{i=1}^{m}\left(\frac{\partial \ell_i(\beta,\gamma)}{\partial\beta}\right)\Bigg|_{\beta=0,\gamma=\tilde{\gamma}} \tag{5.30}$$

where $\tilde{\gamma}$ is the maximum likelihood estimate of γ when $\beta = 0$ in (5.29). Note that $\tilde{\gamma}$ is obtained by fitting the renewal process where the gap times W_{ij} have hazard function $h_0(w)$, as discussed in Section 4.1. Consider (5.30) in the special case where $g(t) = t$; give a variance estimate for U.

[Section 5.2]

5.5. The model (5.29) is a modulated renewal process (4.24) with $z(t) = g(t)$, and we can test $H_0 : \beta = 0$ in the semiparametric model where $h_0(B(t))$ is not specified parametrically, as described in Section 4.2.4. Use both this approach and the approach of Problem 5.4, with $h_0(w) = \gamma_1\gamma_2 t^{\gamma_2 - 1}$ of Weibull form, to test for trend in the air-conditioning failure data of Section 5.2.4. What complication does the fact that $m = 1$ introduce? See also Problem 5.10.

[Sections 5.2 and 5.4]

5.6. Carefully describe why the multistate analysis in Section 5.3.1 provides estimates of mean functions which are robust to state-dependent censoring, when the Markov model holds.

[Section 5.3]

5.7. Show that $\widehat{p}_{0k}(t)$ from (5.20) equals $Y_{\cdot k}(t)/m$ and then prove that the Nelson–Aalen estimate of the mean function $\mu(t)$ is numerically identical to the estimate based on (5.18) and(5.20), in the absence of any censoring before time t.

[Section 5.3; Andersen et al., 1993, Section 4.4.1.4]

5.8. Consider the method of estimating the mean function $\mu(t) = E\{N(t)\}$ described in Section 5.3.2. That is,

$$\widehat{\mu}(t) = \sum_{k=0}^{\infty} k\widehat{p}_{0k}(t) = \sum_{k=1}^{\infty} \widehat{F}_k(t)\,, \tag{5.31}$$

where $F_k(t) = \Pr(T_k \le t)$ and $\widehat{F}_k(t)$ is the Kaplan–Meier estimate based on data (t_{ik}, δ_{ik}), where $\delta_{ik} = I(T_{ik} \le \tau_i)$.

a. Show that if there is no censoring of an individual before time t, then (5.31) equals the Nelson–Aalen estimate (3.17).
b. Discuss why (5.31) will be biased if censoring times τ_i depend on the number of prior events.

[Section 5.3; Pepe, 1991; Cook et al., 2003]

5.9. Consider a clinical trial designed to assess the effect of an experimental treatment versus a placebo control on the prevention of recurrent complications such as asthma exacerbations.

a. Discuss the utility of intensity-based models as a basis for the primary analysis of a treatment comparison on the occurrence of events.
b. Discuss the utility of intensity-based methods for secondary analyses exploring the natural history of the disease process in the control arm.

[Section 5.3]

5.10. Consider a modulated renewal process with intensity function of the form (5.29), and for simplicity consider $h_0(w;\gamma) = \gamma$ and $g(t) = t$.

a. Consider maximum likelihood estimation of (γ, β) based on observation of a single process ($m = 1$) over the period $[0, \tau]$. What conditions seem needed for $(\widehat{\gamma}, \widehat{\beta})'$ to be asymptotically normal as $\tau \to \infty$?
b. Consider the partial score statistic for testing that $\beta = 0$. Show it has a limiting normal distribution as $\tau \to \infty$.
c. Consider the semiparametric model where $h_0(w)$ is arbitrary, with β estimated using (4.25). What seems needed for $\widehat{\beta}$ to have a normal limiting distribution as $\tau \to \infty$?
d. Investigate the various distributions above by simulation.

[Sections 3.7.1, 5.2, 5.4; Feigin, 1976; Cox, 1972b; Oakes and Cui, 1994]

5.11. For the pulmonary exacerbation analysis in Section 5.5.1, explain why it makes sense to assume persons are at risk of an exacerbation event at all followup times for the estimates shown in Table 5.5, but to exclude periods where a person is being treated for an exacerbation with antibiotics, for the models represented in Table 5.4.

[Section 5.5]

6
Multitype Recurrent Events

6.1 Multivariate Event Data

In many studies of chronic disease subjects are at risk of different types of recurrent events. For example, transient ischemic attacks may be classified according to location in cardiovascular trials, migraines may be differentiated by severity in neurological studies, and in respiratory studies asthma exacerbations may be subtyped according to cellular analyses of sputum samples. In other settings, one may record the causes of production stoppages in manufacturing, different types of financial transactions in commerce, and the types of claims filed by insurance policy holders.

It may be sufficient to perform separate analyses for the different event types, especially if they occur more or less independently of each other. However, models for multivariate counting processes are most often needed if events are related or if the occurrence of one event affects the risk of another type of event. The methods discussed in Chapters 2 to 5 can readily be adapted to facilitate joint analyses. For example, intensity-based models can provide a full specification of a multivariate point process. Internal time-dependent covariates may be used to express the effect of one type of event on the occurrence of other types. It is also sometimes helpful to formulate multivariate models through assumptions of conditional independence between events given univariate or multivariate random effects, a natural adaptation of the approaches of Sections 3.5 and 4.2. Finally, when interest lies in rate or mean functions, or other marginal features, the robust methods discussed in Section 3.6 may be adapted.

We consider the approaches just mentioned in Sections 6.2 to 6.4. We also look at two special settings in Sections 6.5 and 6.6, the first involving repeated transitions between two states and the second involving recurrent events in the presence of a terminating event. We present the methodology first, deferring illustrations to Section 6.7.

6.2 Intensity-Based Methods

6.2.1 Notation and Intensity Functions

We start by extending the notation of previous chapters to deal with multiple types of events. Consider a sample of m subjects in which each subject is at risk of J different types of recurrent events. Let i index subjects and j index the event types so $i = 1, \ldots, m$, and $j = 1, \ldots, J$. Let $N_{ij}(t)$ be the number of type j events occurring over the interval $[0, t]$ for subject i, $\Delta N_{ij}(t) = N_{ij}(t + \Delta t^-) - N_{ij}(t^-)$, and let $dN_{ij}(t) = N_{ij}(t) - N_{ij}(t^-)$ indicate whether a type j event occurred for subject i at time $t \geq 0$. The full vector of counting processes is denoted $N_i(t) = (N_{i1}(t), \ldots, N_{iJ}(t))'$, and $dN_i(t) = (dN_{i1}(t), \ldots, dN_{iJ}(t))'$, $i = 1, \ldots, m$. We assume that each subject is under observation for all types of events over the same period of time, although it is possible to relax this assumption. Let $[0, \tau_i]$ denote the period of observation for subject i and as before, let $Y_i(t) = I(t \leq \tau_i)$. The event history for subject i is $H_i(t) = \{N_i(s) : 0 \leq s < t\}$ and the intensity function for type j events is defined as

$$\lambda_{ij}(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta N_{ij}(t) = 1|H_i(t))}{\Delta t}.$$

Let t_{ijk}, $k = 1, \ldots, N_{ij}(t)$, denote the times of type j events over $[0, t]$, $j = 1, \ldots, J$ and $t_{i1}, \ldots, t_{iN_{i\cdot}(t^-)}$ denote the times of all types of events for subject i over $[0, t]$, with $N_{i\cdot}(t) = \sum_{j=1}^{J} N_{ij}(t)$, $\Delta N_{i\cdot}(t) = N_{i\cdot}(t + \Delta t^-) - N_{i\cdot}(t^-)$ and $dN_{i\cdot}(t) = \sum_{j=1}^{J} dN_{ij}(t)$. We assume that at most one event can occur at any given time, with

$$\begin{aligned} \Pr\{\Delta N_{ij}(t) = 1|H_i(t)\} &= \lambda_{ij}(t|H_i(t))\Delta t + o(\Delta t) \\ \Pr\{\Delta N_{i\cdot}(t) = 0|H_i(t)\} &= 1 - \sum_{j=1}^{J} \lambda_{ij}(t|H_i(t))\Delta t + o(\Delta t) \\ \Pr\{\Delta N_{i\cdot}(t) \geq 2|H_i(t)\} &= o(\Delta t). \end{aligned} \tag{6.1}$$

Therefore, if we consider events over a specified time period $[0, \tau]$ for subject i, and the partition $0 = u_0 < u_1 < \cdots < u_R = \tau$ with $\Delta u_r = u_{r+1} - u_r$, then the probability distribution of $N_i(u_0), \ldots, N_i(u_R)$ is

$$\prod_{r=0}^{R} \Pr(N_i(u_r)|H_i(u_r)) = \prod_{r=0}^{R} \Pr(\Delta N_i(u_r)|H_i(u_r))$$

which is given by

$$\prod_{r=0}^{R} \left\{ \prod_{j=1}^{J} \left[\lambda_{ij}\left(u_r|H_i(u_r)\right)\Delta u_r\right]^{\Delta N_{ij}(u_r)} \left[1 - \sum_{j=1}^{J} \lambda_{ij}\left(u_r|H_i(u_r)\right)\Delta u_r\right]^{1-\Delta N_{i\cdot}(u_r)} \right\},$$

plus terms of higher order in the Δu_r. The likelihood is obtained by dividing by $\prod_j \prod_k (\Delta t_{ijk})$ and taking the limit as $R \to \infty$ as in Section 2.1, to give

$$L_i = \left\{ \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \lambda_{ij}(t_{ijk}|H_i(t_{ijk})) \right\} \exp(-\sum_{j=1}^{J} \int_0^{\tau_i} \lambda_{ij}(u|H_i(u))du)$$

$$= \prod_{j=1}^{J} \left\{ \prod_{k=1}^{n_{ij}} \lambda_{ij}(t_{ijk}|H_i(t_{ijk})) \exp(-\int_0^{\tau_i} \lambda_{ij}(u|H_i(u))du) \right\}. \qquad (6.2)$$

The factorization in (6.2) reveals that with intensity-based analyses where the type-specific intensity functions are functionally independent, estimates of the different intensities can be obtained separately by maximum likelihood. As discussed in Section 2.6, the expression (6.2) can also be used as a likelihood function when τ_i is random, provided that it is a stopping time. One may then adopt a particular intensity-based model, as in Chapter 5, for each of the specific event types.

6.2.2 Remarks on Intensity-Based Models

Intensity functions of the Poisson form $\lambda_{ij}(t|H_i(t)) = \rho_{ij}(t)$ imply that the event processes are mutually independent. This is a very special situation, although when covariates are present, the assumption is only one of conditional independence. However, it is often implausible that the available covariates will explain the full extent of association between event types and so we should consider the possibility that the processes are related. Intensity-based models have great flexibility, but we usually look for fairly parsimonious models that capture the dynamics of the individual processes and relationships between them. We mention here a few types of model.

Markov models include ones analogous to (5.14) for each type of event. Specifically, the intensity function for type j events may adopt a new functional form upon the occurrence of each type j event, $j = 1, 2, \ldots, J$. If $Y_{ijk}(t) = I(T_{ijk} \le t < T_{ij,k+1})$, then we may consider

$$\lambda_{ij}(t|H_i(t)) = \sum_{k=0}^{\infty} Y_{ijk}(t)\alpha_{jk}(t)\,, \qquad (6.3)$$

where $\alpha_{jk}(t)$ is the rate of type j events among subjects with k such events over $[0, t]$, $k = 0, 1, \ldots$. This may be modulated by covariate effects. For example, let $z_{ijh}(t)$, $h = 1, 2, \ldots, q_j$ denote internal covariates capturing different aspects of the history for the jth process over $[0, t)$, $x_{ij}(t)$ denote external covariates, $z_{ij}(t) = (z_{ij1}(t), \ldots, z_{ijq_j}(t), x'_{ij}(t))'$, and amend $H_{ij}(t)$ to

$$H_{ij}(t) = \{N_{ij}(s) : 0 \le s < t;\ z_{ij}(s) : 0 \le s \le t\}\,.$$

Then we can consider

$$\lambda_{ij}(t|H_i(t)) = \sum_{k=0}^{\infty} Y_{ijk}(t)\alpha_{jk0}(t)\exp\left(z'_{ij}(t)\beta_j\right) \tag{6.4}$$

where β_j is a vector of coefficients. We may often wish to assume that the $\alpha_{jk0}(t)$ are the same and denoted $\alpha_{j0}(t)$ in (6.4), and use $z_{ij}(t)$ to model the dependence on event history.

This model does not address possible associations between the different types of event occurrences, but generalizations which do are easily obtained. For example, one may set

$$\lambda_{ij}\left(t|H_i(t)\right) = \sum_{k=0}^{\infty} Y_{ijk}(t)\alpha_{jk0}(t)\exp\left(z'_i(t)\beta_j\right)\,, \tag{6.5}$$

where $z_i(t) = (z'_{i1}(t), \ldots, z'_{iJ}(t))'$ contains relevant information on the history of all processes over $[0,t)$, β_j is a vector of coefficients for type j events, and

$$H_i(t) = \left\{N_i(s) : 0 \le s < t;\ z_i(s) : 0 \le s \le t\right\}.$$

Examples of useful internal covariates include the cumulative number of type j events ($z_{ij1}(t) = N_{ij}(t^-)$), the most recent gap time for type j events, ($z_{ij2}(t) = I(N_{ij}(t^-) > 0)(t_{ijk} - t_{ij,k-1})$, where $N_{ij}(t^-) = k$), or the number of type j events in the past s time units ($z_{ij3}(t) = I(t > s)(N_{ij}(t^-) - N_{ij}((t - s)^-))$). Another example is

$$\lambda_{ij}(t|H_i(t)) = \sum_{k=0}^{\infty} I(N_{i\cdot}(t^-) = k)\alpha_{jk}(t)\,, \tag{6.6}$$

which has the intensity for type j events depending on the cumulative number of events of any type.

Semi-Markov models may also be specified. Let $B_i(t)$ denote the backwards recurrence time for the previous event of any type, and $B_{ij}(t)$ denote the backwards recurrence time for the previous type j event. Independent semi-Markov models for each event type have intensities of the form

$$\lambda_{ij}(t|H_i(t)) = h_j(B_{ij}(t))\,, \tag{6.7}$$

where $h_j(\cdot)$ is the hazard function for the gap time between type j events. If

$$\lambda_{ij}(t|H_i(t)) = h_j(B_i(t))\,, \tag{6.8}$$

then renewals occur at the time of each event of any type. Modulated renewal models are obtained by the introduction of covariates dependent on the process history as discussed in Chapter 4. Introducing covariates into (6.7) gives, for example,

$$\lambda_{ij}(t|H_{ij}(t)) = h_j(B_{ij}(t))\exp\left(z'_i(t)\beta_j\right)\,, \tag{6.9}$$

a model with a basic renewal structure for type j events which may be modulated by other information on past events.

In some settings, it may be appropriate to use Markov time scales for modeling the occurrence of some event types and semi-Markov time scales for others. For example, consider a machine which breaks down from two causes. One cause (type 1) may be due to failure of a particular part which is replaced upon failure and the other cause (type 2) may be due to the general wear of another part which is repairable but cannot be easily replaced. A semi-Markov model could be used for type 1 failures because the replacement of the part constitutes a renewal; a Markov time scale could reflect the generally increasing failure rate with increasing wear on the repairable part. Of course, intensity functions with both time scales can also be used (see Section 5.1).

6.3 Random Effect Models for Multitype Events

Intensity-based analyses provide explicit expressions for the dependence between event processes through time-dependent stratification and covariates. Sometimes, however, it is of interest to formulate models through the introduction of random effects. When appropriate, these models can provide a parsimonious representation of dependence or association between event types. Let

$$\lambda_{ij}\left(t|H_i(t),u_{ij}\right)=\lim_{\Delta t\to 0}\frac{\Pr\left(\Delta N_{ij}(t)=1|H_i(t),u_{ij}\right)}{\Delta t} \tag{6.10}$$

denote the event intensity at time t for events of type j, conditional on subject and type-specific random effect $u_{ij}>0$, and covariate and event history $H_i(t)$. Multiplicative random effect models are discussed in Sections 2.2.3 and 3.5. In the present framework we consider

$$\lambda_{ij}\left(t|H_i(t),u_{ij}\right)=u_{ij}\lambda_{ij}\left(t|H_i(t)\right)\,,$$

where $u_{ij}\lambda_{ij}\left(t|H_i(t)\right)$ is referred to as the conditional intensity function for type j events. Let $u_i=(u_{i1},\ldots,u_{iJ})'$ denote the multivariate random effect, which we assume arises from distribution $G(u_i;\phi)$. So-called "genuine multivariate" random effect models involve separate parameters for variances and covariances of the component random effects. These are typically parameterized so that $E(u_{ij})=1$, $\mathrm{var}(u_{ij})=\phi_j$, and $\mathrm{cov}(u_{ij},u_{ik})=\phi_{jk}$, for $j,k=1,2,\ldots,J$.

The derivation of (6.2) with $\lambda_{ij}(t|H_i(t))$ replaced with $\lambda_{ij}\left(t|H_i(t),u_i\right)=u_{ij}\lambda_{ij}\left(t|H_i(t)\right)$ gives the likelihood conditional on u_i as

$$\prod_{j=1}^{J}\left\{\prod_{k=1}^{n_{ij}}u_{ij}\lambda_{ij}\left(t_{ijk}|H_i(t_{ijk})\right)\exp\left(-u_{ij}\int_0^{\tau_i}\lambda_{ij}\left(u|H_i(u)\right)du\right)\right\}\,,$$

and the marginal likelihood for individual i as

$$\int\prod_{j=1}^{J}\left\{\prod_{k=1}^{n_{ij}}u_{ij}\lambda_{ij}(t_{ijk}|H_i(t_{ijk}))\exp\left(-u_{ij}\int_0^{\tau_i}\lambda_{ij}(u|H_i(u))du\right)\right\}dG(u_i;\phi)\,. \tag{6.11}$$

With multivariate random effects, it is often assumed that $\lambda_{ij}\,(t|H_i(t)) = \lambda_{ij}\,(t|H_{ij}(t))$, where $H_{ij}(t) = \{N_{ij}(s) : 0 \le s < t;\ x_{ij}\}$, because the general approach is to assume independence of different event types, given the random effect. Mixed Poisson models are obtained if $\lambda_{ij}\,(t|H_{ij}(t)) = \rho_{ij}(t)$. As in Section 3.5, we then find $E\{N_{ij}(t)\} = \mu_{ij}(t) = \int_0^t \rho_{ij}(u)du$, $\text{var}\{N_{ij}(t)\} = \mu_{ij}(t) + \phi_j\mu_{ij}^2(t)$, and by the conditional covariance formula,

$$\text{cov}\,\{N_{ij}(s_1,t_1), N_{ij}(s_2,t_2)\} = \phi_j\,\mu_{ij}(s_1,t_1)\mu_{ij}(s_2,t_2)\,, \tag{6.12}$$

for nonoverlapping time intervals (s_1,t_1) and (s_2,t_2). Here however, we obtain as well, for $j \ne k$ and arbitrary intervals,

$$\text{cov}\,\{N_{ij}(s_1,t_1), N_{ik}(s_2,t_2)\} = \phi_{jk}\,\mu_{ij}(s_1,t_1)\mu_{ik}(s_2,t_2)\ . \tag{6.13}$$

As before, the variances reflect extra-Poisson variation and association of the event counts from each process, and by (6.13) the covariances between the random effects accommodate association among the different types of events.

In such multivariate mixed Poisson models, the overall likelihood conditional on $u_i = (u_{i1},\ldots,u_{iJ})'$ is of the form

$$\prod_{j=1}^{J}\left\{\prod_{k=1}^{n_{ij}} u_{ij}\rho_{ij}(t_{ijk})\exp\left(-u_{ij}\mu_{ij}(\tau_i)\right)\right\}\,,$$

and the marginal likelihood is given by (6.11). An appealing choice is to take u_i as multivariate log-normal, with an unrestricted covariance matrix; in this case, the vector $(\log u_{i1},\ldots,\log u_{iJ})'$ has a multivariate normal distribution. The marginal likelihood is then

$$\int_0^\infty\cdots\int_0^\infty \prod_{j=1}^{J}\prod_{k=1}^{n_{ij}} u_{ij}\rho_{ij}(t_{ijk})\exp\left(-u_{ij}\mu_{ij}(\tau_i)\right)\,dG(u_i;\phi)\,, \tag{6.14}$$

and is generally intractable. However, with a parametric specification of $\rho_{ij}(t)$, say as $\rho_j(t;\alpha_j)\exp(x_{ij}'(t)\beta_j)$, and if there are not too many different types of events, then (6.14) may be evaluated by numerical integration and maximized with a general-purpose optimization program to obtain $(\widehat{\alpha},\widehat{\beta},\widehat{\phi})$. For larger J, simulation-based methods of evaluating the integral in (6.14) can be used. Standard maximum likelihood results may be used for inference. Profile likelihood methods are convenient for interval estimation and testing of parameters. Alternatively the observed information matrix, approximated by numerical differentiation of the log-likelihood, is given by good optimization software (Appendix B.2), and provides standard errors for estimates.

Semiparametric specifications of $\rho_{ij}(t)$ are possible to fit, but are considerably more challenging numerically; the EM algorithm (see Section 3.5.3)

offers one approach but the required conditional expectations can be difficult to obtain and Monte Carlo methods may prove useful.

A simpler so-called "shared" random effect model is obtained if we set $u_{ij} = u_i,\ j = 1, 2, \ldots, J$, where $E(u_i) = 1$ and $\text{var}(u_i) = \phi$. In this case

$$\text{var}\{N_{ij}(s,t)\} = \mu_{ij}(s,t) + \phi\, \mu_{ij}^2(s,t)\,,$$

$$\text{cov}\{N_{ij}(s_1,t_1), N_{ij}(s_2,t_2)\} = \phi\, \mu_{ij}(s_1,t_1)\mu_{ij}(s_2,t_2)\,,$$

for nonoverlapping intervals (s_1, t_1) and (s_2, t_2), and

$$\text{cov}\{N_{ij}(s_1,t_1), N_{ik}(s_2,t_2)\} = \phi\, \mu_{ij}(s_1,t_1)\mu_{ik}(s_2,t_2)\,,$$

$j \neq k$ and the integration in (6.14) reduces to the one-dimensional integration of Section 3.5. In the shared random effect setting, the gamma distribution (2.28) is convenient to adopt because it leads to a tractable likelihood based on a negative multinomial process; if the values τ_i are fixed, the probability function for the numbers of events $n_{i1}, \ldots, n_{iJ}$ of each type is then

$$L_{1i} = \frac{\Gamma(\phi^{-1} + n_{i\cdot})}{\Gamma(\phi^{-1}) n_{i1}! \ldots n_{iJ}!} \times \left(\frac{1}{1 + \phi\mu_{i\cdot}(\tau_i)}\right)^{\phi^{-1}} \left(\frac{\phi\mu_{i1}(\tau_i)}{1 + \phi\mu_{i\cdot}(\tau_i)}\right)^{n_{i1}} \cdots \left(\frac{\phi\mu_{iJ}(\tau_i)}{1 + \phi\mu_{i\cdot}(\tau_i)}\right)^{n_{iJ}},$$

and the full likelihood contribution is $L_i = L_{1i} L_{2i}$, where

$$L_{2i} = \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \left\{ \frac{\rho_{ij}(t_{ijk})}{\mu_{ij}(\tau_i)} \right\}.$$

The same likelihood L_i, without the $n_{ij}!$ terms in L_{1i}, applies when the τ_i are not prespecified. This is the extension of (3.32) for dealing with J types of events. The semiparametric model with $\rho_{ij}(t) = \rho_{0j}(t) \exp(x'_{ij}\beta_j)$ for $j = 1, \ldots, J$ can be fit using `coxph` in S-PLUS or R with the `frailty` option and a "covariate by `strata`" interaction to accommodate different baseline rates and different regression coefficients for the different event types. Although this type of model leads to convenient analytic forms for the marginal likelihood, it is very restrictive because it implies the same degree of heterogeneity for the different event types, and does not have a separate parameter to characterize associations between processes. Moreover, it is suitable only in settings where a positive association is anticipated between event types.

More flexible models with tractable marginal likelihoods can be obtained by defining additive random effects $u_{ij} = v_i + w_{ij}$ where v_i has mean μ_v and variance ϕ_v and w_{ij}, $j = 1, 2, \ldots, J$, are independently distributed with common mean μ_w and variance ϕ_w. Here $E(v_i + w_{ij}) = \mu_v + \mu_w$, $\text{var}(u_{ij}) = \phi_v + \phi_w$ and $\text{cov}(u_{ij}, u_{ik}) = \phi_v$. Such models are also most useful in settings where positive correlations are anticipated; they are more general than the

shared random effect model because they have a separate parameter for the covariance of the random effects. Models with additive random effects can lead to marginal likelihoods of closed form with suitable choice of distribution for v_i and w_{ij}; in particular, if these are taken to be gamma, closed-form marginal likelihoods are obtained.

The focus here has been on the use of multivariate random effects with conditionally Poisson models. Multivariate random effects can also be used with other conditional intensity functions, including those based on semi-Markov or hybrid time scales.

6.4 Robust Methods for Multitype Events

6.4.1 Methods Based on Working Independence Assumptions

Robust methods for rate and mean functions of multivariate processes, analogous to the methods of Section 3.6.3, are also available. Here we assume

$$E\{dN_{ij}(t)|x_i^{(\infty)}\} = \rho_{0j}(t)dt \exp\left(x'_{ij}(t)\beta_j\right) \qquad t > 0\,, \tag{6.15}$$

$j = 1, 2, \ldots J$, where $x_{ij}(t)$ is a $p_j \times 1$ vector of external covariates and $x_i^{(\infty)} = \{x_i(t) : 0 \le t\}$. The observation or censoring process $\{Y_i(t), 0 \le t\}$ is assumed independent of the event processes. In the following, we let τ denote max τ_i.

If we are not interested in estimation of variance or association parameters, then it is convenient to base estimation of the $\mu_{0j}(t)$ and $\beta = (\beta'_1, \ldots, \beta'_J)'$ on the Poisson estimating functions (3.19) and (3.21) under a "working independence" assumption. In this case, we use the extension of $U_\beta(\beta)$ in Section 3.4.2 with $d\bar{N}_{\cdot j}(s) = \sum_{i=1}^m Y_i(s)dN_{ij}(s)$ and set

$$\sum_{i=1}^m \int_0^\tau Y_i(s)\, x_{ij}(s) \left[dN_{ij}(s) - \frac{d\bar{N}_{\cdot j}(s)}{\sum_{\ell=1}^m Y_\ell(s) \exp(x'_{\ell j}(s)\beta_j)} \exp(x'_{ij}(s)\beta_j)\right] \tag{6.16}$$

equal to zero to obtain $\widehat{\beta}_j$. The extension of (3.23) gives the estimates

$$d\widehat{\mu}_{0j}(s) = \frac{d\bar{N}_{\cdot j}(s)}{\sum_{i=1}^m Y_i(s) \exp(x'_{ij}(s)\widehat{\beta}_j)}\,, \tag{6.17}$$

and we then take $\widehat{\mu}_{0j}(t) = \int_0^t d\widehat{\mu}_{0j}(s)$ as the estimate of the baseline mean function for events of type j, $j = 1, 2, \ldots, J$.

Robust inference regarding β is possible by writing (6.16) as

$$U_{\beta_j}(\beta_j) = \sum_{i=1}^m \int_0^\tau Y_i(s) W_{ij}(s; \beta_j) dN_{ij}(s) \tag{6.18}$$

with

$$W_{ij}(s;\beta_j) = x_{ij}(s) - \frac{\sum_{\ell=1}^{m} Y_\ell(s)\exp(x'_{\ell j}(s)\beta_j)x_{\ell j}(s)}{\sum_{\ell=1}^{m} Y_\ell(s)\exp(x'_{\ell j}(s)\beta_j)}\,. \tag{6.19}$$

Then if $U(\beta) = (U'_{\beta_1}(\beta_1), \ldots, U'_{\beta_J}(\beta_J))'$, $\text{var}\{\sqrt{m}^{-1}U_{\beta_j}(\beta_j)\}$ is given along the lines of Section 3.6.3 by

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^\tau\int_0^\tau Y_i(u)Y_i(v)W_{ij}(u;\beta_j)W'_{ij}(v;\beta_j)\ \text{cov}\,\{dN_{ij}(u), dN_{ij}(v)\} \tag{6.20}$$

and $\text{cov}\{\sqrt{m}^{-1}U_{\beta_j}(\beta_j), \sqrt{m}^{-1}U_{\beta_k}(\beta_k)\}$ is

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^\tau\int_0^\tau Y_i(u)Y_i(v)W_{ij}(u;\beta_j)W'_{ik}(v;\beta_k)\ \text{cov}\,\{dN_{ij}(u), dN_{ik}(v)\}\,. \tag{6.21}$$

Subject to mild regularity conditions and assuming that $m^{-1}\sum Y_i(t)$ approaches a positive limit in probability as $m \to \infty$, for $0 \le t \le \tau$, (6.20) and (6.21) are consistently estimated by

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^\tau \int_0^\tau Y_i(u)Y_i(v)W_{ij}(u;\widehat{\beta}_j)W'_{ij}(v;\widehat{\beta}_j)d\widehat{M}_{ij}(u)d\widehat{M}_{ij}(v) \tag{6.22}$$

and

$$\frac{1}{m}\sum_{i=1}^{m}\int_0^\tau \int_0^\tau Y_i(u)Y_i(v)W_{ij}(u;\widehat{\beta}_j)W'_{ik}(v;\widehat{\beta}_k)d\widehat{M}_{ij}(u)d\widehat{M}_{ik}(v)\,, \tag{6.23}$$

respectively, where $d\widehat{\mu}_{ij}(u) = d\widehat{\mu}_{0j}(u)\exp(x'_{ij}(u)\widehat{\beta}_j)$ and $d\widehat{M}_{ij}(u) = dN_{ij}(u) - d\widehat{\mu}_{ij}(u)$, $j = 1, \ldots, J$. These can be rewritten in a form like that following (3.37). If $\beta = (\beta'_1, \ldots, \beta'_J)'$, then following the arguments of Section 3.6.3, we obtain

$$\widehat{V}(\widehat{\beta}) = \widehat{\text{asvar}}\left\{\sqrt{m}(\widehat{\beta} - \beta)\right\} = A^{-1}(\widehat{\beta})B(\widehat{\beta})\left[A^{-1}(\widehat{\beta})\right]'\,, \tag{6.24}$$

where $\widehat{\beta}$ is $p \times 1$,

$$A(\widehat{\beta}) = -m^{-1}\partial U(\widehat{\beta})/\partial\widehat{\beta}'$$

is a $p \times p$ block diagonal matrix and $B(\widehat{\beta})$ is a $p \times p$ matrix with diagonal blocks given by (6.22) and off-diagonal blocks given by (6.23). The matrices $A(\widehat{\beta})$ and $B(\widehat{\beta})$ estimate the limits of

$$\mathcal{A}(\beta) = m^{-1}E\,\{-\partial U(\beta)/\partial\beta'\}$$

and

$$\mathcal{B}(\beta) = m^{-1} E\left\{U(\beta)U'(\beta)\right\} ,$$

respectively, and the asymptotic variance of $\sqrt{m}(\widehat{\beta}-\beta)$ is $\mathcal{A}^{-1}(\beta)\mathcal{B}(\beta)\mathcal{A}^{-1}(\beta)$.

When the covariates are either fixed or piecewise-constant, the integrals in (6.22), (6.23), and in $A(\widehat{\beta})$ are simply sums. Because $A(\widehat{\beta})$ is block diagonal, the diagonal blocks in (6.24) that give asvar$\{\sqrt{m}(\widehat{\beta}_j-\beta_j)\}$ for $j = 1, \ldots, J$ are the estimated covariance matrices obtained by considering each of the types of events on their own. As discussed in Section 3.6.3 and 3.8.1, the function `coxph` in S-PLUS or R provides these. The off-diagonal blocks in (6.24) if needed, currently have to be computed directly. The following paragraph gives one such instance.

The robust estimate (6.24) allows simultaneous inferences on $\beta_1, \ldots, \beta_J$. For example, suppose $x_{ij}(t) = x_i$ is a scalar taking the value 1 for subjects in a treatment group and 0 for subjects in a control group. Then if the null hypothesis is that there is no effect of treatment on any of the event rates, one may write $H_0 : \beta_j = 0$, $j = 1, 2, \ldots, J$. A global test of H_0 may be based on the statistic

$$T = m\widehat{\beta}'\widehat{V}^{-1}(\widehat{\beta})\widehat{\beta} .$$

This test statistic is asymptotically χ^2_J under H_0 and so the $p-$value is computed as $\Pr(\chi^2_J > t)$ where t is the realized value of T. If it is anticipated that the effect of treatment will be comparable for all event types, then a more powerful test for detecting effects in the same direction is based on a pooled statistic $m\,C'\widehat{V}^{-1}(\widehat{\beta})\widehat{\beta}$ where $C = (1, 1, \ldots, 1)'$, which has variance $m^2 C'\widehat{V}^{-1}(\widehat{\beta})C$ and gives a test statistic

$$T = \frac{C'\widehat{V}^{-1}(\widehat{\beta})\widehat{\beta}}{\sqrt{C'\widehat{V}^{-1}(\widehat{\beta})C}}$$

which is asymptotically standard normal under the null hypothesis H_0. In this case the $p-$value is $2\Pr(U > |t|)$, where U is a standard normal random variable.

6.4.2 Robust Methods with Covariance Functions

With multivariate processes, one can generalize the approaches of Section 3.6.4 to estimate the variances and covariances of event counts. Consider independent bivariate point processes with $N_i(t) = (N_{i1}(t), N_{i2}(t))'$ and increments $dN_i(t) = (dN_{i1}(t), dN_{i2}(t))'$, $i = 1, \ldots, m$. Let $\rho_{ij}(t) = \rho_{0j}(t; \alpha_j)\exp(x'_{ij}(t)\beta_j)$ denote the marginal rate for type j events for subject i and $\mu_{ij}(t) = \int_0^\infty \rho_{ij}(s)ds$ the respective mean function. Let $\theta_j = (\alpha'_j, \beta'_j)'$ denote the vector of parameters indexing the marginal mean for type j events, let $\theta = (\theta'_1, \theta'_2)'$ be of dimension p, and let $\psi = (\theta', \phi')'$ denote the full vector of unknown parameters, where $\phi = (\phi_1, \phi_2, \phi_{12})'$ contains the variance and covariance parameters.

Poisson estimating functions can be used for θ, which we write as

$$U_1(\psi) = \sum_{i=1}^{m} \int_0^{\infty} Y_i(t) a_i(t) \left[dN_i(t) - \rho_i(t) dt \right] , \tag{6.25}$$

where $a_i(t)$ is a $p \times 2$ weight matrix implied by (3.4) and (3.5) in the parametric setting or (3.19) and (3.5) in the semiparametric setting, and $\rho_i(t) = (\rho_{i1}(t), \rho_{i2}(t))'$.

Consider now a model for association based on mixed Poisson processes where, given a bivariate random effect $u_i = (u_{i1}, u_{i2})'$,

$$\begin{aligned} E\{dN_{ij}(t)|u_i\} &= u_{ij}\rho_{ij}(t)dt \\ \operatorname{cov}\{dN_{ij}(s), dN_{ij}(t)|u_i\} &= I(s=t)u_{ij}\rho_{ij}(t)dt \\ \operatorname{cov}\{dN_{i1}(s), dN_{i2}(t)|u_i\} &= 0\,. \end{aligned}$$

Then if $E(u_i) = (1,1)'$ and

$$\operatorname{var}(u_i) = \Sigma_u = \begin{bmatrix} \phi_1 & \phi_{12} \\ \phi_{12} & \phi_2 \end{bmatrix} ,$$

we have $\operatorname{var}(N_{ij}(t)) = \mu_{ij}(t) + \phi_j \mu_{ij}^2(t)$ and the covariances are given by (6.12) and (6.13). Setting (6.25) equal to zero gives $\widehat{\theta}$ and thus $\widehat{\rho}_i(t)$ and $\widehat{\mu}_{ij}(\tau_i)$, $j = 1, 2$. These can be plugged into the generalization of (3.40), to give moment estimates

$$\widehat{\phi}_j = \frac{\sum_{i=1}^{m} \left[(n_{ij} - \widehat{\mu}_{ij}(\tau_i))^2 - \widehat{\mu}_{ij}(\tau_i) \right]}{\sum_{i=1}^{m} \widehat{\mu}_{ij}^2(\tau_i)} \qquad j = 1, 2\,, \tag{6.26}$$

and

$$\widehat{\phi}_{12} = \frac{\sum_{i=1}^{m} \left[(n_{i1} - \widehat{\mu}_{i1}(\tau_i))\,(n_{i2} - \widehat{\mu}_{i2}(\tau_i)) \right]}{\sum_{i=1}^{m} \widehat{\mu}_{i1}(\tau_i)\widehat{\mu}_{i2}(\tau_i)}\,. \tag{6.27}$$

Choices for $a_i(t)$ in (6.25) which may depend on ϕ can lead to more efficient estimation of θ and more efficient estimators of ϕ may be considered for the mixed Poisson working model, but in general optimal specifications are difficult to make. However, an approach that should have high efficiency under certain mixed Poisson processes with gamma random effects is to obtain $\widehat{\phi}_1$ and $\widehat{\phi}_2$ from estimating equations like the one following (3.40), and to use an analogous equation for $\widehat{\phi}_{12}$.

Although the estimates of variances and covariances for multitype event data given here are based on mixed Poisson processes, they do provide some insight into the mean–variance–covariance relationship more generally. Another way to get insight is to examine values of $[n_{ij} - \widehat{\mu}_{ij}(\tau_i)]^2$ versus $\widehat{\mu}_{ij}(\tau_i)$ and $[n_{i1} - \widehat{\mu}_{i1}(\tau_i)][n_{i2} - \widehat{\mu}_{i2}(\tau_i)]$ versus $\widehat{\mu}_{i1}(\tau_i)\widehat{\mu}_{i2}(\tau_i)$. An alternative is to examine the Poisson residuals $[n_{ij} - \widehat{\mu}_{ij}(\tau_i)]/[\widehat{\mu}_{ij}(\tau_i)]^{1/2}$, as is commonly done with log-linear models for count data (see McCullaugh and Nelder, 1989).

6.5 Alternating Two-State Processes

Often recurrent events have a duration associated with them, as mentioned in Section 1.5.3. Such processes may be better characterized as having recurrent episodes rather than recurrent events, but there are considerable similarities in the types of questions posed and the models used. Examples include exacerbations in patients with chronic bronchitis, failures in systems which are inoperable while being repaired, and admissions to hospital for psychiatric illnesses. In the latter example, an admission arises as a consequence of a clinically important deterioration of health, and the number of admissions reflects, to some extent, the burden of the disease. Simply counting the number of hospitalizations is not necessarily sensible, however, because some admissions may lead to long stays and others to shorter stays; in the latter, subjects become at risk for readmission sooner. The duration of previous spells in hospital may also be predictive of future admissions.

In some settings, the duration of an episode is approximately fixed. For example, episodes of infection which are treated with medication may last one week or so. In other settings, the durations can be quite variable and models which incorporate the duration are appealing. Alternating two-state processes are the canonical models of this type; they feature an "active" and an "inactive" state, and two types of recurrent events corresponding to the two types of transitions between these states. The intensity-based models of Section 6.2 are relatively straightforward to adopt for such alternating processes. Models based on gap times or sojourn times in each state often have considerable appeal. As discussed in Chapter 5, models featuring hybrid time scales are also appealing when a basic semi-Markov structure is modified by calendar time trends. When covariate data are limited, it may be useful to introduce random effects to reflect unexplained heterogeneity and model association between sojourn times.

Suppose m individuals have independent two-state processes with states labeled 1 and 2. We refer to the time since the initiation of the process as the Markov or calendar time scale and denote it $t \geq 0$. The time t may be, for example, the time since the diagnosis of a chronic disease, the time since enrollment in a study, or the age of an individual. Let $N_{ij}(t)$ denote the number of $j \to 3-j$ transitions ($j = 1, 2$) for subject i over the interval $[0, t]$, $i = 1, \ldots, m$. Furthermore, let $Y_{ij}(t) = 1$ if individual i is in state j at time t^-, and 0 otherwise. The history of the ith process up to time t is denoted

$$H_i(t) = \{N_{ij}(s) : 0 \leq s < t;\, z_{ij}(s) : 0 \leq s \leq t,\, j = 1, 2\} \,,$$

where $z_{ij}(t)$ is a $p_j \times 1$ vector of covariates reflecting the process history, external time-varying conditions, and fixed subject characteristics.

Intensity-based models are often most natural in a modulated alternating renewal form

$$\lambda_{ij}(t|H_i(t)) = Y_{ij}(t) h_{0j}(B_{ij}(t)) \exp(z'_{ij}(t)\beta_j) \,,$$

where $B_{ij}(t)$ is the time since the latest transition into state j for subject i, and β_j is a vector of regression coeffients. In some settings it may be desirable to allow the baseline functions to be different depending on the cumulative number of sojourns, but typically covariate effects are assumed to be the same.

If we introduce random effects we define conditional transition intensities as $\lambda_{ij}(t|H_i(t), u_{ij})$. Figure 6.1 is a representation of this process where we drop the index for individuals. We assume that the covariates and random effects act multiplicatively on the baseline transition intensities so that given the random effect $u_i = (u_{i1}, u_{i2})'$ and covariates, the conditional intensity for the $j \to 3 - j$ transition is

$$\lambda_{ij}\left(t|H_i(t), u_i\right) = Y_{ij}(t)u_{ij}h_{0j}(B_{ij}(t))\exp(z'_{ij}(t)\beta_j) \tag{6.28}$$

where $h_{0j}(B_{ij}(t))$ is a baseline transition intensity indexed by a parameter α_j, $j = 1, 2$. We denote $\theta_j = (\alpha'_j, \beta'_j)'$ and assume the u_i are i.i.d. for $i = 1, \ldots, m$. Finally, let $\theta = (\theta'_1, \theta'_2)'$ denote the full vector of parameters.

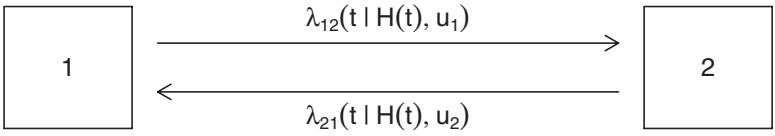


Fig. 6.1. State diagram for an alternating two-state process.

An attractive distribution for (u_{i1}, u_{i2}) is the bivariate log-normal distribution. In that case, $v_{ik} = \log u_{ik}$ and $v_i = (v_{i1}, v_{i2})'$ are i.i.d. $i = 1, \ldots, m$ as $N(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \phi_1 & \phi_{12} \\ \phi_{12} & \phi_2 \end{pmatrix}$$

is an unknown 2×2 covariance matrix parameterized by $\phi = (\phi_1, \phi_2, \phi_{12})'$. If the random effects are negatively correlated, subjects with shorter active periods will tend to have longer inactive periods, and vice versa. A model admitting a negative correlation is often desirable, for example, for chronic diseases in which states 1 and 2 represent good and poor health.

Consider data on m independent subjects, obtained by observing subject i over the time period $[\tau_{i0}, \tau_i]$, $i = 1, \ldots, m$. Let n_i denote the total number of transitions observed to take place over $[\tau_{i0}, \tau_i]$ for subject i, and let n_{ij} denote the total number of complete sojourns in state j. Usually τ_{i0} is the time of entry to the study and τ_i is a censoring time denoting the end of followup.

We assume that τ_{i0} and τ_i are stopping times, as discussed in Section 2.6. Let $t_{i1} < \cdots < t_{in_i}$ be the observed transition times for subject i occurring during the course of followup.

It follows from (6.2) that the likelihood for subject i, conditional on u_i and denoted $L_i(\theta|u_i)$, is given by

$$\prod_{j=1}^{2}\left[\prod_{k\in\mathcal{S}_{ij}}\lambda_{ij}(t_{ik}|H_i(t_{ik}),u_{ij})\exp\left(-\int_{\tau_{i0}}^{\tau_i}Y_{ij}(s)\lambda_{ij}(s|H_i(s),u_{ij})ds\right)\right], \quad (6.29)$$

where $\mathcal{S}_{ij} = \{k|dN_{ij}(t_{ik}) = 1\}$, and $Y_{ij}(s) = I$ (individual i is in state j at time s^-), for $j = 1, 2$; $i = 1, \ldots, m$. The marginal likelihood for subject i is the expectation of (6.29) with respect to the random effect,

$$L_i(\psi) = E_{u_i}\left\{L_i(\theta|u_i)\right\},$$

where $\psi = (\theta', \phi')'$. The marginal log-likelihood based on all m subjects is

$$\ell(\psi) = \sum_{i=1}^{m}\ell_i(\psi) = \sum_{i=1}^{m}\log(L_i(\psi)).$$

It is generally necessary to use numerical integration to compute such likelihoods. When u_{i1} and u_{i2} are independent, the likelihood factors into two pieces for transitions of type $j = 1$ and $j = 2$, and in that case software such as `coxph` in S-PLUS or R can often be used. When u_{i1} and u_{i2} are correlated, there is no such factorization and then we use general optimization software to implement the analysis. An example of an alternating process is considered in Section 6.7.2.

In many studies, the time of randomization does not coincide with a transition time or the origin of the process. In that case matters are more complicated for models with random effects because the relevant distribution of u_i over which we must average is the conditional distribution of u_i, given $H_i(\tau_{i0})$; see Section 4.5.

6.6 Recurrent Events with a Terminal Event

In many settings, interest lies in characterizing the incidence of recurrent events in conjunction with an event which terminates the recurrent event process. For simplicity we return to the setting where recurrent events are of a single type, and we think of the terminal event as a second type of event. Examples are ubiquitous and include any setting involving patients with a serious disease which is associated with both recurrent complications and high mortality. In neurovascular trials, for example, one may be interested in reducing the occurrence of transient ischemic attacks and mild strokes, but deaths from major strokes or any other cause may also occur. In oncology, one may

be interested in characterizing the use of health services following diagnosis of cancer, but use of such services terminates upon death. Methods for the analysis of recurrent and terminal events are also of interest in reliability, where certain types of failure necessitate the retirement of a piece of equipment. A multistate diagram representing this sort of process is given in Figure 1.9. We next consider several approaches for this important setting.

6.6.1 Intensity-Based Approaches

General models for recurrent events and failures are provided by the theory of multivariate counting processes used with multiple events. As before, we let $\Delta N_i(t)$ denote the number of recurrent events over the small interval $[t, t+\Delta t)$. Let T_i denote the terminal event time for subject i and define $D_i(t) = I(t \leq T_i)$ and $\bar{Y}_i(t) = D_i(t)I(t \leq C_i)$, where C_i is a censoring time corresponding to the end of followup. If $H_i(t) = \{(N_i(s), D_i(s)) : 0 \leq s < t\}$ represents the process history up to time t, a full model for the process may be expressed in terms of the event intensity functions

$$\lambda_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N_i(t) = 1|H_i(t)\}}{\Delta t} \tag{6.30}$$

$$\gamma_i(t|H_i(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{T_i < t + \Delta t|H_i(t), D_i(t) = 1\}}{\Delta t}\,. \tag{6.31}$$

Covariates can also be added to the histories $H_i(t)$ as in previous sections.

As can be seen from Figure 1.9, the presence of a terminal event creates a recurrent quasi-competing risk problem in the sense that whenever patients are at risk of another recurrent event, they are also at risk of the terminal event. If n_i recurrent events are observed at times $t_{i1}, \ldots, t_{in_i}$ over $[0, \tau_i]$ where $\tau_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$, then under independent censoring the likelihood function is, by (6.2), proportional to

$$\prod_{j=1}^{n_i} \lambda_i\left(t_{ij}|H_i(t_{ij})\right)\left[\gamma_i(\tau_i|H_i(\tau_i))\right]^{\delta_i} \tag{6.32}$$
$$\times \exp\left(-\int_0^{\tau_i} \left[\lambda_i(u|H_i(u)) + \gamma_i(u|H_i(u))\right] du\right)$$

and inferences may be based on partial likelihoods arising from the factorization of this into two pieces (assuming (6.30) and (6.31) do not share any parameters).

Models for the recurrent event intensity (6.30) can be based on the approaches in Chapters 2 to 5, although it should be kept in mind that the termination time that leads to the competing risk problem must be addressed when computing summary statistics for marginal features of the process. The

hazard or intensity for the terminal event can include dependency on the history of recurrent events, for example, as in

$$\gamma_i(t|H_i(t)) = \gamma_0(t) \exp\left(x_i'(t)\beta + \zeta N_i(t^-)\right) .$$

More general models for the terminal event could include some form of stratification, say $\gamma_i(t|H_i(t)) = \gamma_{0k}(t) \exp(x_i'(t)\beta)$ if $N_i(t^-) = k$, or information on recent gap times. These are all Cox regression models with internal time-dependent covariates reflecting the recurrent event history.

Intensity-based models are appealing in their flexibility but they can lead to complicated expressions for marginal features that are often of interest, such as the marginal survivor function for T and the cumulative mean function for the number of recurrent events. We next consider random effects models for which marginal features are readily obtained. We show, however, that here too such features do not typically take a simple form.

6.6.2 Random Effects Models

The approach described in Section 6.3 can be specialized to the present setting. Let u_{i1}, u_{i2} be positive-valued random effects with joint distribution function $G(u_{i1}, u_{i2})$ and suppose that conditional on $u_i = (u_{i1}, u_{i2})'$ and external covariates $z_i(t)$, the bivariate process $\{N_i(t), D_i(t), 0 \le t\}$ has intensities

$$\lambda_i(t|u_i, H_i(t)) = u_{i1}\lambda_i(t|H_i(t))$$

and

$$\gamma_i(t|u_i, H_i(t)) = u_{i2}\gamma_i(t|H_i(t))$$

for recurrent and terminal events, respectively. Such models are most useful when the random effects allow event history to be dropped from the conditional intensities, and when marginal features such as $E\{N_i(t)|T_i > t\}$ or $E\{N_i(T_i)\}$ have fairly simple forms. Therefore, we make the additional assumptions that

$$\lambda_i(t|u_i, H_i(t)) = u_{i1}D_i(t)\lambda_0(t) \exp\left(z_i'\beta\right) , \tag{6.33}$$

$$\gamma_i(t|u_i, H_i(t)) = u_{i2}D_i(t)\gamma_0(t) \exp\left(z_i'\alpha\right) . \tag{6.34}$$

That is, the model for recurrent events is conditionally Poisson, although it is terminated at the time of death. In addition, the association between time of death T_i and the recurrent events is entirely driven by the association between u_{i1} and u_{i2}. Time-varying covariates can also be accommodated in (6.33) and (6.34) but for simplicity we denote them as fixed here.

It is often assumed that $u_{i1} = u_{i2}$ or that u_{i1} and u_{i2} are related parametrically, but it is generally preferable to avoid this, because overdispersion in the recurrent events and their association with T_i is then determined by a single

factor. If C_i is a censoring time and $\tau_i = \min(T_i, C_i)$ then by (6.32) the conditional likelihood function based on data $\{\delta_i = I(T_i \le C_i), \tau_i, z_i, N_i(t), 0 \le t \le \tau_i\}$ for individual i is

$$\left\{\prod_{j=1}^{n_i} u_{i1}\lambda_i(t_{ij})\right\} \exp\left(-u_{i1}\Lambda_i(\tau_i)\right) \cdot \{u_{i2}\gamma_i(\tau_i)\}^{\delta_i} \exp\left(-u_{i2}\Gamma_i(\tau_i)\right) ,$$

where $n_i = N_i(\tau_i)$ and for simplicity we write $\lambda_i(t)$ for $\lambda_0(t)\exp(z_i'\beta)$ and $\gamma_i(t)$ for $\gamma_0(t)\exp(z_i'\alpha)$, and where $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$ and $\Gamma_i(t) = \int_0^t \gamma_i(s)ds$. The marginal likelihood based on the observable data is then

$$L_i(\psi) = \left\{\prod_{j=1}^{n_i} \lambda_i(t_{ij})\right\} \gamma_i(\tau_i)^{\delta_i} \times \int_0^\infty \int_0^\infty u_{i1}^{n_i} u_{i2}^{\delta_i} \exp\left(-u_{i1}\Lambda_i(\tau_i) - u_{i2}\Gamma_i(\tau_i)\right) dG(u_{i1}, u_{i2}) ,$$

where ψ includes parameters in $\lambda_i(t), \gamma_i(t)$, and $G(u_{i1}, u_{i2})$.

As discussed in Section 6.3, a flexible and feasible approach is to assume that $(\log u_{i1}, \log u_{i2})$ has a zero-mean bivariate normal distribution, and to adopt parametric models for $\lambda_0(t)$ and $\gamma_0(t)$ in (6.33) and (6.34). Other models, such as additive gamma random effects models mentioned in Section 6.3, can also be considered.

It should be noted that although such models allow a simple conditional (on the random effect) interpretation of covariate effects on the mean function for recurrent events, they do not offer simple marginal features. For example, a feature of interest is

$$\mu_i(s;t) = E\left\{N_i(s) | T_i = t\right\} , \qquad 0 \le s \le t .$$

This is given by

$$\begin{aligned} \mu_i(s;t) &= E\left[E\left\{N_i(s)|T_i = t, u_i\right\}\right] \\ &= E\left(u_{i1}|T_i = t\right)\Lambda_i(s) . \end{aligned} \tag{6.35}$$

Because T_i is conditionally independent of u_{i1}, given u_{i2}, the conditional mean of u_{i1} given $T_i = t$ is found to be

$$E\left(u_{i1}|T_i = t\right) = \frac{\int_0^\infty f_i(t|u_{i2})g_2(u_{i2})E(u_{i1}|u_{i2})du_{i2}}{\int_0^\infty f_i(t|u_{i2})g_2(u_{i2})du_{i2}} , \tag{6.36}$$

where $f_i(t|u_{i2})$ denotes the density of T_i given u_{i2}, and $g_2(u_{i2})$ denotes the marginal density of u_{i2}. In general, (6.36) will be a complicated function of the covariates z_i and so the effect of covariates in (6.35) will be complicated. A similar calculation applies for $E\{N_i(s)|T_i \ge t\}$ for $0 \le s \le t$. Note that

when u_{i1} and u_{i2} are independent we have, without loss of generality, that $E(u_{i1}|u_{i2}) = 1$, in which case (6.36) equals one. However, this case is uninteresting to us here, because T_i is then independent of the event process and the distribution of $\{N_i(s),\ 0 \le s \le t\}$ does not depend on t. The special case where $u_{i1} = u_{i2}$ is no simpler; in that case $E(u_{i1}|u_{i2}) = u_{i2}$ in (6.36), but $E(u_{i1}|T_i = t)$ is still a complicated function of covariates. Even in the case where there are no covariates, we cannot say much without making further assumptions. For example, if $E(u_{i1}|u_{i2})$ is an increasing function of u_{i2} it might be expected that (6.35) would be an increasing function of t, but this appears to require additional conditions.

We next consider the use of marginal methods that provide direct estimates of features of common interest.

6.6.3 Robust Methods for Marginal Features

Here we consider models based on the marginal distribution of the termination time, and the rate function for the recurrent events, conditional on the termination time. For now, we consider covariates to be fixed. The termination time distribution function given covariate vector x_i is

$$\Pr(T_i \le t|x_i) = F_i(t) = F(t|x_i)\,, \tag{6.37}$$

and may be specified and analyzed in ways that are well known in survival analysis.

Consider next the rate of occurrence functions and associated mean functions that condition on survival to a specified time, given by

$$\rho_i(s;t)ds = E\left\{dN_i(s)|T_i \ge t,\ x_i\right\} \qquad s \le t\,, \tag{6.38}$$

and

$$\mu_i(s;t) = \int_0^s \rho_i(u;t)du = E\{N_i(s)|T_i \ge t, x_i\}\,, \tag{6.39}$$

respectively, $i = 1, 2, \ldots, m$. The rate function (6.38) and mean function (6.39) are easily interpreted and directly related to observed data. For example, a nonparametric estimate of (6.39) is obtained in the case where there are no covariates by taking all subjects with failure time known to be at least t, and computing the average number of events up to each time $s \le t$.

A special marginal rate function is given by

$$\rho_i^*(s)ds = \rho_i(s;s)ds = E\{dN_i(s)|T_i \ge s\}\,. \tag{6.40}$$

This can be estimated nonparametrically from the data, but integrating $\rho_i^*(s)$ does not in general yield anything interpretable unless the recurrent events are independent of termination time. The quantity

$$E\{N_i(t)\} = \int_0^t \rho_i^*(u) S_i(u) du\,, \tag{6.41}$$

where $S_i(u) = 1 - F_i(u)$, is always interpretable, however; it is the marginal expected number of recurrent events up to time t per subject, incorporating the fact that subjects who fail cannot experience any further events.

In the absence of covariates, if there are data on m independent individuals we can estimate (6.41) as follows, assuming that the censoring time C_i is independent of the recurrent event process and T_i. The rate function increments $\rho^*(u)du = \rho_i^*(u)du$ are estimated by the increments of the Nelson–Aalen estimate (3.17), given here by

$$\widehat{\rho}^*(u)du = \frac{\sum_{i=1}^m I(C_i \geq u) dN_i(u)}{\sum_{i=1}^m I(C_i \geq u, T_i \geq u)}\,, \tag{6.42}$$

assuming the denominator is positive. The common survivor function $S_i(u) = S(u)$ can be estimated with an ordinary Kaplan–Meier estimate, and then the two estimates can be plugged into (6.41). The resulting estimate $\widehat{\mu}(t)$ takes the form of a sum for any $t > 0$, because $\widehat{\rho}^*(u)du$ equals zero except at times u at which an event is observed. Variance estimates or pointwise confidence intervals for $\widehat{\mu}(t)$ are most simply obtained using the nonparametric bootstrap. An alternative is to use a variance estimate given by Ghosh and Lin (2000).

A concern with marginal estimates of this type is that they become invalid if censoring times are not independent of the event processes. For example, if the censoring process were dependent on the event history, or if there were unmeasured covariates that affected censoring as well as the recurrent events or termination time, then (6.42) should not be used. In Section 7.2, we discuss how to use weights based on a model for the censoring process, in order to obtain usable estimates. We also discuss, in the following Section 6.6.4, an approach that provides estimates of marginal features when censoring may depend on past event history in a specific way.

Regression models for the marginal distribution of T_i and for conditional rate functions can also be formulated. For example, if we consider the model

$$\rho_i^*(t) = \rho_0^*(t) \exp(x_i'\beta) \tag{6.43}$$

then the estimating functions from a Markov intensity-based model in which (6.30) is given by (6.43) are still unbiased and valid for estimation, provided that given x_i, C_i is independent of T_i and the recurrent event process. The estimating functions in question are equivalent to (3.4) and (3.5) in Section 3.2. If $\rho_0^*(t)$ is specified in terms of a parameter α, then they are

$$U_\beta(\beta, \alpha) = \sum_{i=1}^m x_i \int_0^\infty \bar{Y}_i(t) \{dN_i(t) - \rho_i^*(t;\alpha)dt\} \tag{6.44}$$

$$U_\alpha(\beta, \alpha) = \sum_{i=1}^m \int_0^\infty \bar{Y}_i(t) \frac{\partial \log \rho_0^*(t;\alpha)}{\partial \alpha} \{dN_i(t) - \rho_i^*(t;\alpha)dt\}\,, \tag{6.45}$$

where $\bar{Y}_i(t) = I(t \leq T_i, t \leq C_i)$. Setting $U_\beta(\beta, \alpha)$ and $U_\alpha(\beta, \alpha)$ equal to zero and solving provides consistent estimates of β and α. The semiparametric model in which $\rho_0^*(t)$ is an arbitrary positive-valued function can also be handled, as described in Section 3.6.3. In particular, as shown there, the estimates $\widehat{\beta}$ and the estimate of $R_0^*(t) = \int_0^t \rho_0^*(s)ds$ are those from the standard Andersen–Gill model, as given by the function `coxph` in S-PLUS or R, for example. The robust variance estimates described in Section 3.6.3 are also given by the software when the `cluster` option is used.

In the present setting, the generalized Nelson–Aalen estimate $\widehat{R}_0^*(t)$, given by (3.24), is not meaningful. As described above, a meaningful quantity is the expected number of events experienced up to time t, allowing that an individual has no more events once they experience the terminating event. This is given by

$$\widehat{\mu}(t; x) = \widehat{E}\{N_i(t)|x_i = x\} = \int_0^t \widehat{S}(u|x)d\widehat{R}^*(u|x)\,, \tag{6.46}$$

where $d\widehat{R}^*(u|x) = \exp(x'\widehat{\beta})d\widehat{R}_0^*(u)$ and where $\widehat{S}(u|x)$ is the estimated survivor function in a model for the termination time T_i, given covariates x. A variety of models, including proportional hazards, can be handled with standard survival analysis software. We note that external time-varying covariates $x(t)$ can also be accommodated with the approach just described. Bootstrap variance estimates or confidence limits can be obtained for $\mu(t; x)$.

Regression models for the conditional rate and mean functions (6.38) and (6.39) could also be considered, as could models for rate functions that condition on the value $T_i = t$. Models for $\mu_i(s; t) = E\{N_i(s)|T_i = t, x_i\}$ for $0 \leq s \leq t$ can be valuable, and offer a representation of the dependence of the recurrent event process on the termination time. However, these require assumptions that are not easily checked, and are harder to fit; Cook and Lawless (1997) provide some discussion. Moreover, most models do not give easily interpreted covariate effects, as illustrated by (6.35). Ghosh and Lin (2002) consider models with $E\{N(t)|x\} = \mu_0(t)\exp(x'\beta)$. These require the use of weighted estimating functions, which are considered in Section 7.2 for dealing with dependent censoring or loss to followup. Such models on the surface provide a more easily interpreted estimate of the effects of covariates on $\mu(t; x) = E\{N(t)|x\}$ than (6.46). However, the models are purely descriptive; if a covariate value is associated with a large value of $\mu(t; x)$; for example, it will not be clear whether this is due to increased survival time, an increased event rate when alive, or some combination of the two. The models that lead to (6.46) provide insight into both of these factors.

6.6.4 Partially Conditional Methods

Another approach to the joint analysis of recurrent and terminal events is based on the multistate formulation of Section 5.3. Figure 6.2 displays a multistate diagram in which the states indicate the status of an individual at any

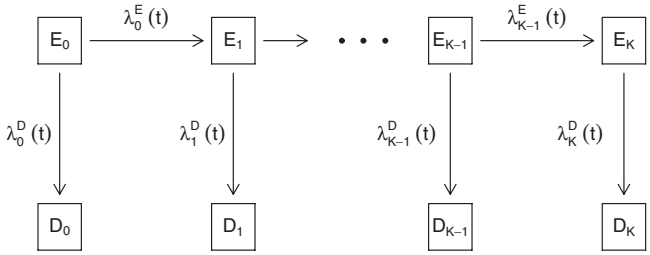


Fig. 6.2. A multistate diagram for a Markov model for recurrent events and a terminal event.

time and the arrows indicate the possible directions of transitions. Specifically, an individual is in state E_k if she is alive and has experienced k events, whereas an individual enters a "D" state when she dies. The subscript on the "D" indicates the cumulative number of events that individual experiences before she dies and so a subject absorbed into state D_k would have had k events prior to death. Although in principle there may not be a maximum number of events, in any specific application it is convenient to assume there is a maximum, denoted K, so that the process generates no further events after the Kth. For a given dataset we can without loss of generality set K equal to the maximum number of events observed for any single individual, because estimated transition intensities beyond this are all zero.

We assume that all individuals begin in state E_0 at $t = 0$, that states $E_0, \ldots, E_K$ are transient, and that states D_k, $k = 0, 1, \ldots, K$ are absorbing. At times it will be convenient to number $E_0, \ldots, E_K$ as states $0, \ldots, K$ and states $D_0, \ldots, D_K$ as states $K + 1, \ldots, 2K + 1$. An individual's life history tracks movement through these states.

Let $Z(t)$ represent the state occupied by an individual at time t and $H(t) = \{Z(s) : 0 \leq s < t\}$ denote the process history over $[0, t)$. The transition intensity functions for events are defined as

$$\lambda_k^E(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(Z((t+\Delta t)^-) = E_{k+1}|Z(t^-) = E_k, H(t))}{\Delta t},$$

$k = 0, \ldots, K - 1$, and for death as

$$\lambda_k^D(t|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\mathrm{P}(Z((t+\Delta t)^-) = D_k|Z(t^-) = E_k, H(t))}{\Delta t},$$

$k = 0, \ldots, K$. A model must be specified to characterize the way information in $H(t)$ influences the instantaneous probability of transition at time $t > 0$. Markov models provide a natural framework for many settings. As discussed

in Section 5.3, for Markov models the transition intensities depend only on the current state occupied and the time since the origin of the process and may therefore be viewed as transition rates. Under a Markov model we have

$$\lambda_k^E(t|H(t)) = \lambda_k^E(t) \qquad k = 0, \ldots, K-1 \tag{6.47}$$

and

$$\lambda_k^D(t|H(t)) = \lambda_k^D(t) \qquad k = 0, \ldots, K \tag{6.48}$$

as shown in Figure 6.2.

When we wish to consider general transition intensities we use $q_{k\ell}(t)$ to denote the transition rate from state k to state ℓ, $\ell \neq k$, and let $Q_{k\ell}(t) = \int_0^t q_{k\ell}(u)du$ denote the corresponding cumulative transition rate. The transition probability functions are, for $s \geq t$,

$$p_{k\ell}(s,t) = \Pr(Z(t) = \ell | Z(s) = k)\,, \tag{6.49}$$

which we write in matrix form as $\mathcal{P}(s,t) = [p_{k\ell}(s,t)]$. The transition probability functions and transition intensities are again linked via the product integral, as given earlier in (5.17),

$$\mathcal{P}(s,t) = \prod_{(s,t]} (\mathcal{I} + \mathcal{Q}(u)du)\,, \tag{6.50}$$

where here $\mathcal{I}$ is a $(2K+2) \times (2K+2)$ identity matrix and $\mathcal{Q}(t)$ is the $(2K+2) \times (2K+2)$ matrix of transition rates with entries $q_{kk}(t) = -\lambda_k^E(t) - \lambda_k^D(t)$, $k = 0, 1, \ldots, K-1$; $q_{KK}(t) = -\lambda_k^D(t)$; $q_{k,k+1}(t) = \lambda_k^E(t)$, $k = 0, 1, \ldots, K-1$; $q_{k,(K+1)+k}(t) = \lambda_k^D(t)$, $k = 0, 1, \ldots, K$ and zero elsewhere. Assuming subjects are under observation at the start of the process (i.e. at $t = 0$), the state occupancy probabilities $p_{0k}(t)$ at time t are given by the top row of $\mathcal{P}(0,t)$. Note that they give the probabilities $\Pr(N(t) = k, T > t) = \Pr(Z(t) = E_k)$ and $\Pr(T \leq t, N(t) = k) = \Pr(Z(t) = D_k)$ for $k = 0, 1, \ldots, K$.

Multistate models of this sort can be used to address many questions. For example, in some applications it is common to focus on "event-free survival" time, in which case interest lies in $\Pr(Z(t) = 0 | Z(0) = 0) = p_{00}(0,t)$. Another feature of interest is the cumulative mean function (6.41). In the present framework, under the assumption that the probability of more than K events is zero, $\mu(t) = E\{N_i(t)\}$ is given by

$$\mu(t) = \sum_{k=1}^{K} k \left[\Pr(Z(t) = E_k | Z(0) = 0) + \Pr(Z(t) = D_k | Z(0) = 0)\right]. \tag{6.51}$$

One may also consider features such as the expected number of events over an individual's lifetime. The *cumulative lifetime mean* $E\{N(T)\}$ is

$$\mu(\infty) = \lim_{t \to \infty} \mu(t) = \sum_{k=1}^{K} k \cdot \Pr\left(Z(\infty) = D_k | Z(0) = 0\right). \tag{6.52}$$

As described in Section 5.3.2, nonparametric estimation is straightforward when there are no covariates and the transition rates in (6.47) and (6.48) are unrelated. Estimates of (6.51), for example, may be obtained by replacing state occupancy probabilities $p_{0k}(t)$ in the right-hand side with their corresponding estimates, as given by (5.19) and (5.20) of Section 5.3.2. This is adapted to the present setting as follows.

Consider a dataset of m independent subjects in which subject i is followed over $[0, C_i]$ where C_i is a censoring time independent of $\{Z_i(t), 0 \le t\}$. Let $Y_{ik}(t) = I(Z_i(t^-) = k)$, $k = 0, 1, \ldots, K-1$ indicate whether subject i is at risk of a transition out of transient state k at time u, and $Y_i(t) = I(t \le C_i)$ indicate whether subject i is under observation at time u. Let $dN_{ik\ell}(t) = 1$ if subject i makes a transition from state k to ℓ at time t, and $dN_{ik\ell}(t) = 0$ otherwise. Finally, let $N_{ik\ell}(t) = \int_0^t dN_{ik\ell}(u)$ indicate whether a transition from k to ℓ occurred for subject i over $[0, t]$. For the corresponding observable quantities we let $\bar{Y}_{ik}(t) = Y_i(t)Y_{ik}(t)$, $d\bar{N}_{ik\ell}(t) = \bar{Y}_{ik}(t)dN_{ik\ell}(t)$, and $\bar{N}_{ik\ell}(t) = \int_0^t \bar{Y}_{ik}(u)dN_{ik\ell}(u)$.

The Nelson–Aalen estimate of the $k \to \ell$ cumulative rate function is

$$\widehat{Q}_{k\ell}(t) = \int_0^t \frac{I(\bar{Y}_{\cdot k}(u) > 0)d\bar{N}_{\cdot k\ell}(u)}{\bar{Y}_{\cdot k}(u)}\,, \tag{6.53}$$

where $d\bar{N}_{\cdot k\ell}(u) = \sum_{i=1}^m d\bar{N}_{ik\ell}(u)$, $\bar{Y}_{\cdot k}(u) = \sum_{i=1}^m \bar{Y}_{ik}(u)$, and $0/0$ is defined to be 0. The Aalen–Johansen estimate of the transition probability matrix is then

$$\widehat{\mathcal{P}}(s,t) = \prod_{(s,t]} (I + d\widehat{Q}(u))\,, \tag{6.54}$$

where $dQ(u) = \mathcal{Q}(u)du$ in (6.50) and $d\widehat{Q}_{k\ell}(u) = I(\bar{Y}_{\cdot k}(u) > 0)d\bar{N}_{\cdot k\ell}(u)/\bar{Y}_{\cdot k}(u)$. The functions $p_{0k}(t)$ are then estimated by the first row of $\widehat{\mathcal{P}}(0,t)$. These estimates are, of course, valid when the present Markov model is suitable, but they are also valid more generally if censoring is independent of the event process, as described in Section 5.3. It should be noted that caveats stated in Section 5.3, regarding the inestimability of $dQ_{k\ell}(u)$ over intervals where $\bar{Y}_{\cdot k}(u) = 0$, also apply here.

6.7 Applications and Illustrations

6.7.1 Cerebrospinal Fluid Shunt Failures

In Section 5.4.1 we considered data on recurrent failures of shunts that drain excess cerebrospinal fluid for 839 children with hydrocephalus. Failures were classified by cause into three types: Obstruction, Infection, and Other, which account for about 70%, 15%, and 15% of failures, respectively. Table 5.4 showed the distribution of number of failures per child, but we must bear

in mind that followup in these observational data ranges from about one to eleven years across individuals.

The analysis in Section 5.4.1 considered all causes of failures combined, and showed the importance of covariates associated with etiology, shunt type, other surgery, and age at surgery in an examination of successive gap times between failures. It is natural to focus on the gap times, because failure of a shunt results in its full or partial replacement. We consider in this section the analysis of the three types of failure, along with death as a terminating event. For shunt failures, semi-Markov models of the form $D_i(t)$ times (6.9) are used, where $D_i(t) = I(T_i \geq t)$ indicates that individual i is alive at time t^-, $B_{ij}(t) = B_i(t)$ is the time since the last previous failure of any type, and $j = 1, 2, 3$ denotes the Obstruction, Infection, and Other types of failure, respectively. If W_{ik} denotes the kth gap time and C_{ik} denotes the corresponding cause of failure for individual i, then the models considered can equivalently be expressed in terms of the cause-specific intensity functions $\lambda_{ijk}(w)$ given by

$$\lim_{\Delta w \downarrow 0} \left\{ \frac{\Pr\left(W_{ik} < w + \Delta w, C_{ik} = j | W_{ik} \geq w, z_{ik}, T_i > t_{i,k-1} + w\right)}{\Delta w} \right\}, \quad (6.55)$$

where $t_{i0} = 0$, $t_{ik} = w_{i1} + \cdots + w_{ik}$ is the time of the kth failure, and z_{ik} is a covariate vector that may contain information on previous failure times and causes. Note that (6.55) is conditional on the individual being alive at the time in question. In addition, we consider models for the death intensity function of the form

$$\lambda_{iD}(t) = \lim_{\Delta t \downarrow 0} \left\{ \frac{\Pr\left(T_i < t + \Delta t | T_i \geq t, z_i(t)\right)}{\Delta t} \right\}, \quad (6.56)$$

where $z_i(t)$ is a covariate vector that may include information on previous shunt failures.

We first show the results for the risk of death. Multiplicative Cox regression models for (6.56) that included baseline covariates for etiology, age at first surgery, and other factors were considered, along with time-varying covariates that indicated whether there had been prior shunt failures. The survival time T_i is defined here as the time from insertion of the first shunt (i.e. first surgery) to death, because it seems most natural to assess mortality in terms of the (approximate) time since the condition necessitating the shunts arose. Table 6.1 shows results for the model deemed most satisfactory; it includes a binary covariate $z(t)$ ("Prior failure") that indicates whether (yes = 1, no = 0) a shunt failure has occurred by time t.

Table 6.1 indicates a much higher risk of death for persons in the Tumor etiology group, and a significantly lower risk of death for persons in the IVH and MMC groups. Of the 121 deaths, 62 were in the Tumor group; this group comprises 23% of the 839 individuals. The existence of a prior failure is strongly associated with an increased risk of death. The elevation in risk

Table 6.1. Multiplicative intensity model for death in persons with shunts.

Covariate	EST.	S.E.	RR	$p-$value[†]
Age at 1st surgery (years)				
< 0	0.540	0.407	1.716	0.180
0–1	0.206	0.269	1.229	0.440
> 1	–	–	–	–
Etiology				
Adsten	-0.550	0.528	0.577	0.300
IVH	-1.747	0.595	0.174	0.003
Men	-1.163	0.761	0.312	0.130
MMC	-0.840	0.384	0.432	0.029
Trauma	0.122	0.655	1.129	0.850
Tumor	1.111	0.363	3.037	0.002
Other	-0.130	0.371	0.878	0.730
Con	–	–	–	–
Prior failure	0.697	0.197	2.007	0.0004

[†] $p-$value is based on $2\Pr(Z > |\widehat{\beta}|/\text{s.e.}(\widehat{\beta}))$, where Z is standard normal.

is, from other models fitted, similar for persons in the Tumor and non-Tumor etiology groups. There is also mild but not highly significant evidence that obstruction failures elevate the risk of death more than infection or other failures. Baseline age is not significant in Table 6.1, but it should be noted that it is strongly associated with shunt failure, and hence with the covariate Prior failure. Finally, an examination of the times of death and the baseline cumulative intensity function for the fitted model indicates that the risk of death is highest soon after first surgery but does persist later on, especially for persons in the Tumor etiology group.

For shunt failures, we consider models (6.55) of the multiplicative form

$$\lambda_{ijk}(w) = \lambda_{0jk}(w) \exp\left(z'_{ik}\beta_{jk}\right) , \tag{6.57}$$

where $j = 1, 2, 3$ denotes the type of failure, $k = 1, 2, \ldots$ denotes the gap, and z_{ik} is a vector of covariates which for $k \geq 2$ may include information on previous failure times. The stratified form of (6.57) is consistent with our all-causes analysis in Section 5.4.1, and it is sensible to continue to allow the effects of covariates to change with successive shunts, as done previously.

After exploration of a variety of factors, we settled on the models shown in Tables 6.2 to 6.4, which give results for Obstruction, Infection, and Other Failures, respectively. Models for first, second, and third shunt failures are shown. The covariates include age at current surgery (i.e. at the time of shunt insertion), etiology, shunt type, concurrent other surgery and, for second and third shunts, the time they started to operate (i.e. the start time of gaps

2 and 3, respectively). Very few factors are significant for second and third shunts, but we retain all covariates for comparison across all three shunts. For Obstruction, comprising approximately 70% of the failures, there was an indication that effects of age at current surgery and shunt type might be time-varying, and we allow for this in the first two shunts. There was also an indication that second Obstruction failures might be related to the cause of the first shunt failure, and we incorporate this as well.

Table 6.2. Estimates from regression models for obstruction shunt failures.

Covariate	SHUNT 1 EST.	S.E.	SHUNT 2 EST.	S.E.	SHUNT 3 EST.	S.E.
Age at current surgery						
< 0 (year 1)	1.08	0.26	0.67	0.44	0.43	0.79
< 0 (year 2)	1.57	0.43	1.22	1.07	0.43	0.79
< 0 ($>$ year 2)	-0.13	0.56	1.43	1.07	0.43	0.79
0–1	0.74	0.20	0.12	0.25	0.38	0.29
> 1	–	–	–	–	–	–
Etiology						
Adsten	0.55	0.30	-0.07	0.44	0.06	0.63
IVH	0.68	0.25	0.49	0.36	0.58	0.50
Men	0.40	0.34	0.56	0.42	0.34	0.55
MMC	0.59	0.23	0.23	0.33	0.19	0.50
Trauma	0.93	0.44	0.61	0.54	-0.64	1.10
Tumor	0.65	0.28	0.05	0.39	0.05	0.53
Other	0.39	0.25	0.25	0.35	-0.11	0.54
Con	–	–	–	–	–	–
VP shunt type						
Year 1	-0.79	0.21	- 0.13	0.30	0.19	0.39
Year 2	-0.50	0.62	0.03	0.43	0.19	0.39
$>$ Year 2	0.39	0.60	- 0.38	0.49	0.19	0.39
Concurrent surgery	0.28	0.20	0.64	0.31	0.33	0.44
Time of shunt surgery ($\div$ 100)	–	–	0.06	0.03	-0.005	0.02
Cause of first failure						
Infection	–	–	0.48	0.21	–	–
Obstruction	–	–	0.86	0.28	–	–
Other	–	–	–	–	–	–

Table 6.2 shows results for Obstruction failures; we simply comment on a few points. First, age at current surgery is important for first shunts, with excess risks both for infants born prematurely and for those under one year

of age. The high excess risk for premature infants persists only for about two years; this may also be true for infants under one year of age, but a test for a time-varying covariate effect does not give significant evidence. Shunt type has a similar time-dependent effect for first shunts. For second and third shunts, there is no strong indication of association with the time of previous failure, although we should note that this is ambiguous because of the relationship between time of first failure and baseline covariates. Interestingly, the risk of an obstruction failure following a first failure due to obstruction is significantly lower than if the first failure were due to infection or other causes.

Table 6.3. Estimates from regression models for infection shunt failures.

Covariate	SHUNT 1 EST.	SHUNT 1 S.E.	SHUNT 2 EST.	SHUNT 2 S.E.	SHUNT 3 EST.	SHUNT 3 S.E.
Age at current surgery						
< 0	1.22	0.48	-0.29	0.96	2.19	1.51
0–1	0.70	0.42	0.15	0.56	0.49	0.84
> 1	–	–	–	–	–	–
Etiology						
Adsten	1.7	0.61	-0.24	0.92	-1.41	1.37
IVH	1.03	0.57	0.60	0.69	-1.76	1.01
Men	0.94	0.71	-0.03	0.92	-8.21	31.6
MMC	1.09	0.54	-0.65	0.71	-1.61	1.00
Trauma	2.10	0.81	-5.02	14.0	1.18	1.00
Tumor	1.27	0.61	0.70	0.77	-0.69	0.92
Other	0.88	0.58	-0.50	0.82	-1.68	1.21
Con	–	–	–	–	–	–
VP shunt type (year 1)	0.66	0.60	0.72	0.75	6.85	20.8
Concurrent surgery	0.57	0.33	-0.33	0.76	1.37	0.77
Time of shunt surgery ($\div$ 100)	–	–	-0.14	0.08	-0.03	0.07

Tables 6.3 and 6.4 show analogous results for Infection and Other failures. There are significant effects due to age at current surgery and etiology for first shunt failures due to infection, and a few other miscellaneous significant effects for failures due to other causes, and for second and third failures due to infection. These effects are not consistent across first, second, and third shunts, although an effect due to shunt type appears for second and third failures for the Other cause category.

Note that each shunt (first, second, third, etc.) is at risk for the three causes of failure, and an alternative way to present the regression results would be to show the three causes together for each shunt (first, second, third). It should also be noted that Infection failures virtually all occur within a year of shunt

Table 6.4. Estimates from regression models for other shunt failures.

Covariate	SHUNT 1 EST.	SHUNT 1 S.E.	SHUNT 2 EST.	SHUNT 2 S.E.	SHUNT 3 EST.	SHUNT 3 S.E.
Age at current surgery						
< 0	0.75	0.54	1.07	0.89	2.91	1.18
0–1	0.18	0.38	0.30	0.57	0.33	0.77
> 1	–	–	–	–	–	–
Etiology						
Adsten	0.04	0.71	-5.50	18.3	-0.39	1.52
IVH	- 0.35	0.64	1.72	1.09	0.07	1.21
Men	1.50	0.53	2.22	1.14	-5.60	17.1
MMC	-0.60	0.57	-0.42	1.24	0.02	1.12
Trauma	0.18	1.11	-5.56	32.0	-5.90	28.7
Tumor	1.01	0.54	1.73	1.11	0.63	1.10
Other	0.02	0.55	0.71	1.16	-0.96	1.44
Con	–	–	–	–	–	–
Shunt type (year 1)	-0.12	0.39	-1.20	0.52	-1.44	0.61
Concurrent surgery	0.88	0.37	-0.36	0.79	0.78	0.79
Time of shunt surgery ($\div$ 100)	–	–	-0.003	0.05	0.04	0.04

insertion, whereas the risk for the other two causes of failure is highest in the first year after insertion, but persists over time. There is also a suggestion, as in the earlier all-causes analysis of Section 5.4.1, that the baseline risk for failure of second and subsequent shunts for each of the three causes may be higher than for first shunts. This may reflect heterogeneity of risk across the individuals, although the time of prior failures is not significant in Tables 6.3 and 6.4.

6.7.2 Exacerbations in Patients with Chronic Bronchitis

We consider the analysis of data from a multicenter randomized trial involving patients with chronic bronchitis, where the disease process consists of the development and resolution of acute exacerbations of chronic bronchitis (AECB) over time (see Section 1.5.3). To have been eligible for the study, patients must have been 18 years or older, diagnosed with chronic bronchitis, able to maintain a daily diary, able to understand and complete detailed health status questionnaires, and currently experiencing an acute exacerbation. Upon study entry patients were randomized to receive either Ciprofloxacin or standard care for the treatment of exacerbations for a period of one year of followup. Kaplan–Meier estimates for the time from study entry to resolution of the initial exacerbation are given for the Ciprofloxacin and control groups in Figure

6.3; a log-rank test gives $p = 0.062$, indicating some evidence of more rapid symptom resolution in the Ciprofloxacin arm. This is a simple marginal comparison of treatment groups on which one can base causal inferences about treatment effects. Interpretation of the treatment effect is somewhat ambiguous, however, because there was considerable variation in the lengths of time from the start of the initial exacerbation to randomization.

Figure 1.10, which portrays the event histories for a number of subjects from this study, shows that analysis based on the first exacerbation ignores a considerable amount of the information collected. Because of the relapsing and remitting nature of chronic bronchitis, it provides an example of the two-state process considered in Section 6.5, and here we consider analysis of the full data with a view to identifying and characterizing the effect of important prognostic variables. We define an individual to be in state 1 when she is experiencing an exacerbation, and in state 2 when she is not.

Patients were required to visit the participating clinic when they perceived that a new exacerbation was beginning, or when they determined that an exacerbation was resolved. Patient followup was to continue for 365 days, but early termination could occur if a subject refused to complete her symptom diary or to return for further followup visits, or if she died. There were 115 eligible patients randomized to take Ciprofloxacin and 107 randomized to standard care. Patients were accrued from November 1993 to June 1994 and the average duration of followup was 357 and 350 days in the Ciprofloxacin and standard care groups, respectively. Days are used as the time scale in the following discussion.

Episodic conditions of this sort often exhibit both semi-Markov and Markov behavior. For example, the distribution of sojourns in the exacerbation state is often primarily governed by the time since the exacerbation began. Markov or calendar time trends may be present if the sojourn distributions depend on the history of the underlying chronic condition; for example, patients who have had the disease longer may tend to have longer exacerbations when they occur. For the onset of new exacerbations, the time since disease onset may be the most natural time scale because the exacerbations often arise as a consequence of a deterioration in lung function due to the underlying disease. In addition, many chronic respiratory disease processes also exhibit seasonal trends. Gap time analyses remain attractive, however, inasmuch as an aim of treatment is to delay the onset of exacerbations and hence extend the interexacerbation periods.

We let t denote the calendar time scale for a subject, originating at her time of recruitment into the study. We refer to this as the study time. Let $d_i(0)$ denote the duration of the underlying disease (time since diagnosis with chronic bronchitis) for subject i upon entry, and $d_i(t)$ the duration of disease at study time t. In this dataset the time from the diagnosis of chronic bronchitis to study entry ranged from 1 to 54 years, and here we partition the time-scale into five-year intervals and construct time-varying covariates indicating in which interval the duration of a patient's chronic bronchitis

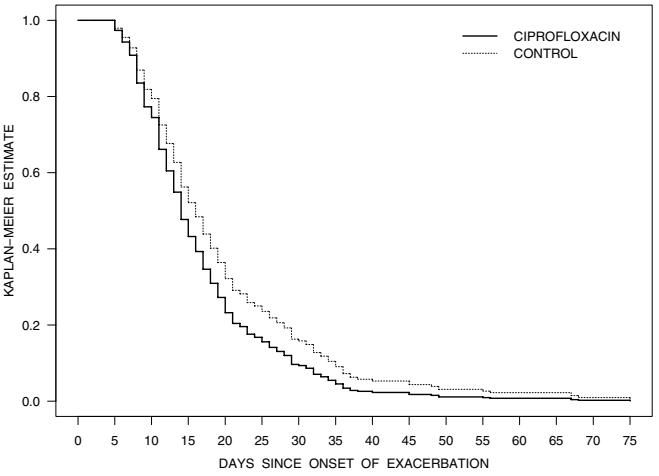


Fig. 6.3. Kaplan–Meier plots of the distribution of the time from study entry to resolution of the first exacerbation.

is. So if $b_k = 5k$, $k = 0, \ldots, 11$, we define $v_{ik}(t) = I(b_{k-1} < d_i(t) \leq b_k)$, $k = 1, 2, \ldots, 11$, and $v_i(t) = (v_{i2}(t), \ldots, v_{i11}(t))'$. Seasonal effects are modeled by partitioning the calendar year into four quarters: January to March, April to June, July to September, and October to December. We define $s_{ik}(t) = 1$ if study time t is in the kth quarter of the calendar year, $k = 2, 3, 4$. Then we let $s_i(t) = (s_{i2}(t), s_{i3}(t), s_{i4}(t))'$, and $z_{ij}(t) = (s_i'(t), v_i'(t), x_{ij}'(t))'$, where $x_{ij}(t)$ could contain time-dependent covariates, but here contains only fixed covariates including treatment (Ciprofloxacin versus standard care), a variable (symptoms) recording the number of days that symptoms had been present at the time of entry into the study, sex (female versus male), and clinical designation of the severity of the chronic bronchitis (severe versus not severe).

To be eligible for inclusion in the study, subjects were required to be experiencing an exacerbation at the time of entry (i.e. they were required to be in state 1 at $t = 0$). Thus the time of randomization did not coincide with a transition time and patients are picked up in the middle of a sojourn in state 1. Recalling the discussion in Section 4.5.2, we note that in this study the time $t_{i0} < 0$ of the start of the initial exacerbation (represented by the variable "symptoms") is known for each subject. However, treatment assignment was made at the time of randomization ($t = 0$). We partially address this by modeling separately the distribution of the time from study entry to resolution of the first exacerbation at randomization. No stratification is made for subsequent sojourns in states 1 or 2. We assume that seasonal effects are the same for the first and subsequent exacerbations but estimate

separate effects of all other covariates. The duration that symptoms were present for the exacerbation experienced at the time of study entry is included as a covariate to address the selection effects arising from the inclusion criteria. We comment more on this in Section 7.3.1.

Table 6.5. Estimates from semiparametric random effect models with independent gamma frailties, for the bronchitis data.

Covariate	AECB to AECB-Free EST.	S.E.	$p-$value	AECB-Free to AECB EST.	S.E.	$p-$value
	First Observed Duration					
Treatment	0.513	0.159	0.001	–	–	–
Sex	-0.199	0.160	0.212	–	–	–
Severity	-0.282	0.248	0.256	–	–	–
Symptons	-0.109	0.016	< 0.001	–	–	–
	Second and Subsequent Durations					
Treatment	0.060	0.131	0.644	-0.037	0.124	0.766
Sex	-0.089	0.135	0.508	0.236	0.127	0.063
Severity	0.006	0.179	0.974	0.562	0.172	0.001
Symptons	-0.011	0.010	0.261	–	–	–
Season						
Jan–March	–	–	–	–	–	–
April–Jun	0.291	0.120	0.016	-0.100	0.161	0.536
July–Sept	0.180	0.147	0.221	0.382	0.148	0.010
Oct–Dec	0.204	0.124	0.101	0.022	0.150	0.886
Disease duration	–	–	0.160	–	–	0.015
Variance (ϕ_j)	$\widehat{\phi}_1 = 0.1639$			$\widehat{\phi}_2 = 0.1977$		
Log-likelihood	-2845.418			-2151.162		

Table 6.5 contains the estimates obtained by fitting a semiparametric model with random effects as in (6.28) but with $h_{01}(t)$ and β_1 differing according to whether $N_{i1}(t) > 0$. The random effect u_{i1} and the effects of season and disease duration are assumed to be common for the cases $N_{i1}(t^-) = 0$ and $N_i(t^-) > 0$, but the other covariate effects are allowed to differ. Previous analyses (Ng and Cook, 1999b) suggest there is no need to model the association between the random effects and so we use independent gamma random effects in this model. This allows the models to be fitted using the function `coxph` in S-PLUS or R. The code for fitting this model is given in Appendix C. When controlling for sex, severity of chronic bronchitis, the duration of symptoms at randomization, and disease duration (i.e. $v_i(t)$), the treatment is associated with a substantial increase in the rate of resolution ($RR = 2.72, p < 0.001$) for the exacerbation experienced at randomization. Moreover, there is evidence that, when controlling for other factors, those subjects with a longer

duration of symptoms upon recruitment had a significantly longer time to resolution ($p < 0.001$) of the first exacerbation. There is no evidence that sex or severity of symptoms affect the intensity for symptom resolution for the first or subsequent exacerbations, and the treatment and symptom duration at randomization are not associated with the duration of second and subsequent exacerbations either. There is some evidence of seasonal variation in the recovery rates; when controlling for other covariates, it appears that exacerbations in the spring have somewhat shorter durations than those originating in the winter months. The $p-$value in Table 6.5 for disease duration is based on a test that the regression coefficients for $v_{i2}(t), \ldots, v_{i11}(t)$ are all zero; no significant effect of disease duration on the time to resolution of the exacerbations is indicated.

In terms of the prognostic variables for the onset of exacerbations, patients with a diagnosis of severe chronic bronchitis have much higher risk of new exacerbations ($p < 0.001$), but treatment, sex, and the duration of symptoms at randomization do not appear to be related to the onset of new exacerbations. We remark that randomization and treatment initiation took place in the middle of the initial exacerbation; in contrast, therapeutic treatment was initiated near the onset of subsequent exacerbations, when patients presented themselves to the clinic. Only a subset of patients was observed to experience a second exacerbation, however, so treatment comparisons on second and subsequent exacerbations are subgroup analyses in which causal inferences about treatment effects are difficult.

There is evidence that exacerbations occur more frequently in the summer months than the winter months ($p = 0.010$), and the disease duration also appears to be a significant factor ($p = 0.015$); the longer the history of chronic bronchitis the greater the risk of developing a new exacerbation. There is significant evidence of heterogeneity in both the duration of exacerbations and the times between exacerbations. Likelihood ratio tests of $H_0 : \phi_j = 0$ give $p < 0.0001$ and $p < 0.0001$ for $j = 1$ and $j = 2$, respectively.

The difficulties of dealing with the first exacerbation can be avoided if one only uses data following the resolution of the exacerbation at study entry because this resolution time is a stopping time. There is an obvious loss of information incurred by doing this but we considered this approach here as a type of sensitivity analysis. The results concerning the effects of covariates on onset and resolution of subsequent durations are broadly similar to those given in Table 6.5.

6.7.3 Skeletal Complications in Metastatic Cancer

Cancer patients with skeletal metastases are at risk of complications such as fractures or bone pain relating to decreased integrity of the bone. Bisphosphonate therapy is designed to strengthen bone and reduce the occurrence of complications arising from bone lesions. Mortality is quite high in these pop-

Table 6.6. Frequency distribution for total number of complications on study.

Number of Skeletal Events	Placebo No.	Placebo Percent	Pamidronate No.	Pamidronate Percent
0	69	35.4	99	53.5
1	41	21.0	39	21.1
2	34	17.4	17	9.2
3	18	9.2	13	7.0
4	12	6.2	10	5.4
5	7	3.6	3	1.6
6	5	2.6	1	0.5
7	4	2.1	1	0.5
≥ 8	5	2.6	2	1.1
Total	195		185	

ulations, however, and the occurrence of skeletal complications is of course terminated by death.

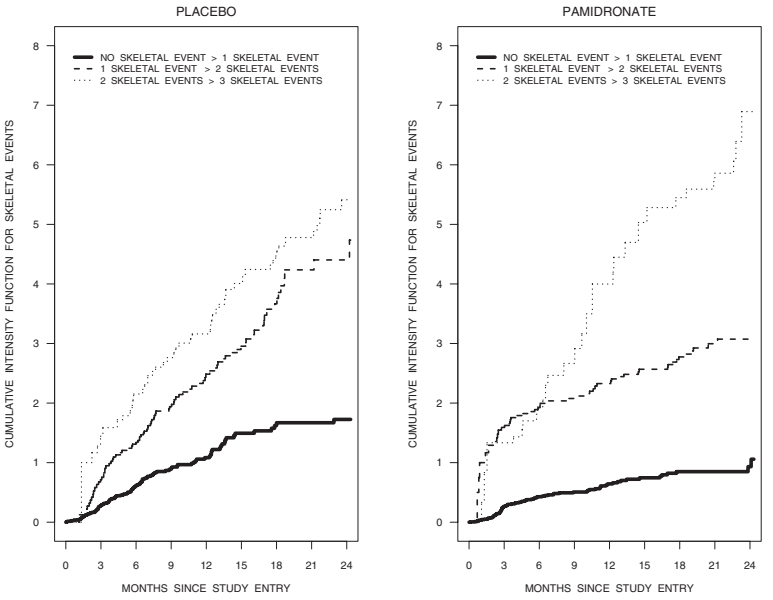


Fig. 6.4. Generalized Nelson–Aalen estimates of the cumulative intensity functions for first, second, and third skeletal events.

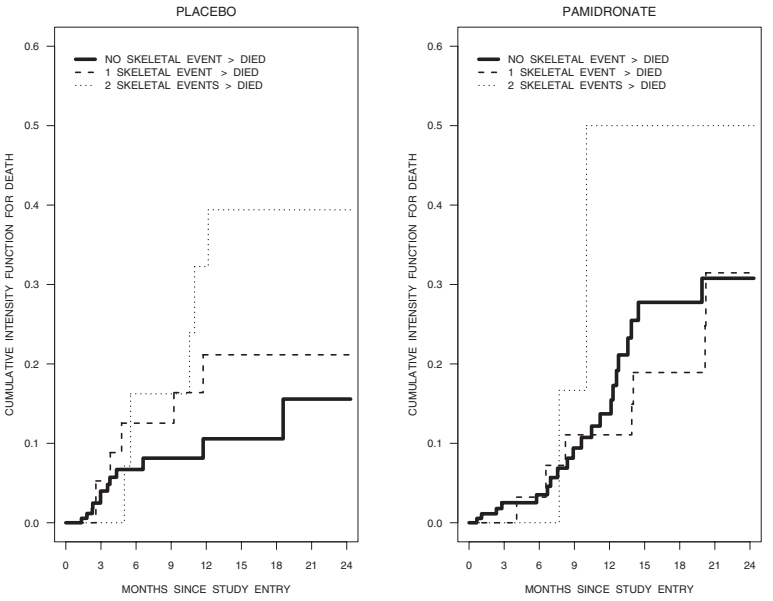


Fig. 6.5. Generalized Nelson–Aalen estimates of the cumulative intensity functions for death with no, one, or two skeletal events.

Here we consider the analysis of data arising from a randomized trial reported in Hortobagyi et al. (1996). Patients were accrued between January 1991 and March 1994 from 97 study sites in the United States, Canada, Australia, and New Zealand. Patients with stage IV breast cancer receiving cytotoxic chemotherapy with at least one predominantly lytic bone lesion greater than or equal to one centimeter in diameter were randomized to treatment groups. A total of 382 women were enrolled in the study with 185 and 197 randomized to receive an experimental treatment and the placebo control, respectively; two patients in the placebo arm were found to be ineligible and are excluded from the analyses that follow. Patients assigned to the experimental arm received 90 mg of a bisphosphonate, pamidronate disodium, via a two-hour infusion every four weeks, whereas patients randomized to the placebo control received dextrose infusions. Patients on a three-week chemotherapy regimen were permitted to receive the study drug every three weeks. At monthly visits patients were assessed and the occurrence of skeletal complications was recorded. The skeletal complications of interest include pathologic fractures, spinal cord compression with vertebral fracture, the need for surgery to treat or prevent fractures, and the need for radiation for the treatment of bone pain. After the planned followup of one year, the observation period was extended for an additional year; the results are published in Hortobagyi et al. (1998). Table 6.6 gives the frequency distribution of the

total number of skeletal complications from randomization to last contact, by treatment group. Tabulations of this sort are difficult to interpret when there may be differences in survival distributions, so this is only a crude data summary.

In this section we consider simple marginal analysis of the occurrence of skeletal complications, with death as a terminating event, performed separately for the two treatment groups. The purpose is to gain insight into the frequency of events through the multistate model depicted in Figure 6.2. We consider estimation based on Markov assumptions. Nelson–Aalen estimates (6.53) of the cumulative transition are given in Figure 6.4 for the occurrence of skeletal events in Figure 6.5 for death. These reveal increasing risk of new skeletal complications with the occurrence of each skeletal event, and a generally higher risk of death with additional skeletal-related events, although this pattern does not hold as clearly for treated patients; there are fewer patients at risk of death from the states representing two or three events.

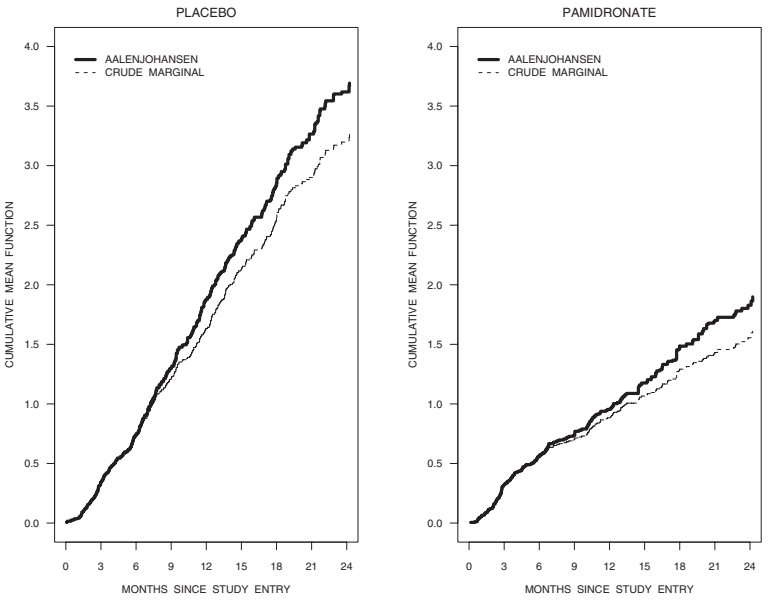


Fig. 6.6. Estimated mean functions for skeletal events based on Aalen–Johansen estimates (6.54) and "crude" marginal estimates (6.46).

As mentioned in Section 5.3, Aalen–Johansen estimates of the prevalence functions may also be obtained. We let $p_{0k}(0,t) = \Pr(Z(t) = E_k | Z(0) = E_0)$ and $p'_{0k}(0,t) = \Pr(Z(t) = D_k | Z(0) = E_0)$ and estimate these using (6.54). The estimated marginal mean function $\mu(t)$ can then be estimated using (6.51) and

this is plotted for placebo and treated patients in Figure 6.6. Also plotted in Figure 6.6 is an estimate of $\mu(t)$ based on the estimated conditional rate (6.42) and the Kaplan–Meier estimate of the survival function for each group (see (6.46) of Section 6.6.3 with no covariates). Finally, the usual marginal Kaplan–Meier and the Aalen–Johansen estimates of the survivor function (where the latter is given by $\sum_{k=0}^{\infty} \widehat{p}_{0k}(0,t)$) are displayed in Figure 6.7. The marginal and multistate estimates in both Figures 6.6 and 6.7 exhibit some differences. Such differences, and in particular the drop in the Aalen–Johansen estimate in the left panel of Figure 6.8, are influenced by there being few individuals at risk at later time points, and even fewer at risk from each transient state, but one would not expect these differences to be too large under completely independent censoring.

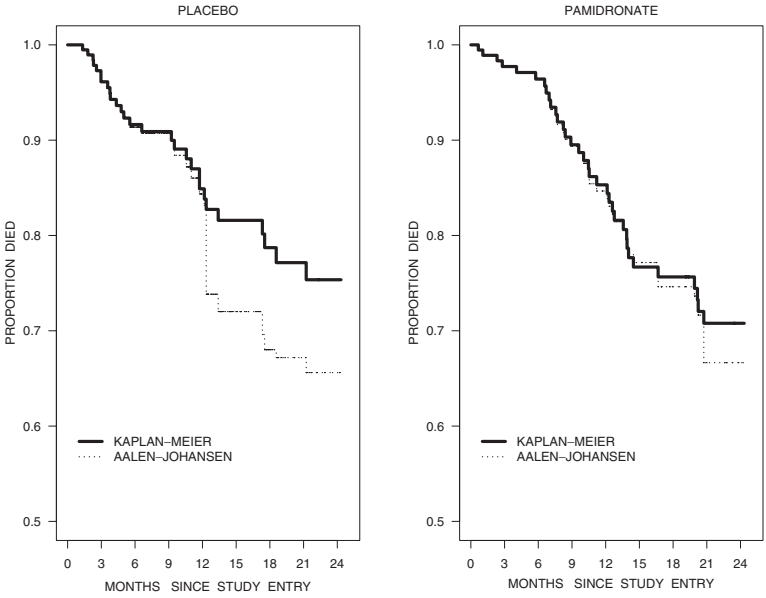


Fig. 6.7. Kaplan–Meier and Aalen–Johansen estimates of the survivor distribution for placebo and pamidronate groups.

To investigate this further, we plot the estimated cumulative intensities for censoring from each state (Figure 6.8). These plots show that subjects with a higher number of skeletal complications are at increased risk of withdrawal from the study. This is an example of state-dependent censoring which means that the rate of withdrawal or censoring from the study depends on the state occupied. Under these circumstances, the crude marginal estimates as in Figure 6.6 and the marginal Kaplan–Meier estimates as in Figure 6.7 are

inconsistent. Estimates of the survival distribution and mean function based on the multistate analysis are therefore preferable; they are consistent if the Markov model is appropriate.

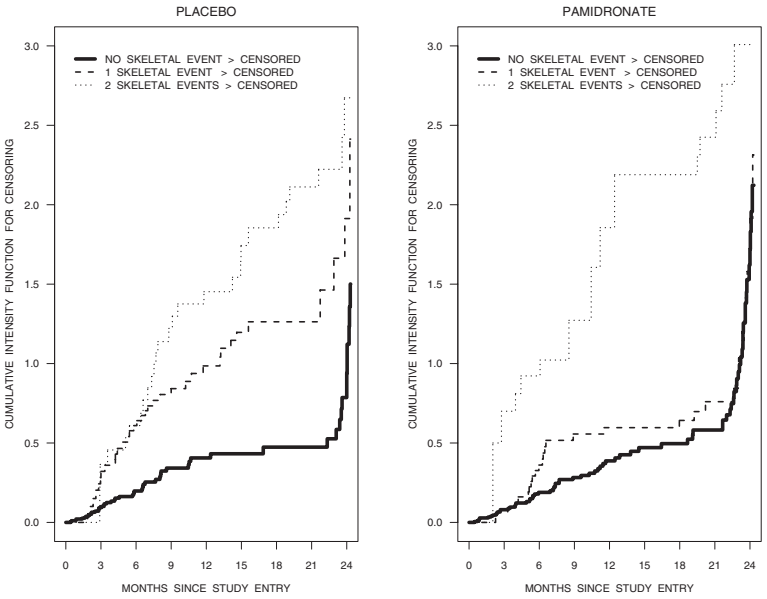


Fig. 6.8. Cumulative intensity functions for censoring with no, one, or two skeletal events.

An alternative approach to estimation in this context is to model marginal features of interest directly, but to introduce weights into the associated estimating functions which correct for effects of dependent censoring; see Section 4.4. We discuss the use of "inverse probability of censoring" weights in more detail in Section 7.2.

6.7.4 Relationships Between Skeletal Complications

In studies like the one in the preceding section, we may also be interested in characterizing the risk for different types of skeletal events in patients with bone metastases. The skeletal complications patients may experience are of several types, but the most common are pathological fracture and severe bone pain requiring radiotherapy. Here we consider analyses of the data on control patients from Hortobagyi et al. (1998) to study the occurrence of these complications. We define type 1 events as fractures and type 2 events as bone pain requiring radiotherapy. Figure 6.9 shows a multistate diagram which can

be used to characterize possible paths for event occurrence based on these outcomes. In addition to the two skeletal complications, patients on study are at risk of death, but we suppress this state in Figure 6.9 for convenience.

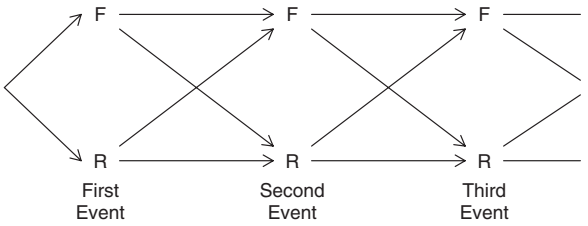


Fig. 6.9. Multistate diagram for a bivariate recurrent event process involving fractures (F) and radiotherapy (R).

We consider several different forms for the intensity functions. If events are independent one could consider Markov models with

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_j(t)\exp(x'_{ij}\beta_j)\,, \tag{6.58}$$

where T_i is the time of death, $D_i(t) = I(t \le T_i)$ as before, and $\alpha_j(t)$ is a baseline rate function for events of type $j = 1, 2$. Dependence on the history of the same type of event can be incorporated by considering, for example,

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_j(t)\exp(x'_{ij}\beta_j + \gamma_{1j}N_{ij}(t^-))\,, \tag{6.59}$$

or dependence on the history of both types of events by

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_j(t)\exp(x'_{ij}\beta_j + \gamma_{1j}N_{ij}(t^-) + \gamma_{2j}N_{i,3-j}(t^-))\,. \tag{6.60}$$

As stressed previously, it is important to recognize that the interpretation of the regression coefficient β_j is different in (6.58) and models such as (6.59) or (6.60) which involve additional conditioning. If x_{ij} is a scalar, for example, then $\exp(\beta_j)$ represents the multiplicative effect of a one-unit increase in x_{ij} on the rate of type j events in (6.58); in (6.60) $\exp(\beta_j)$ represents the effect of a one-unit increase in x_{ij} on the rate of type j events among subjects with the same cumulative number of fractures and rounds of radiotherapy for bone pain. There may be considerable differences in the estimates of regression coefficients obtained in these models when the additional terms in (6.59) and (6.60) indicate a significant dependence on prior event history.

More general models are possible within this Markov framework by introducing some form of stratification. One can, for example, let

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_{jk_j}(t)\exp(x_i'\beta_j)\,, \tag{6.61}$$

where $N_{ij}(t^-) = k_j$, $j = 1, 2$, which is analogous to (6.59) but where the effect of the previous type j events is dealt with through stratification; that is, a separate baseline intensity is introduced for each cumulative number of type j events. The impact of past events of type $3 - j$ can be assessed parametrically through models such as

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_{jk_j}(t)\exp(x_i'\beta_j + \gamma_{2j}N_{i,3-j}(t^-))\,, \tag{6.62}$$

where the two processes are independent if the $\gamma_{2j} = 0$. More highly stratified Markov models such as

$$\lambda_{ij}\left(t|H_i(t)\right) = D_i(t)\alpha_{jk_1k_2}(t)\exp(x_i'\beta_j)\,, \tag{6.63}$$

where $N_{i1}(t^-) = k_1$ and $N_{i2}(t^-) = k_2$, feature separate baseline transition intensities for different combinations of the cumulative event counts. These models lend themselves to informative graphical summaries if the number of event-dependent strata is not too large, as we illustrate below.

Table 6.7. Estimates obtained for regression analyses of fractures and rounds of radiotherapy in unstratified and stratified models.

	Unstratified Models						Stratified Model		
	Model 1			Model 2			Model 3†		
	EST.	S.E.	p	EST.	S.E.	p	EST.	S.E.	p
	Fractures (Type 1 Events)								
Positive PR status	-0.921	0.178	< 0.0001	-0.581	0.190	0.002	-0.599	0.233	0.010
Unknown PR status	0.108	0.151	0.477	0.204	0.162	0.208	0.132	0.219	0.545
Prior fracture (Y/N)	0.857	0.135	< 0.0001	0.498	0.148	0.001	0.270	0.193	0.163
Cumulative No.									
Fractures	–	–	–	0.167	0.017	< 0.0001	–	–	–
Radiotherapies	–	–	–	0.124	0.055	0.024	–	–	–
	Radiotherapy Events (Type 2 Events)								
Prior radiation	0.422	0.175	0.016	0.332	0.177	0.062	0.308	0.181	0.088
Cumulative No.									
Radiotherapies	–	–	–	0.384	0.066	< 0.0001	–	–	–
Fractures	–	–	–	0.010	0.032	0.769	0.058	0.036	0.113

†Model 3 for pathological fractures is stratified by cumulative number of fractures and rounds of radiotherapy and Model 3 for rounds of radiotherapy is stratified on cumulative rounds of radiotherapy.

Table 6.7 contains the results of fitting several models for the occurrence of fractures (top half) and the need for radiotherapy for bone pain (bottom

half). The covariates considered included age (1 if at least 50 yrs ; 0 otherwise), prior fracture status (1 if they had at least one fracture prior to study entry; 0 otherwise), prior radiation (1 if they had at least one round of radiotherapy prior to study entry; 0 otherwise), prior chemotherapy (1 if they had prior chemotherapy experience; 0 otherwise), prior hormone (1 if they had prior hormonal treatment; 0 otherwise), estrogen receptor (ER) status (0 if negative; 1 if positive; 2 if unknown), and progesterone receptor (PR) status (0 if negative; 1 if positive; 2 if unknown).

We first discuss the results from fitting models for the occurrence of fractures. Model 1 was obtained by selecting those covariates that were significant in a multivariate analysis based on (6.58), which were PR status and prior fracture. In Model 2 we add the cumulative number of fractures and rounds of radiotherapy as in (6.60) and find both significantly increase the risk of subsequent fractures. A test of the assumption of multiplicative effects for these variables via `cox.zph` gives $p < 0.001$ and $p = 0.030$ for the cumulative number of fractures and radiotherapies, respectively, indicating some model generalization is warranted. The multiplicative assumption can be relaxed with models where these covariate effects are allowed to be time-dependent by considering interactions with defined time-dependent covariates. Here we base Model 3 on (6.63) and rely on simple graphical summaries for interpretation instead. The tests of multiplicative effects of PR status and prior fracture do not suggest problems with Model 3, where PR status remains a significant prognostic variable indicating that among patients with the same number of fractures and rounds of radiotherapy, a postive PR status confers a protective effect; those with a positive PR status have lower risk of fractures than those with negative PR status ($RR = 0.549$; 95% CI (0.348, 0.867); $p = 0.010$). Among patients with the same cumulative number of fractures and rounds of radiotherapy, the prior fracture status at study entry no longer conveys useful prognostic information and could be dropped from the model. The addition of other baseline variables to Model 3 does not produce anything significant.

Figure 6.10 displays a variety of integrated baseline intensities for Model 3 where the baseline intensity corresponds to a patient with a negative PR status and no prior fracture. The top left plot gives the estimated cumulative transition intensities for the first three successive fractures among patients who have not had any radiation therapy for bone pain. The increasing trend is consistent with the regression findings for Models 1 and 2 in Table 6.7 in that the risk of a new fracture increases with each occurrence of each fracture. We see in Figure 6.10 that this is mainly a factor over the first few months on study, the intensities being roughly the same after four months. The top right plot in Figure 6.10 reveals that among patients with no on-study fractures, the risk of first on-study fracture does not change much following the first two bouts of radiation for bone pain. Comparison of the top left and bottom left plots of Figure 6.10 suggests that patients with a single round of radiotherapy have a reduced risk of a second fracture, and a smaller reduction in risk of the third fracture. There are, however, rather few individuals in the "2 fractures,

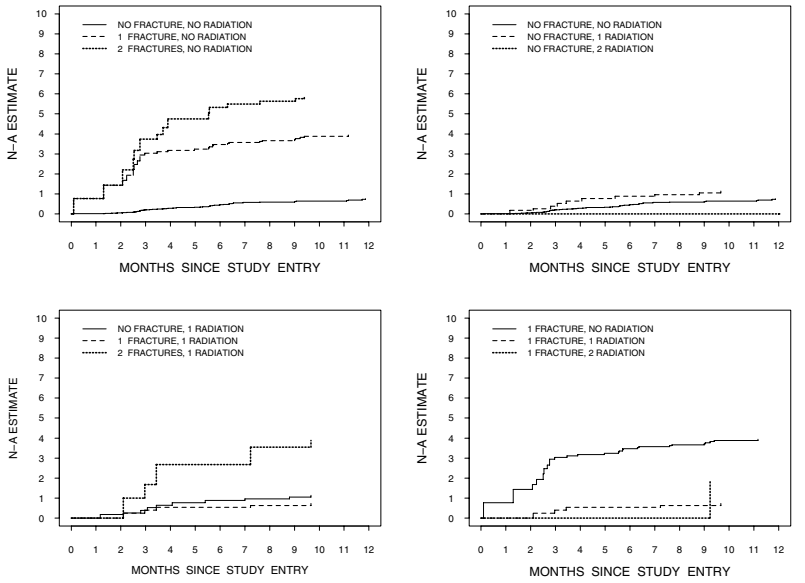


Fig. 6.10. Generalized Nelson–Aalen estimates of the cumulative baseline transition intensities based on a stratified Markov model for pathological fractures (Model 3).

1 radiation" risk sets and the estimate has a large standard error. Caution is warranted when interpreting such plots without confidence intervals, and awareness of the size of risk sets is important.

The bottom portion of Table 6.7 contains the results of corresponding analyses for the radiotherapy event. The estimates for Model 1 reveal that only prior radiotherapy was significant following backwards elimination based on (6.58). When controlling for the cumulative number of fractures and rounds of radiotherapy in Model 2 based on (6.60), this prior radiation effect was somewhat attenuated as one might expect. Tests based on `cox.zph` for Model 2 reveal evidence against the multiplicative effect of the cumulative number of rounds of radiotherapy ($p = 0.036$) but no such evidence for prior radiotherapy or the cumulative number of fractures. We therefore consider Model 3 based on (6.62) and find that prior radiotherapy and the cumulative number of fractures are not significant. Figure 6.11 gives plots for the Nelson–Aalen estimates of the integrated intensities for the first four radiotherapy events corresponding to patients with no prior radiotherapy and no on-study fractures. They reveal a greater risk of new radiotherapy among subjects with one previous event, but the risk following second and third rounds of radiotherapy appears about the same.

It should be noted that death is a competing risk for the recurrent skeletal events here, and the recurrent event intensities at time t are here, as in (6.30) and (6.58)–(6.60), conditional on $T_i \geq t$.

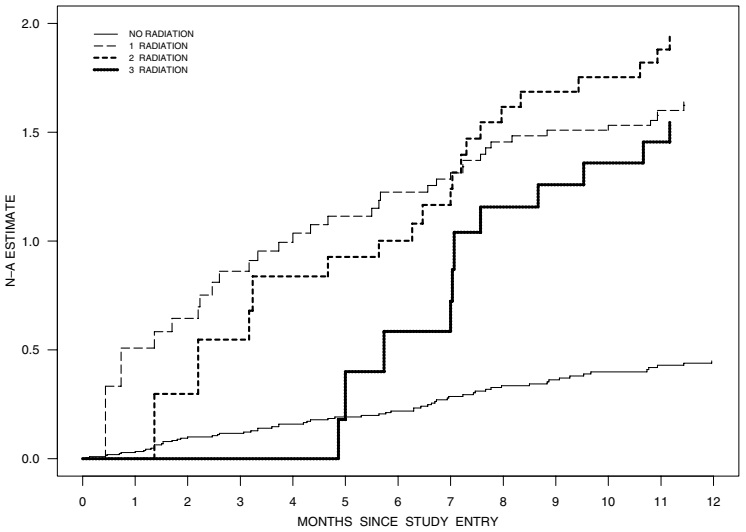


Fig. 6.11. Generalized Nelson–Aalen estimates of the cumulative baseline transition intensities based on a stratified Markov model for radiotherapy (Model 3).

6.8 Bibliographic Notes

Intensity-based models for multitype recurrent event data are considered in the multivariate counting process literature. Andersen et al. (1993) provide comprehensive coverage of this area, with an emphasis on Markov models. Other references on intensity models are given in the Bibliographic Notes for Chapters 3–5, and many of the models discussed can be adapted for multiple events.

Abu-Libdeh et al. (1990) consider multivariate mixed Poisson models with parametric conditional event rates and multivariate random effect distributions. Chen et al. (2005) describe Gibbs sampling algorithms for fitting mixed Poisson models with multivariate log-normal random effects; the problem motivating this work involved interval-censored data which we consider in Chapter 7. Ng and Cook (1999b) follow the approach of Lawless and Nadeau (1995)

and Nadeau and Lawless (1998) and derive estimating functions and optimality criteria for semiparametric rate and mean functions for multitype recurrent events. In their framework working covariance models are considered. Cai and Schaubel (2004b) formulate estimating functions for marginal analysis but do not incorporate a model for the association structure.

The two-state setting of Section 6.5 has been considered both in the literatures on multistate processes and alternating renewal processes. References germane to the treatment here include Xue and Brookmeyer (1996), Cook et al. (1999), and Lawless and Fong (1999).

The analysis of recurrent events in the presence of a terminal event is considered in Cook and Lawless (1997). There are connections with analysis of quality-adjusted lifetime data (e.g. Zhao and Tsiatis, 1997), health cost analyses (e.g. Lin et al., 1997), and any other settings where interest lies in accumulating processes with termination. This topic is discussed in Chapter 8. Strawderman (2000) discusses the connections between these problems and approaches. Ghosh and Lin (2000) develop robust tests for treatment effects based on rate functions for recurrent events with dependent termination, and Chen and Cook (2004) extend these tests to accommodate multitype recurrent events. Ghosh and Lin (2002) give marginal regression models for the number of recurrent events. Shared random effects models for recurrent events and a terminating event are considered by Wang et al. (2001), Huang and Wang (2004), Liu et al. (2004), Ye et al. (2007), and others. Approaches based on accelerated time models are considered by Ghosh and Lin (2003) and Huang and Wang (2003). Lawless et al. (2001) contains a detailed analysis of the shunt data discussed in Section 6.7.1 as well as further remarks on gap-time analyses with multitype events.

In some cases with multitype event data the type of event is not always available. The related problem of competing risk data with missing cause of failure has received considerable attention in the literature. See Craiu and Reiser (2006), Craiu and Duchesne (2004), and Dewanji and Sengupta (2003) for recent work and references. For the recurrent event context, Schaubel and Cai (2006a) describe an estimation procedure under the assumption that the mean functions for the different types of events are proportional; generalizations are also described to relax this assumption. An alternative approach based on multiple imputation is described in Schaubel and Cai (2006b).

6.9 Problems and Supplements

6.1. Section 3.5.2 discussed the use of mixture models to account for an excess of zeros relative to what one would expect from a given Poisson model. Consider a bivariate point process in which, for $k = 1, 2$, $\{N_k(t), 0 \le t\} | Z_k = 1$ follows a Poisson process with rate $\rho_k(t)$ and $N_k(t) = 0$ if $Z_k = 0$, where Z_1 and Z_2 are unobserved random variables with joint probability function $\pi_{k\ell} = \Pr(Z_1 = k, Z_2 = \ell)$, $k, \ell = 0, 1$.

a. Give the likelihood function based on a random sample of m individuals, where the data from individual i consist of $\{N_{i1}(s), N_{i2}(s),\ 0 \le s \le \tau_i\}$.
b. Develop an EM algorithm to obtain maximum likelihood estimates of the rate functions $\rho_k(s)$, $k = 1, 2$, and $\pi = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ by treating $(Z_1, Z_2)'$ as missing data.
c. Consider extensions where the $\rho_k(t)$ and $\pi_{k\ell}$ may depend on covariates.

[Section 6.1]

6.2. Consider a bivariate two-point mixture model such that given $Z_k = \ell$, $\{N_k(t), 0 \le t\}$ is a Poisson process with rate $\rho_{k\ell}$, $k = 1, 2$, $\ell = 1, 2$. Let $J(s) = 1$ if $dN_1(s) = 1$ and $J(s) = 2$ if $dN_2(s) = 1$ with $J(s) = 0$ otherwise. If $\mathcal{S}_i = \{t_{i1}, \ldots, t_{iN_i(\tau)}\}$ denote the times of events for subject i, and $J_i(s)$ is missing for some $s \in \mathcal{S}_i$, discuss strategies for maximum likelihood estimation.

[Section 6.1]

6.3. Suppose that $\{N_{ik}(t), 0 \le t\} \,|Z_k = 1$ follows a renewal process with gap time hazard functions $h_k(w)$, for $k = 1, 2$. Reconsider Problem 6.1 in this context.

[Section 6.1]

6.4. Consider the random effects model at the end of Section 6.3, where events of types $j = 1, \ldots, J$ for individual i follow independent Poisson processes with rates $u_{ij}\rho_{ij}(t)$, given the vector $u_i = (u_{i1}, \ldots, u_{iJ})'$ of random effects. Suppose $u_{ij} = v_i + w_{ij}$, where the v_i and w_{ij} are mutually independent gamma random variables with means equal to 1 and variances ϕ_v and ϕ_w, respectively. Derive the likelihood function for ϕ_v, ϕ_w and the parameters θ that specify the $\rho_{ij}(t)$.

[Section 6.3]

6.5. Consider a pseudo-score statistic, denoted here $U_j(0)$, for testing $H_0 : \beta_j = 0$, based on (6.16). Derive an estimate of a robust covariance matrix for $(U_1(0), \ldots, U_J(0))$ using (6.21) and discuss how to construct a global test of $H_0 : \beta_1 = \cdots = \beta_J = 0$.

[Section 6.4]

6.6. Suppose that, conditional on a random effect u_i, a recurrent event process and a terminating event process are independent, where (i) the recurrent events follow a Poisson process with rate function $u_i\,\rho(t)$ and (ii) the time T_i to the terminating event has hazard function $u_i\,h(t)$. Suppose that u_i has a gamma distribution with mean 1 and variance ϕ.

a. Given the recurrent event history $\{N_i(s) : 0 \le s < t\}$ and that $T_i \ge t$, determine the probability that $T_i > t'$ where $t' > t$.

b. For the case of a homogeneous Poisson process, with $\rho(t) = \rho$, determine the two mean functions $E\{N_i(t)\}$ and $E\{N_i(t)|T_i \geq t\}$. Plot the two functions for the case where $\rho = 1$, $h(t) = 1$, and $\phi = 1$. Repeat this for the case where $\rho = 1$, $h(t) = 1$, and $\phi = 0.04$. What happens when $\phi = 0$?

This process is not conceptually appealing because it assumes the event process continues after termination. Compare it with the model in (6.33) and (6.34), when $u_{i1} = u_{i2}$.

[Section 6.6; Huang and Wang, 2004; Ye et al., 2007]

6.7. Motivate the estimates (6.26) and (6.27) by using the fact that if τ_i $(i = 1, \ldots, m)$ are fixed values independent of the event processes, and if the processes are of mixed Poisson type, then if $n_{ij} = N_{ij}(\tau_i)$ and $\mu_{ij} = \mu_{ij}(\tau_i)$,

$$\text{var}(n_{ij}) = \mu_{ij} + \phi_j \mu_{ij}^2 , \qquad \text{and} \qquad \text{cov}(n_{i1}, n_{i2}) = \phi_{12} \mu_{i1} \mu_{i2} .$$

[Sections 6.3, 6.4]

6.8. Data on pulmonary exacerbations in persons with cystic fibrosis were considered in Sections 1.2.3, 4.3.2, and 5.5.1; see Appendix D.3. Individuals are not at risk of a new exacerbation while they are being treated with antibiotics for an existing one, so the process is actually as in Figure 6.1. Previous discussion focused on the occurrence of new exacerbations (transitions into state 2).

Use the methods of this chapter to examine both types of transitions, including the possibility that the durations of sojourns in the two states may be related.

[Section 6.5]

6.9. Suppose that a terminating event for a recurrent event process can occur only at the time of each recurrent event. In particular, given that the terminating event has not occurred previously, the probability that it occurs at the time of the jth recurrent event is π_j.

a. For the case in which $\{N(t), 0 \leq t\}$ is a Poisson process with rate function $\rho(t; \alpha)$, give the maximum likelihood estimating equations for α and π_j, $j = 1, 2, \ldots$, based on complete data for independent individuals $i = 1, 2, \ldots, m$, with individual i observed over $[0, \tau_i]$.
b. Determine the marginal survivor function $S(t)$ for the time T of the terminating event, and the marginal mean function $E\{N(t)\}$ for the case where $\pi_j = \pi$, $j = 1, 2, \ldots$. Give an expression for $E\{N(t)|T = t\}$.
c. Generalize parts (a) and (b) for the case where π_j may depend on the time of the event, so that

$$\Pr(\text{termination at time } t | j\text{th event at } t) = \pi_j(t) .$$

[Section 6.6]

6.10. Model checking is often carried out in more complex multivariate processes by fitting expanded models. Describe how intensity-based residuals and methods discussed in Section 5.2.3 can be used for the settings in this chapter.

7

Observation Schemes Giving Incomplete or Selective Data

Thus far we have assumed that the event history process can be observed continuously over some followup period for each individual, and it has been assumed that the selection of individuals and their followup periods are independent in the sense described in Sections 1.4 and 2.6, and thus ignorable. In some settings these conditions are violated; for example, an individual may be observed only intermittently, or loss-to-followup may not be independent in the sense needed for ordinary analysis. In addition, individuals may be selected for a study because their event history satisfies some condition. In the following sections we discuss how to deal with these issues.

7.1 Intermittent Observation During Followup

In Section 1.4.4 we discussed how in some studies it is only possible to inspect or observe individuals intermittently. If the followup period for individual i is $[\tau_{i0}, \tau_{i1}]$, then suppose the individual is seen at the times $\tau_{i0} = b_{i0} < b_{i1} < \cdots < b_{ik_i} = \tau_{i1}$. We assume for now that any covariates x_i are fixed; time-varying covariates are considered in subsequent sections. The b_{ij} can be fixed or random, but they have to satisfy conditions discussed below in order to make standard methods valid.

Let H_{i0} denote any event or covariate history at time b_{i0}^+ that is needed for the models considered, and let $\triangle H_{ij}$ denote the event history data over the interval $B_{ij} = (b_{i,j-1}, b_{ij}]$, which is not observed until time b_{ij}. Let $\tilde{H}_{ij} = \{b_{i0}, H_{i0}, b_{i1}, \triangle H_{i1}, \ldots, b_{ij}, \triangle H_{ij}\}$ denote the observed history up to time b_{ij}. We note that given H_{i0}^+,

$$\Pr(\tilde{H}_{ik_i}) = \prod_{j=1}^{k_i} \Pr(\triangle H_{ij} | \tilde{H}_{i,j-1}, b_{ij}) \prod_{j=1}^{k_i} \Pr(b_{ij} | \tilde{H}_{i,j-1}) \,. \tag{7.1}$$

We assume that the following two conditions hold in conjunction with the methods below.

(i) $\Pr(b_{ij}|\tilde{H}_{i,j-1})$ does not contain any parameters that appear in the model for the event process.
(ii) $\Pr(\triangle H_{ij}|\tilde{H}_{i,j-1}, b_{ij}) = \Pr(\triangle H_{ij}|\triangle H_{i1}, \ldots, \triangle H_{i,j-1})$, which stands for the probability that applies when the times $b_{i1}, \ldots, b_{ij}$ are fixed, and so completely independent of the event process.

Condition (i) means that the inspection process is noninformative and so there is no loss of information regarding the event process if we ignore the second term in (7.1). Condition (ii) states that although b_{ij} may depend on the prior history $\tilde{H}_{i,j-1}$ of event or inspection times up to time $b_{i,j-1}$, it cannot depend on events after $b_{i,j-1}$ and the conditional probability distribution of the event history in B_{ij} must be the same as if b_{ij} were a function of prior inspection times alone.

Based on assumptions (i) and (ii), and given data on m independent individuals, we implicitly condition on the b_{ij} values and consider the likelihood function for the parameters θ in the event history process as

$$L(\theta) = \prod_{i=1}^{m} \prod_{j=1}^{k_i} \Pr(\triangle H_{ij}|\triangle H_{i1}, \ldots, \triangle H_{i,j-1}) \,. \tag{7.2}$$

If only condition (ii) is satisfied this is a partial likelihood, and if both (i) and (ii) hold it is an ordinary likelihood. Either way, it can be used for maximum likelihood inference through the usual procedures.

Sometimes the entire event process over the period $(b_{i,j-1}, b_{ij}]$ can be determined retrospectively at time b_{ij}. This is the case in studies where individuals keep daily diaries on event occurrence, but the diaries are only available at periodic clinic visits. In such settings the likelihood (7.2) is the same as if we observed the event process continuously, and the methods of earlier chapters apply. For the remainder of this section, we deal with the case where event times cannot be determined retrospectively, but instead only the total number n_{ij} of events in B_{ij} can be determined. This kind of data arises in many medical settings where the event of interest is not symptomatic, but is detectable by radiographic examination, MRI or PET scans, for example, and is often referred to as interval-count data. As discussed in Chapter 2, the joint distribution of interval counts is intractable for most models. This is particularly true for renewal processes, where missing information about event times makes calculation even more difficult than usual; Problem 7.1 provides an illustration. For Poisson and mixed Poisson processes, however, interval counts are readily handled, and robust methods for rate and mean functions are also straightforward. We describe the appropriate methodology in the following sections.

7.1.1 Methods Based on Poisson Processes

We assume in this section that $\{N_i(t), 0 \leq t\}$ are independent Poisson processes, conditional on unobserved random effects u_i and covariate vectors x_i,

with multiplicative conditional intensity functions

$$\lambda(t|u_i, x_i) = u_i \rho_0(t) \exp(x_i'\beta)\,. \tag{7.3}$$

The u_i are independent and identically distributed, with $E(u_i) = 1$ and $\text{var}(u_i) = \phi$. An ordinary Poisson process without random effects is given by the limiting case $\phi = 0$, in which case each u_i in (7.3) equals one. This is the model of Section 3.5, and when event times are observed both parametric and nonparametric specifications for $\rho_0(t)$ are readily handled as described there. With only interval counts available, semiparametric maximum likelihood is considerably more challenging, however, and so we assume here that $\rho_0(t)$ is specified parametrically as $\rho_0(t;\alpha)$. Flexibility can be achieved, if desired, by taking $\rho_0(t)$ as piecewise-constant (e.g. Lawless and Zhan, 1998) or modeling it with splines or other flexible parametric forms.

We assume that the u_i have distribution function $G(u;\phi)$ and let $\theta = (\alpha', \beta', \phi)'$ denote the vector of unknown parameters. We also define

$$\begin{aligned}\mu_{ij}(\alpha,\beta) &= E\{N_i(b_{i,j-1}, b_{ij})\} \\ &= \{\mu_0(b_{ij};\alpha) - \mu_0(b_{i,j-1};\alpha)\} \exp(x_i'\beta)\,,\end{aligned} \tag{7.4}$$

where $\mu_0(t;\alpha) = \int_0^t \rho_0(s;\alpha)ds$. For simplicity we continue to assume that the covariate vector x_i is fixed, but straightforward modifications of the expressions that follow allow x_i in (7.4) to be replaced with x_{ij}, so that time-varying covariates which change values at observation times can be accommodated. Finally, it is assumed that conditions (i) and (ii) following (7.1) hold; this requires that b_{ij} be independent of u_i, given $\tilde{H}_{i,j-1}$.

The likelihood function under these conditions is given by the joint distributions of $(n_{i1}, \ldots, n_{ik_i})$ given x_i, where $n_{ij} = N_i(b_{i,j-1}, b_{ij})$. This is proportional to

$$L(\theta) = \prod_{i=1}^{m} \int_0^\infty \prod_{j=1}^{k_i} (u_i \mu_{ij})^{n_{ij}} \exp(-u_i \mu_{ij}) dG(u_i;\phi)\,, \tag{7.5}$$

where we write μ_{ij} for $\mu_{ij}(\alpha,\beta)$ and omit terms $(n_{ij}!)^{-1}$ in this likelihood. The log-likelihood $\ell(\theta) = \log L(\theta)$ can be maximized and an asymptotic covariance matrix for $\widehat{\theta}$ obtained, by using general optimization software described in Appendix B. As in Section 3.5.3, we note that when the u_i have a gamma distribution (2.28) with mean 1 and variance ϕ, the integrals in (7.5) have closed-form expressions, giving

$$L(\theta) = \prod_{i=1}^{m} \left(\prod_{j=1}^{k_i} \mu_{ij}^{n_{ij}} \right) \frac{\Gamma(n_{i\cdot} + \phi^{-1})\phi^{n_{i\cdot}}}{\Gamma(\phi^{-1})(1 + \phi\mu_{i\cdot})^{n_{i\cdot}+\phi^{-1}}}\,, \tag{7.6}$$

where $\mu_{i\cdot} = \sum_{j=1}^{k_i} \mu_{ij}, n_{i\cdot} = \sum_{j=1}^{k_i} n_{ij}$, and $\Gamma(\cdot)$ is the gamma function.

The likelihood score functions for α, β, and ϕ based on (7.4) and (7.6) can, after some algebra, be written as

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{m} \sum_{j=1}^{k_i} \frac{(n_{ij} - \mu_{ij})}{\mu_{ij}} \left(\frac{\partial \mu_{ij}}{\partial \alpha} \right) - \sum_{i=1}^{m} \frac{(n_{i.} - \mu_{i.})\phi}{1 + \phi \mu_{i.}} \left(\frac{\partial \mu_{i.}}{\partial \alpha} \right) \tag{7.7}$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{m} \frac{(n_{i.} - \mu_{i.})}{1 + \phi \mu_{i.}} x_i \tag{7.8}$$

$$\frac{\partial \ell}{\partial \phi} = \sum_{i=1}^{m} \left\{ \frac{n_{i.} - \mu_{i.}}{\phi(1 + \phi \mu_{i.})} + \frac{\log(1 + \phi \mu_{i.})}{\phi^2} - \frac{\sum_{j=1}^{n_{i.}} [1 + \phi(j-1)]^{-1}}{\phi} \right\}. \tag{7.9}$$

They can be utilized in a maximizer for $\ell(\theta)$ if desired or, alternatively, software that solves the equations $\partial \ell / \partial \alpha = 0, \partial \ell / \partial \beta = 0, \partial \ell / \partial \phi = 0$ can be used to obtain $\widehat{\alpha}$, $\widehat{\beta}$, and $\widehat{\phi}$. A test of the need for random effects is conveniently carried out using the likelihood ratio statistic for the hypothesis $H_0 : \phi = 0$. This is given by $W = 2\ell(\widehat{\alpha}, \widehat{\beta}, \widehat{\phi}) - 2\ell(\widehat{\alpha}_0, \widehat{\beta}_0, 0)$, where $\widehat{\alpha}_0$ and $\widehat{\beta}_0$ are the maximum likelihood estimates under the Poisson model. Under H_0, W asymptotically has distribution $.5I(W = 0) + .5\chi^2_1$. An illustration involving the negative binomial model just described is given in Section 7.1.3. In the next section we consider robust methods of estimating the parameters α, β in the marginal rate function $\rho_0(t; \alpha) \exp(x'\beta)$ and mean function $\mu_0(t; \alpha) \exp(x'\beta)$. As in Section 3.6, these methods provide estimates that are applicable to other processes besides Poisson or mixed Poisson, provided the inspection times are determined independently of the event processes, given the x_i.

7.1.2 Robust Estimation of Rate and Mean Functions

As in Section 3.6, we assume that the process mean function takes the form

$$E\{N_i(t) | x_i\} = \mu_0(t; \alpha) \exp(x_i' \beta) \tag{7.10}$$

and the corresponding rate function is $\rho_0(t; \alpha) \exp(x_i' \beta)$. Provided the b_{ij} are determined independently of the event process, it follows that, conditional on the b_{ij}, $E(n_{ij}) = \mu_{ij}$, using the notation of the preceding section. The expectations of (7.7) and (7.8) are then equal to zero under the model (7.10), so that solving $\partial \ell / \partial \alpha = 0, \partial \ell / \partial \beta = 0$ gives consistent estimators $\tilde{\alpha}$ and $\tilde{\beta}$ for any specified value of ϕ. The simplest procedure is to use $\phi = 0$, which corresponds to Poisson maximum likelihood, and we use this here. We must, however, consider variance estimation for $\tilde{\alpha}$ and $\tilde{\beta}$ on the assumption that the negative binomial process giving (7.7) and (7.8) may not hold.

Let us rewrite the estimating functions (7.7) and (7.8) with $\phi = 0$ as

$$U(\alpha,\beta) = \begin{pmatrix} \sum_{i=1}^{m} U_{1i}(\alpha,\beta) \\ \sum_{i=1}^{m} U_{2i}(\alpha,\beta) \end{pmatrix} = \begin{pmatrix} U_1(\alpha,\beta) \\ U_2(\alpha,\beta) \end{pmatrix},$$

where

$$U_{1i}(\alpha,\beta) = \sum_{j=1}^{k_i} \frac{n_{ij} - \mu_{ij}}{\mu_{ij}} \left(\frac{\partial \mu_{ij}}{\partial \alpha}\right)$$
$$U_{2i}(\alpha,\beta) = (n_{i.} - \mu_{i.}) x_i \,.$$

It follows from large sample results for estimating equations (see Appendix A) that under mild conditions, as $m \to \infty$ the distribution of $\sqrt{m}(\tilde{\alpha}' - \alpha', \tilde{\beta}' - \beta')'$ approaches a multivariate normal distribution with mean 0 and covariance matrix that is estimated consistently by

$$\widehat{\text{var}} \begin{pmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{pmatrix} = \tilde{A}^{-1} \tilde{B} \tilde{A}^{-1} \,, \tag{7.11}$$

where

$$A = \frac{1}{m} \begin{pmatrix} -\partial U_1/\partial \alpha' & -\partial U_1/\partial \beta' \\ -\partial U_2/\partial \alpha' \quad \text{and} & -\partial U_2/\partial \beta' \end{pmatrix}, \qquad B = \frac{1}{m} \sum_{i=1}^{m} U_i U_i' \,,$$

and $\tilde{A}$, $\tilde{B}$ are obtained by replacing parameters α, β with $\tilde{\alpha}$, $\tilde{\beta}$. Because $U(\alpha,\beta)$ is the likelihood score function for a Poisson process, the matrix $\tilde{A}$ is simply the observed information matrix from the Poisson model with mean function (7.10) and $\theta = (\alpha', \beta')'$.

The model (7.10) and estimating functions $U(\alpha,\beta) = 0$ can thus be handled by standard optimization software, as follows. Let

$$\ell_p(\theta) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} (n_{ij} \log \mu_{ij} - \mu_{ij}) \,, \tag{7.12}$$

where $\mu_{ij} = [\mu_0(b_{ij};\alpha) - \mu_0(b_{i,j-1};\alpha)] \exp(x_i'\beta)$, denote the Poisson process log-likelihood. Using an optimization routine, maximize $\ell_p(\theta)$ to obtain $\tilde{\theta} = (\tilde{\alpha}', \tilde{\beta}')'$, the solution to $U(\alpha,\beta) = 0$. Good software also gives the information matrix $A(\theta) = (-\partial^2 \ell / \partial\theta \partial\theta')$ evaluated at $\tilde{\theta}$, which is $\tilde{A}$ needed for (7.11). Finally, obtain $\tilde{B}$ by direct calculation, and then get the covariance matrix estimate (7.11).

An alternative to the preceding approach would be to use generalized linear model software for count data. The model (7.10) is not of a form that allows log-linear model software for counts to be used on the n_{ij}, however, because the form of $\mu_{ij} = E(n_{ij})$ is not in general linear in the parameters for any link function. An exception is when $\mu_0(t)$ is linear and the rate function is

constant; then, with $\mu_0(t) = \alpha t$ we have $\mu_{ij} = (b_{ij} - b_{i,j-1}) \exp(x_i'\beta + \log \alpha)$. To use generalized linear model software in other cases, we could approximate μ_{ij} by a generalized linear form, by replacing $\mu_0(b_{ij}) - \mu_0(b_{i,j-1})$ in μ_{ij} with a function evaluated at some point in B_{ij}. In particular, by the Mean Value Theorem we can write (7.4) as

$$\mu_{ij} = \mu(b_{ij}) - \mu(b_{i,j-1}) = \{(b_{ij} - b_{i,j-1})\mu_0'(b_{ij}^*)\} \exp(x_i'\beta)\,,$$

where b_{ij}^* is a value in $(b_{i,j-1}, b_{ij}]$. For certain forms of $\mu_0(t)$ this can be expressed in log-linear form. Because robust variance estimates still have to be computed, software that does this is preferred.

We remark as well that estimating functions which use a parametric working covariance matrix for $(n_{i1}, \ldots, n_{ik_i})'$ can also be considered (e.g. Lawless and Zhan, 1998). Section 7.1.4 outlines this approach for the more general setting of multiple event types; in favorable circumstances this can provide more efficient estimates of α and β plus estimates of association. Problem 7.3 gives results for the present context of a single event type. Finally, although the methods given here are robust to departures from a Poisson process they do, unlike the methods of Section 3.6, rely on a parametric model for $\rho_0(t)$. It is possible, but complicated, to develop nonparametric and semiparametric methods; see Sun (2006, Ch. 8) and other references in the Bibliographic Notes for this chapter. Parametric models are much easier to handle and, if desired, flexible specifications using piecewise-constant, spline-based, or other weakly parametric baseline rate functions can be used, as indicated before (7.4). Model checks involving model expansion and comparison of the n_{ij} and estimates $\widehat{\mu}_{ij}$ from parametric models should of course be made. The next section illustrates the methodology.

7.1.3 Illustration: Superficial Tumors of the Bladder

Lawless and Zhan (1998) discuss data given by Byar (1980), which arose in a clinical trial for patients with bladder cancer. All subjects had superficial bladder tumors at the time they entered the trial; these tumors were removed and they were then assigned randomly to receive one of three treatments: Pyridoxine, Thiotepa, or a placebo. At subsequent followup visits any new tumors were removed and treatment continued. The data (see Appendix D) consist of the months from the beginning of the study to each followup inspection and the number of tumors found at each inspection. The number of initial tumors at the beginning of the study, and the diameter of the largest such tumor, were also recorded as covariates. In the discussion here we treat the development of a tumor as an event; because of the intermittent inspection of subjects, only interval counts are available. Details about how the inspection times b_{ij} $(j = 1, \ldots, k_i)$ were determined for each subject are not given; we assume the conditions (i) and (ii) following (7.1) are satisfied. We consider here only the Thiotepa and placebo groups for simplicity; there are 38 and

47 subjects, respectively, in the two groups. The time in study ranged from 1 to 53 months, and the number of inspections k_i ranged from 1 to 38. Figure 7.1 shows a plot of the total number of tumors n_i versus time in study τ_i for each subject. There are many subjects who experienced no tumors: 20 in the Thiotepa group and 18 in the placebo group.

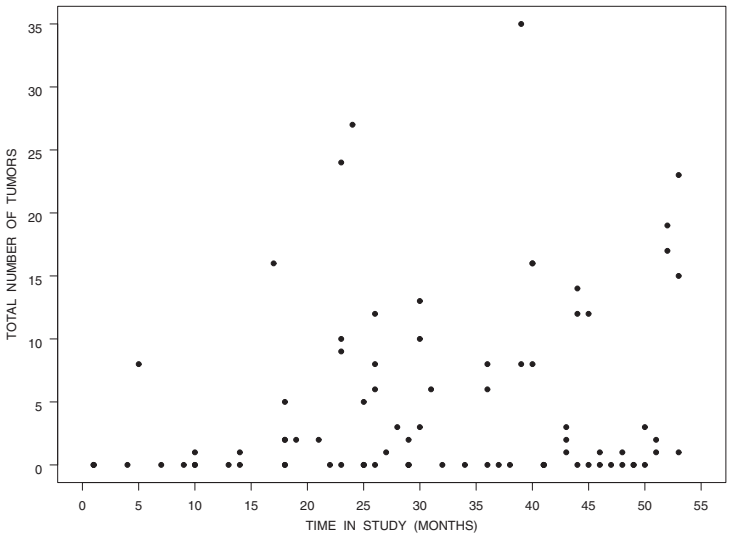


Fig. 7.1. Total number of tumors versus time in study for each subject.

A primary objective is to compare the two treatment arms with respect to the frequency tumors are present at followup visits (Byar, 1980), and several authors (e.g. Wei et al., 1989; Therneau and Grambsch, 2000, Section 8.5.4; Kalbfleisch and Prentice, 2002, Section 9.4.3; Nelson, 2003, Section 1.2) have simply treated followup times at which one or more tumors were present as exact event times. In that case, methods in the preceding chapters can be applied. Such analysis is problematic, however, because the followup times are apparently scheduled. We consider here analyses where the number of tumors n_{ij} observed at inspection time b_{ij} is treated as an event count for the interval $(b_{i,j-1}, b_{ij}]$. Some of these tumors may be recurrences of previous tumors which were removed and some may be new tumors, but in any case we can, if desired, estimate the probability of one or more tumors being present at given times.

We consider analyses based on models for the interval counts n_{ij}. Three covariates are considered: x_{i1} =I(subject i received Thiotepa), x_{i2} =number

of initial tumors present at randomization, and x_{i3} =diameter (cm) of largest initial tumor. We consider models for which

$$E\left(n_{ij}|x_i\right) = \mu_{ij} = \alpha_1 \left(b_{ij}^{\alpha_2} - b_{i,j-1}^{\alpha_2}\right) \exp\left(x_i'\beta\right) \tag{7.13}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, k_i$, where $\alpha_1 > 0$ and $\alpha_2 > 0$; this model arises from a continuous time model with mean function $\mu(t|x) = \alpha_1 t^{\alpha_2} \exp(x'\beta)$. We also consider, following Section 3.3 and Lawless and Zhan (1998), a model in which $\mu(t|x)$ is a piecewise-linear function. Two approaches to estimation are considered: (i) a negative binomial model, with likelihood function (7.6) used to estimate parameters θ in the mean function and the variance parameter ϕ, and (ii) robust estimation of θ based on Poisson estimating functions, as described in Section 7.1.2.

Table 7.1 shows results from four approaches, as follows.

(a) The negative binomial model (7.6) with μ_{ij} given by (7.13), fitted by maximum likelihood.
(b) The negative binomial model (7.6) with $\mu_0(t)$ in (7.10) piecewise-linear, as discussed in Lawless and Zhan (1998, Section 5); there are 8 pieces, with changes of slope at $t = 5.5$, 10.5, 15.5, 20.5, 25.5, 30.5, and 40.5 months. The final piece terminates at $t = 53$, which is the largest followup time.
(c) The mean function model (7.13), fitted by using the robust Poisson estimating equations; this involves obtaining θ by maximizing (7.12) and then obtaining a robust covariance matrix estimate (7.11).
(d) The piecewise-linear mean function of (b), combined with a mixed Poisson covariance structure for $n_i = (n_{i1}, \ldots, n_{ik_i})'$ discussed in Problem 7.3. This assumes that $\text{var}(n_{ij}) = \mu_{ij} + \phi\mu_{ij}^2$ and that $\text{cov}(n_{ij}, n_{i\ell}) = \phi\mu_{ij}\mu_{i\ell}$ for $j \neq \ell$.

The estimates of the regression coefficients and their standard errors, as well as the variance parameter ϕ, are almost the same under the two-parameter model (a) for the baseline mean functions in (7.13) and the eight-parameter piecewise-linear model (b). The robust Poisson estimating equations (c) give slightly different estimates of covariate effects, especially for β_1 (treatment) and β_2 (number of initial tumors), although the differences are not practically important. The robust estimating function approach (d) gives regression coefficient estimates that are in very close agreement with the negative binomial models. The estimate of the variance parameter ϕ, which is obtained from a robust estimating equation in this case, is a little smaller than the negative binomial maximum likelihood estimates.

Model checks can be based on an examination of models (7.10) with additional covariate structure, and by comparisons of the n_{ij} and corresponding fitted values $\widehat{\mu}_{ij}$. The sparseness of the counts makes formal testing difficult, but there do not appear to be any major problems with the models considered.

The analyses all indicate that subjects in the Thiotepa treatment group have a lower rate of tumor occurrence than subjects in the placebo group,

Table 7.1. Maximum likelihood estimates from negative binomial models and robust estimates from mean function models for bladder tumor appearance.

	(a) Neg. Bin. and (7.13)		(b) Neg. Bin. P.W. Linear†		(c) Robust with (7.13)		(d) Robust with P.W. linear†	
Parameter	EST.	S.E.	EST.	S.E.	EST.	S.E.	EST.	S.E.
β_1 (trt)	-1.22	0.40	-1.22	0.38	-0.79	0.32	-1.21	0.32
β_2 (number)	0.38	0.11	0.38	0.10	0.26	0.07	0.38	0.09
β_3 (size)	-0.02	0.14	-0.01	0.13	-0.03	0.10	-0.01	0.11
α_1	0.17	0.08	–	–	0.22	0.04	–	–
α_2	0.83	0.05	–	–	0.82	0.04	–	–
ϕ	2.34	0.49	2.37	0.50	–	–	1.85	0.40

† From Lawless and Zhan (1998, Table 7), which also gives an estimate for the baseline mean function.

and that subjects with a large number of initial tumors have a higher rate of occurrence. The maximum size of the initial tumors is not significant. It is also noted that there is strong evidence of extra-Poisson variation. Standard errors for the regression coefficient estimates under the Poisson model are much smaller than the robust standard errors, and lead to confidence intervals that are too narrow and $p-$values that are too small.

The baseline mean function $\mu_0(t)$ can also be estimated under the models above. For the models (a) and (c) this is $\mu_0(t) = \alpha_1 t^{\alpha_2}$ and for the piecewise-constant rate function models (b) and (d) it is

$$\mu_0(t) = \sum_{r=1}^{R} \alpha_r \Delta_r(t)\,,$$

where $\rho_0(t) = \alpha_r$ for $a_{r-1} < t \le a_r$ $(r = 1, \ldots, R)$ and $\Delta_r(t) = \min(a_r, t) - \min(a_{r-1}, t)$ is the overlap between $(a_{r-1}, a_r]$ and $[0, t]$. The models (b) and (d) in Table 7.1 used $R = 8$, with $(a_0, a_1, \ldots, a_8)$ equal to (0, 5.5, 10.5, 15.5, 20.5, 25.5, 30.5, 40.5, 53.0). The estimates of $\mu_0(t)$ under the two models are very similar. For example, at $t = 15.5$, 30.5, and 53.0 the estimates are (1.71, 3.00, 4.74) for models (a) and (c), and (1.55, 2.98, 4.70) for models (b) and (d). Both models suggest that the baseline rate function decreases slightly with time on study.

The robust mean function analyses considered here require that the inspection and loss-to-followup processes are independent of the tumor occurrence process, given the covariates. A look at the raw data reveals that individuals in the Thiotepa treatment group are inspected more frequently than those in the placebo group; this is related to the way the treatment is administered (Byar et al., 1986) and does not on its own bias the analyses. We should consider, however, whether there is any evidence of a relationship between inspection or drop-out times and prior event history, because that could bias

the robust analysis. For example, if subjects with many prior tumors were more likely to drop out of the study, then the robust marginal estimates of the mean and rate functions would tend to be biased downwards. This might, for example, explain the smaller covariate effects in Table 7.1 under model (c), whose estimation did not account for association among the event counts for an individual. We discuss this issue further and consider ways to deal with it in Section 7.2.3. For the present study, there were a few withdrawals for patients who experienced bladder cancer disease progression, but there is no strong indication of event-dependent censoring.

If the negative binomial process is a satisfactory model for tumor occurrence, then event-dependent inspection or loss-to-followup does not create bias, provided condition (ii) in Section 7.1 holds. An assessment of the negative binomial model's adequacy shows no major problems. We observe that the robust and negative binomial estimates of covariate effects in Table 7.1 are very similar for the piecewise-linear mean models, and are also close to the negative binomial estimates with model (7.13).

7.1.4 Interval-Count Data for Multiple Events

The methods of Sections 7.1.1 and 7.1.2 can be extended to the case of J different types of recurrent events, as in Section 6.4. Let $u_i = (u_{i1}, \ldots, u_{iJ})'$ be a vector of random effects and assume that conditional on u_i and covariate vector x_i the events occur according to J independent Poisson processes, with intensity functions

$$\lambda_j(t|u_i, x_i) = u_{ij}\rho_{0j}(t)\exp(x_i'\beta_j)\,, \qquad j = 1, \ldots, J\,. \tag{7.14}$$

The u_i are assumed to be i.i.d. with multivariate distribution function $G(u;\phi)$. In some cases $u_{i1}, \ldots, u_{iJ}$ may not be functionally independent, and we can rewrite u_i in terms of fewer than J random variables (see Section 6.3). The likelihood function based on the counts of type j events ($j = 1, \ldots, J$) over intervals $B_{ij1}, \ldots, B_{ijk_{ij}}$, where $B_{ij\ell} = (b_{ij,\ell-1}, b_{ij\ell}]$, is by direct extension of (7.5),

$$L(\theta) = \prod_{i=1}^{m}\int_0^\infty \prod_{j=1}^{J}\prod_{\ell=1}^{k_{ij}}(u_{ij}\mu_{ij\ell})^{n_{ij\ell}}\exp(-u_{ij}\mu_{ij\ell})dG(u_i;\phi)\,, \tag{7.15}$$

where $n_{ij\ell} = N_{ij}(b_{ij,\ell-1}, b_{ij\ell})$ is the number of type j events in time interval $B_{ij\ell}$,

$$\mu_{ij\ell} = E(n_{ij\ell}) = [\mu_0(b_{ij\ell}) - \mu_0(b_{ij,\ell-1})]\exp(x_i'\beta_j)\,,$$

and the dimension of the integral equals the number of functionally independent random variables in u_i. In some settings the $B_{ij\ell}$ are the same for each $j = 1, \ldots, J$ but we use the general notation to reflect that certain events may not be counted at every inspection time.

The likelihood (7.15) in general requires numerical integration. Chen et al. (2005) consider a model in which u_i has a multivariate log-normal distribution. The integrals in (7.15) can be calculated using Gauss–Hermite quadrature (e.g. Naylor and Smith, 1982), and parametric models are readily fitted with general optimization software. Semiparametric models are not easily fitted in this setting but weakly parametric forms for $\rho_{0j}(t;\alpha)$ may be used. In some cases a possible alternative approach is to fit separate mixed Poisson models for each event type $j = 1, \ldots, J$, thus obtaining estimates of β_j and $\mu_{0j}(t) = \int_0^t \rho_{0j}(s)ds$, and also of any parameters in the marginal distribution of u_{ij}. Then, these estimates can be plugged into (7.15), and the likelihood maximized for the remaining parameters in the distribution $G(u_i;\phi)$. For example, if $\log u_i$ were multivariate normal with mean 0 and covariance matrix $V = [\phi_{jj'}]$, the variances $\phi_{jj} = \phi_j$ could be estimated from separate analyses of each event of type, and then the correlations or the covariances $\phi_{jj'}$ $(j \neq j')$ could be estimated at the second stage. Such a procedure would need modification when the u_{ij} were not functionally independent; in that case one might estimate all of the parameters in $G(u;\phi)$ at the second stage.

We can also extend the methods in Section 7.1.2 for robust estimation of rate and mean functions. The rate functions for the J types of event are assumed to be of the form

$$\rho_{ij}(t) = \rho_{0j}(t;\alpha)\exp(x_i'\beta_j) \qquad j = 1, \ldots, J\,, \tag{7.16}$$

and the corresponding mean functions are denoted $\mu_{ij}(t) = \mu_{0j}(t)\exp(x_i'\beta_j)$, where $\mu_{0j}(t) = \int_0^t \rho_{0j}(s)ds$. In Section 7.1.2 we considered simple estimating functions for $\theta = (\alpha', \beta')'$ based on Poisson models. Here, we extend the approach by allowing a model for the variances and covariances of the interval counts $n_{ij\ell}$ that is based on the random effects formulation of this section, but without any need for a specific distribution for the u_{ij}.

We consider the case where $u_i = (u_{i1}, \ldots, u_{iJ})'$ has $J \times J$ nonsingular covariance matrix $\Phi = [\phi_{jj'}]$. In that case, it follows from (7.14) and the conditional variance formula that the means of the $n_{ij\ell}$ are $\mu_{ij\ell} = \mu_{ij}(b_{ij\ell}) - \mu_{ij}(b_{ij,\ell-1})$ and that

$$\mathrm{var}(n_{ij\ell}) = \mu_{ij\ell} + \phi_{jj}\mu_{ij\ell}^2 \tag{7.17}$$

$$\mathrm{cov}(n_{ij\ell}, n_{ij'\ell'}) = \phi_{jj'}\mu_{ij\ell}\mu_{ij'\ell'}, \qquad (j,\ell) \neq (j',\ell')\,, \tag{7.18}$$

where $\ell = 1, \ldots, k_{ij}$ and $j, j' = 1, \ldots, J$. Let $\theta = (\alpha', \beta')'$ denote the parameter vector, let n_i denote the counts $n_{ij\ell}$ across $j = 1, \ldots, J$ and $\ell = 1, \ldots, k_{ij}$ listed in vector form, and let $\mu_i = E(n_i)$ and $V_i = \mathrm{cov}(n_i)$ denote the corresponding mean vector and covariance matrix for n_i. The entries of V_i are given by (7.17) and (7.18). Unbiased generalized estimating equations (GEE) for θ are then given by

$$U_1(\theta) = \sum_{i=1}^{m} D_i' V_i^{-1}(n_i - \mu_i) = 0\,, \tag{7.19}$$

where $D_i = \partial \mu_i / \partial \theta'$ (Chen et al., 2005). If values $\phi_{jj'}$ are given, these equations can be solved to obtain an estimate $\widehat{\theta}$. Because $E\{U_1(\theta)\} = 0$ as long as the specified mean functions and rate functions (7.16) are correct, we can just specify fixed values for $\phi_{jj'}$. A common procedure is to take $\phi_{jj'} = 0$ for $j \neq j'$, corresponding to a "working independence" assumption which simplifies the calculation of V_i^{-1}. If we take $\phi_{jj'} = 0$ for all $j, j' = 1, 2, \ldots, J$, then V_i is simply a diagonal matrix with entries $\mu_{ij\ell}$ on the diagonal. When $J = 1$ (7.19) reduces to the Poisson estimating equations used in Section 7.1.2.

If all $\phi_{jj'}$ are allowed to be nonzero and treated as separate unknown parameters, then we can also estimate them using generalized estimating equations. The simplest approach is to use the moment equations

$$\sum_{i=1}^{m} \sum_{\ell=1}^{k_{ij}} \{(n_{ij\ell} - \mu_{ij\ell})^2 - (\mu_{ij\ell} + \phi_{jj} \mu_{ij\ell}^2)\} = 0 \tag{7.20}$$

$$\sum_{i=1}^{m} \sum_{\ell=1}^{k_{ij}} \sum_{\ell'=1}^{k_{ij'}} \{(n_{ij\ell} - \mu_{ij\ell})(n_{ij'\ell'} - \mu_{ij'\ell'}) - \phi_{jj'} \mu_{ij\ell} \mu_{ij'\ell'}\} = 0\,, \tag{7.21}$$

for $j \neq j'$, $j, j' = 1, \ldots, J$. These give closed-form estimates of the ϕ_{jj} and $\phi_{jj'}$. Alternatives that may give more efficient estimates in some settings are to use estimating equations analogous to the one following (3.40); see Lawless and Zhan (1998).

The equations (7.19), (7.20), and (7.21) involve both θ and the parameter vector ϕ containing the $\phi_{jj'}$. As discussed in Section 7.1.2 for the univariate case, an effective procedure is to adopt initial values $\phi_{jj'} = 0$ for all j, j' and to solve (7.19) for $\widehat{\theta}$. Then, estimates of the $\phi_{jj'}$ can be obtained by (7.20) and (7.21) with θ replaced by $\widehat{\theta}$. The procedure may now be repeated, with (7.19) used with V_i based on $\widehat{\phi}$, to yield an updated $\widehat{\theta}$. This process, when iterated, normally converges to give estimates of θ and ϕ. One advantage of this procedure is that an estimate of the covariance matrix of u_i is obtained, which allows an assessment of association among the different event types, as well as extra-Poisson variation in the $n_{ij\ell}$. This is achieved without compromising the robustness of the rate function parameter estimates $\widehat{\theta}$. Estimating function theory establishes that under mild conditions, the asymptotic covariance matrix of $\sqrt{m}(\widehat{\theta} - \theta)$ is estimated consistently by $\widehat{A}^{-1} \widehat{B} \widehat{A}^{-1}$, where

$$\widehat{A} = \frac{1}{m} \sum_{i=1}^{m} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i \quad \text{and} \quad \widehat{B} = \frac{1}{m} \sum_{i=1}^{m} \widehat{D}_i' \widehat{V}_i^{-1} (n_i - \widehat{\mu}_i)(n_i - \widehat{\mu}_i)' \widehat{V}_i^{-1} \widehat{D}_i\,,$$

with the hats indicating that parameters in D_i, V_i, and μ_i are replaced by their estimates. This is true whether or not the variance specifications (7.17) and (7.18) are correct. A final point is that if these variance specifications are satisfactory, the equations (7.19) will produce more efficient estimates asymptotically than simpler equations, say with $\phi_{jj'}$ set equal to zero. However, it

is often difficult to assess the adequacy of these or other variance specifications very precisely, and experience suggests that substantial efficiency gains through the use of detailed modeling of V_i in (7.19) are difficult to realize.

7.1.5 Illustration: Joint Damage in Psoriatic Arthritis

Patients with arthritic conditions are at risk of developing debilitating joint damage and it is common to use the total joint count as a global summary of damage. We consider data from the University of Toronto Psoriatic Arthritis (PsA) Clinic which is comprised of several hundred patients enrolled since 1978. In this clinic damage is assessed through clinical examinations scheduled semiannually, or radiographic examination scheduled biannually. The two methods of damage assessment are directed at different features of damage. Joints are classified as damaged based on clinical examination if there is evidence of deformity or a tendency to flail, whereas radiographic or x-ray examination detects erosions of the bone surfaces forming the joints, joint-space narrowing, or total joint destruction. Here we restrict attention to data as of March 2001, for 250 patients with complete covariate information and at least two assessments. We focus on 64 peripheral joints of the hands, wrists, feet, ankles, knees, elbows, shoulders, as well as temporomandibular joints. The average duration of followup (from clinic entry to last assessment) was 8.2 years.

We consider a bivariate marginal regression model for rate functions of the form (7.16), where events for processes 1 and 2 are joints classified as damaged, based on clinical and radiological assessment, respectively. The time origin in this analysis is the time of clinic entry. Covariates of interest include sex, age of onset of psoriasis (≥ 40 versus < 40 years), age of onset of psoriatic arthritis (≥ 60 versus < 60 years), arthritis pattern (spinal involvement versus peripheral disease alone), and sedimentation rate (ESR; abnormal versus normal). Baseline rates with three constant pieces were specified, where the cutpoints were determined based on the tertiles of the respective assessment times. It should be noted that the total number of possible events of either type is 64 (the number of joints considered) and, strictly speaking, the piecewise-constant rate function model does not reflect this. However, we consider a maximum of 24 years of followup after clinic entry and over this period the model is satisfactory. In the following analysis, the variance and association parameters in (7.17) and (7.18) are also obtained, by solving (7.20) and (7.21) iteratively with (7.19). Starting values are obtained by maximizing (7.5) separately for each event type.

The first few columns of Table 7.2 give the results of fitting models involving all covariates. The estimated regression coefficients and standard errors give a similar picture for the events based on clinical and radiological damage, with the exception that spinal involvement appears to be a significant risk factor for the development of damage by the clinical criteria ($p = 0.009$) but not by radiological criteria ($p = 0.411$). Having an abnormal sedimentation rate

significantly increases the rate of both clinical ($p < 0.001$) and radiological ($p < 0.001$) damage. Both processes exhibit substantial overdispersion relative to a Poisson process, and the two types of event counts show strong positive association. The reduced model in Table 7.2, with three covariates dropped, gives similar estimates for spinal involvement and abnormal sedimentation rate. Figure 7.2 displays the estimated mean functions based on (7.16) for clinical and radiological damage joint counts, for the four groups of patients defined by the covariates in the reduced model.

Table 7.2. Results from fitting rate function and association models to data from the Toronto Psoriatic Arthritis Clinic.

	Full Model				Reduced Model			
	Clinical		Radiological		Clinical		Radiological	
Covariate	EST.	S.E.	EST.	S.E.	EST.	S.E.	EST.	S.E.
Sex								
Male vs. female	0.042	0.199	0.109	0.156	–	–	–	–
Onset of psoriasis								
$\geq$ 40 vs. $<$ 40 yrs	0.179	0.209	-0.038	0.173	–	–	–	–
Onset of PsA								
$\geq$ 60 vs. $<$ 60 yrs	0.548	0.442	0.576	0.317	–	–	–	–
Spinal disease	0.523	0.201	0.127	0.154	0.527	0.194	0.160	0.157
Abnormal ESR	0.636	0.198	0.591	0.160	0.655	0.196	0.581	0.161
Association								
ϕ_j $j = 1, 2$	2.237	0.468	1.673	0.265	2.359	0.460	1.734	0.274
ϕ_3	1.259	0.220	–	–	1.299	0.210	–	–

7.2 Dependent Censoring or Inspections

7.2.1 Dependent Censoring and Weighted Estimating Functions

We now return to the setting where events are observed in continuous time, subject to right censoring. The methods presented to this point have all required that censoring, meaning the end of followup for an individual, be independent of their event process in the sense discussed in Section 2.6. In some settings there is good reason to question the assumption of independent censoring. In this section we explore the independence concept in more detail and consider methods for dealing with dependent censoring by the introduction of weights into estimating functions. A clear distinction should be noted here between censoring and dependent termination, which was discussed in Section 6.6. In particular, following a censoring time the event process is assumed to

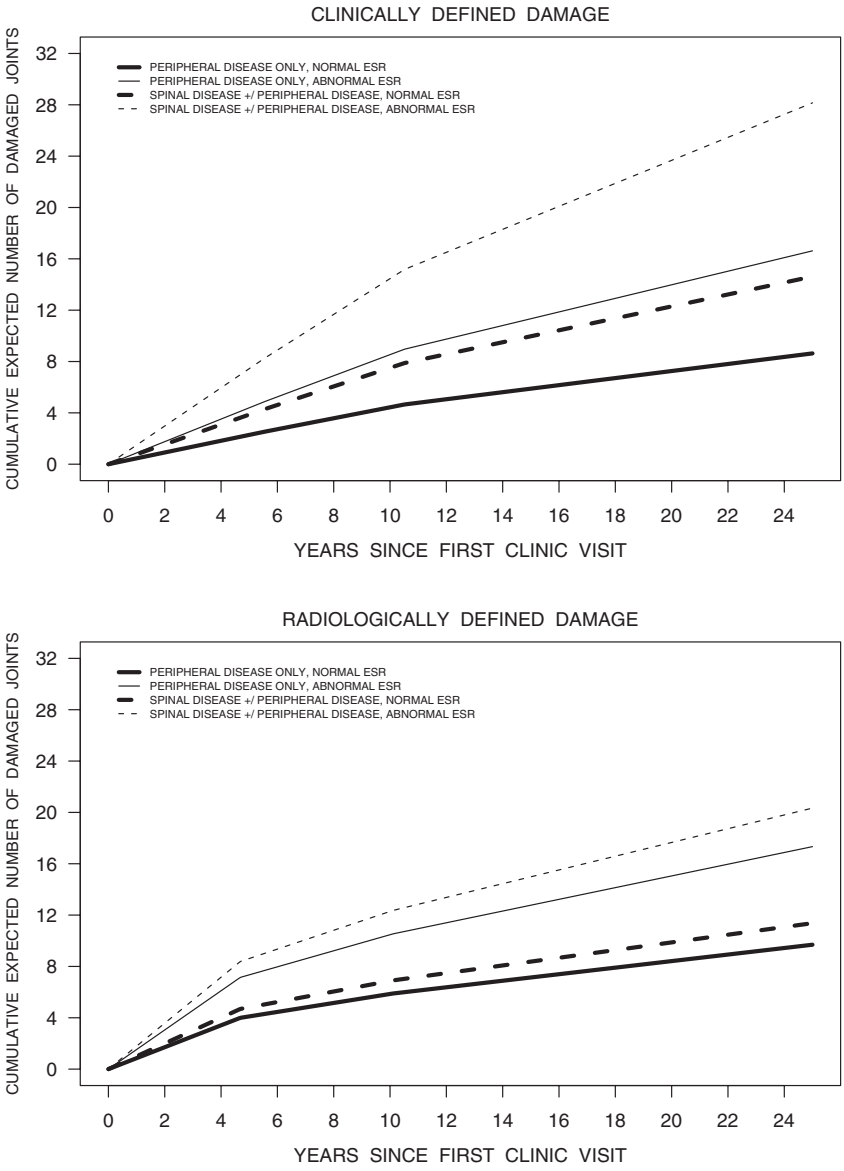


Fig. 7.2. Cumulative mean functions for the number of damaged joints by clinical and radiological assessment based on the reduced model of Table 7.2.

continue. In contrast, at a termination time the event process is known to terminate.

As usual, let $\{N_i(t), 0 \leq t\}$ denote the counting process of events for individual i, and let C_i denote their notional censoring time. To consider different types of models for intensity or rate functions, and the effects of censoring or estimation, we proceed as follows. Let $x_i(t)$ represent external covariates for individual i, and define $x_i^{(t)} = \{x_i(s) : 0 \leq s \leq t\}$. Let $H_i(t) = \{N_i(s) : 0 \leq s < t; x_i^{(t)}\}$ denote the event and covariate history up to time t. It is assumed that

$$E\{dN_i(t)|H_i(t), C_i \geq t\} = E\{dN_i(t)|H_i(t)\}\,, \tag{7.22}$$

where we use notation discussed in Problem 2.15. This is a definition of independent censoring; see (2.54).

Issues concerning dependent censoring arise when we consider intensity or rate function models that do not involve the full event and covariate history $H_i(t)$. Suppose that $z_i(t)$ is a time-varying covariate that may include information from $x_i^{(t)}$ as well as event history up to t, and that our objective is to consider the model

$$E\{dN_i(t)|z_i(t)\} = \lambda(t|z_i(t);\theta)dt\,. \tag{7.23}$$

This covers many of the settings considered in previous chapters. For example, if $z_i(t)$ contains sufficient information that $E\{dN_i(t)|z_i(t)\} = E\{dN_i(t)|H_i(t)\}$ then we are in the realm of intensity-based models. If, on the other hand, $z_i(t)$ contains no event history, we are in the realm of the rate function models of Sections 3.6 and 6.4.

The methods in this book are based on maximum likelihood for fully specified models, or on estimating functions that are usually maximum likelihood score equations under some model. When exact event times are available over the period of followup and we do not want to rely on the correctness of a full process model, but merely wish to consider the model (7.23), we base estimation of θ in (7.23) on estimating functions of the form

$$U_i(\theta) = \int_0^\infty Y_i(t)a_i(t;\theta)\{dN_i(t) - \lambda(t|z_i(t);\theta)dt\}\,, \tag{7.24}$$

where $Y_i(t) = I(C_i \geq t)$ and $a_i(t;\theta)$ is a function that may depend on $z_i(t)$. For example, compare the estimating functions (3.4) and (3.5) and the general form of the likelihood (5.4). In order for (7.24) to have expectation zero and thus for $U(\theta) = 0$ to provide a consistent estimator $\widehat{\theta}$, we require that the integrand have expectation zero at each time t. A sufficient condition is that

$$E\{dN_i(t)|z_i(t), C_i \geq t\} = E\{dN_i(t)|z_i(t)\} = \lambda(t|z_i(t);\theta)dt\,. \tag{7.25}$$

This is equivalent to saying that $Y_i(t)$ is conditionally independent of $dN_i(t)$, given $z_i(t)$.

When $z_i(t)$ contains sufficient information from $H_i(t)$ and (7.22) holds, (7.25) is plausible. However, when we consider rate function models for which $z_i(t)$ does not contain any previous event history, (7.25) may well be violated; it would be, for example, if censoring at t is not independent of previous event history, given $z_i(t)$. An extreme case occurs when $z_i(t)$ is empty, and we consider the marginal rate functions $\lambda(t;\theta)dt = \rho(t)dt = E\{dN_i(t)\}$ discussed in Section 3.6.1. In that case censoring has to be completely independent of event history. This would be violated if, for example, censoring were related to an omitted covariate that also affected the event process. It is straightforward to assess whether censoring is related to prior event history by modeling the censoring times in terms of event and covariate history.

If the independent censoring assumption does not seem plausible for (7.24), we can use the idea of inverse probability of censoring (IPC) weights discussed in Section 4.4 to adjust the estimating function. This requires the following modeling assumptions concerning censoring. Assume that given $C_i \geq t$ and $H_i(t)$, the event $C_i < t + \Delta t$ is independent of $H_i(s)$ for $s > t$. We further assume that the hazard function for C_i is given by

$$\begin{aligned}\lambda_c(t|H_i(t)) &= \lim_{\triangle t \downarrow 0} \frac{\Pr(C_i < t + \triangle t | C_i \geq t, H_i(t))}{\triangle t} \\ &= \lambda_c(t|v_i(t))\,, \end{aligned} \tag{7.26}$$

where $v_i(t)$ is a specified time-varying covariate vector containing information from $H_i(t)$. Define the product integral

$$G_i(t) = \prod_{[0,t)} \{1 - d\Lambda_c(s|v_i(s))\}\,, \tag{7.27}$$

where $d\Lambda_c(s|v_i(s)) = \lambda_c(s|v_i(s))ds$ when $\lambda_c(s|v_i(s))$ is continuous. The expression (7.27) also allows discrete models in which there is a positive probability of censoring occurring at certain times s, and then $d\Lambda_c(s|v_i(s))$ equals $\Pr(C_i = s|C_i \geq s, H_i(s))$ at such time points.

We now amend the estimating function (7.24) to

$$U_i^W(\theta) = \int_0^\infty \frac{Y_i(t)a_i(t;\theta)}{G_i(t)} \{dN_i(t) - \lambda(t|z_i(t);\theta)dt\}\,. \tag{7.28}$$

Note first that by (7.26) and the independent censoring assumption (7.22), we have $G_i(t) = \Pr(C_i \geq t|H_i(t))$. Then, by evaluating the expectation of the integrand in (7.28) at t, conditional on $z_i(t)$, as the iterated expectation $E_{H_i(t)}[E_{Y_i(t)|H_i(t)}(\cdot)]$, we find the expectation to be $E[a_i(t;\theta)E\{dN_i(t) - \lambda(t|z_i(t);\theta)dt|z_i(t)\}] = 0$ and so (7.28) is unbiased.

The unbiasedness of (7.28) depends on the validity of the censoring model expressed in (7.26) and (7.27). In practice, we have to estimate $\lambda_c(t|v_i(t))$ and $G_i(t)$, and replace $G_i(t)$ with its estimate $\widehat{G}_i(t)$. We can, for example, use a semiparametric proportional hazards model, although other models are also useful. The following section illustrates the procedure.

7.2.2 Rate or Mean Function Estimation with Event-Dependent Censoring

Suppose we aim to estimate a marginal mean function $\mu(t)$ independent of any covariates. Thus, $\mu(t) = E\{N_i(t)\}$ for each $i = 1, \ldots, m$, as in Section 3.6.1, and the Nelson–Aalen estimate (3.17) given there is valid provided censoring times C_i are completely independent of the event processes. Suppose, however, that C_i was thought to be related to the number of events, and that this can be expressed in a model (7.26) for which $v_i(t) = N_i(t^-)$. The estimate (3.17) can be viewed as arising from estimating functions of the form (7.24), where in heuristic terms, θ is associated with increments $d\mu(t)$ across different values of t, $\lambda(t;\theta)dt = d\mu(t)$, and $a_i(s;\theta) = I(t = s)/d\mu(s)$ gives the estimating function for $d\mu(s)$. Amending this to (7.28), we consider the estimating equations

$$\sum_{i=1}^{m} \frac{Y_i(s)}{\widehat{G}_i(s)}\{dN_i(s) - d\mu(s)\} = 0 \qquad s > 0\,,$$

which yield the estimates

$$d\widehat{\mu}(s) = \sum_{i=1}^{m} \frac{Y_i(s)dN_i(s)}{\widehat{G}_i(s)} \Big/ \sum_{i=1}^{m} \frac{Y_i(s)}{\widehat{G}_i(s)} \tag{7.29}$$

and the corresponding estimate

$$\widehat{\mu}(t) = \int_0^t d\widehat{\mu}(s)\,. \tag{7.30}$$

In (7.29) and (7.30) it is assumed that $Y_i(s) > 0$ for at least one individual at each $s \leq t$, in which case all $\widehat{G}_i(s) > 0$. When censoring is completely independent of the event process, it is preferable to use the Nelson–Aalen estimate (7.17) over (7.30), because it is usually more efficient. When this is not the case, (7.30) with a plausible model for $G_i(s)$ is preferred because the Nelson–Aalen estimate may be biased.

Asymptotic properties for $\widehat{\mu}(t)$ may be developed under assumptions that ensure the number of individuals under observation at each time $s \leq t$ becomes arbitrarily large in probability. Variance estimates and confidence intervals or bands for $\mu(t)$ are most simply obtained using the nonparametric bootstrap.

In the absence of any covariates, we can estimate the censoring process nonparametrically. Let

$$\lambda_{cj}(t) = \lim_{\triangle t \downarrow 0} \frac{\Pr\{C_i < t + \triangle t | C_i \geq t, N_i(t^-) = j\}}{\triangle t}$$

and $\Lambda_{cj}(t) = \int_0^t \lambda_{cj}(s)ds$. Then $\Lambda_{cj}(t)$ is estimated nonparametrically as for the cumulative event intensity (5.19) in Section 5.3.2, with

$$d\widehat{\Lambda}_{cj}(t) = \frac{\sum_{i=1}^{m} I(C_i = t, N_i(t^-) = j)}{\sum_{i=1}^{m} I(C_i \geq t, N_i(t^-) = j)}\,. \tag{7.31}$$

This is effective if the censoring hazard function depends only on the number of prior events. If it were thought to depend on other features of the event process, such as the time since the latest event, other models (7.26) could be formulated.

An alternative approach to estimation of $\mu(t)$ when censoring is event-dependent is to use the multistate modeling methodology of Section 5.3. When the event process is Markov as described there, and the censoring hazard depends only on the number of previous events, the Markov estimates for $\Pr\{N(t) = r\}$ given by (5.20) can be used, and the corresponding estimate for $\mu(t)$ is obtained from (5.18).

The two estimates based on (7.30) and (5.18) epitomize the two main approaches to estimation of a marginal process feature $\psi(t)$ when censoring may be dependent on event or covariate history. One is to adopt a simple model for $\psi(t)$ and to adjust estimating functions for $\psi(t)$ by the inclusion of inverse probability of censoring weights designed to make the estimating functions unbiased. The other approach is to model the event process intensity in sufficient detail that censoring is conditionally independent, and then to estimate $\psi(t)$ as a (perhaps complicated) function of the estimated intensity. The estimate based on expression (5.18) for $\mu(t)$ exemplifies the latter approach. Either approach can be complicated, depending on the amount of detail about the process history that needs to be considered in modeling the intensity or the censoring hazard function. The IPCW methods are generally simpler, and a model for censoring is usually easier to check than a model for the event process. They may, however, be less efficient than methods based on an appropriate model. Guidelines on which framework gives more efficient estimators are currently not available.

Other marginal features that have been discussed in previous chapters include $\Pr\{N_i(t) = r\}$ and the distribution $\Pr\{W_{ij} \geq w\}$ of the jth gap time. For the latter, IPC weights were introduced in Section 4.4.1 to deal with the dependent censoring created for second and subsequent gap times when gap times are correlated. The method of estimating $p_{0k}(t) = \Pr\{N_i(t) = k | N_i(0) = 0\}$ based on Markov models in Section 5.3.2 is valid if both the event and censoring processes are Markov, as described there and illustrated in Figure 5.7. Other estimates may be devised with IPC weights, such as the weighted prevalence estimate

$$\widehat{p}_{0k}(t) = \sum_{i=1}^{m} \frac{Y_i(t) I(N_i(t) = k)}{\widehat{G}_i(t)} \Big/ \sum_{i=1}^{m} \frac{Y_i(t)}{\widehat{G}_i(t)}\,, \tag{7.32}$$

where it is assumed that at least one $\widehat{G}_i(t)$ is > 0. Datta and Satten (2002) and Cook et al. (2003) discuss various estimators and provide references. Problem 7.6 gives one such method.

7.2.3 Intermittent Observation

When individuals are observed intermittently as in Section 7.1, the conditions (i) and (ii) that follow (7.1) are generally assumed in order to draw inferences on the event process. Condition (ii) is more crucial and also might be violated in some settings, for example, if the time b_{ij} an individual is inspected is related to the number or pattern of events after $b_{i,j-1}$. Another concern in observational studies is that an individual who is following a schedule of periodic inspections or followup visits may fail to appear for an inspection and then not be seen for a considerable length of time. Sometimes individuals may be truly lost to followup, but sometimes they may have simply chosen to skip inspections because they were not experiencing any events. When an individual has not been seen for a long time before an administrative end-of-followup time, a decision has to be made about the definition of C_i, the duration of followup.

An example of this occurs in the study of cerebrospinal fluid shunt failures for pediatric patients, discussed in Section 6.7.1. In this case patients were to return to a hospital clinic for observation roughly every six months or year. The first shunts were received between January, 1987 and December, 1996, and the administrative endpoint for the database was the end of 1997. Therefore, the potential durations of followup range from one to eleven years, although some patients died before the administrative end of followup. Figure 7.3 is a plot of the final inspection times for each patient, against the date of receipt of their first shunt. The final inspection times are shown as time (in days) from receipt of the first shunt, or time in study. The final inspection time may correspond to an inspection, a shunt replacement, or death; dots indicate a live inspection and the plus symbols denote a time of death. If patients were really returning to the clinic for annual inspections, then the dots in Figure 7.3 should all lie within a band of width roughly one year along the line denoting potential duration of followup. This is not so, and we see that many individuals were last seen a considerable length of time before the administrative end-of-followup. The question then arises as to how to define the duration of followup C_i. In the analyses of Section 6.7.1 and previously, we used $C_i = D - \tau_{i0}$ where D is December 31, 1997 and τ_{i0} is the date of receipt of the first shunt for individual i. This was done on the assumption that persons who did not return for regular inspections had not experienced any shunt failures. It is possible that some patients may have moved far enough away that a shunt failure would have been dealt with at another hospital, but this was judged to be a rare occurrence. Another possibility would have been to define C_i as $\tau_i - \tau_{i0}$, where τ_i is the date of the last inspection for individual i. This was not chosen because τ_i was not judged to be the true end of followup for the individual. In fact, it would be a dependent censoring time (i.e. not a stopping time) and introduce bias, if it was determined because no future events occurred.

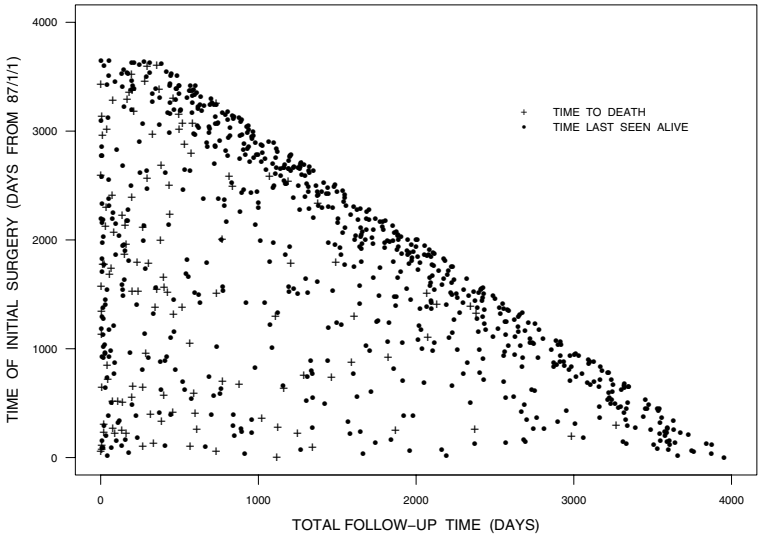


Fig. 7.3. Final inspection times versus date of first shunt, for pediatric patients with cerebrospinal fluid shunts.

If there are no losses to followup prior to an administrative censoring time, and it is possible to ascertain retrospectively the event history ΔH_{ij} over $(b_{i,j-1}, b_{ij}]$ at the time b_{ij}, then no problems arise if an intensity-based model is used, because event-dependent choice of b_{ij} is ignorable. The case where only the number of events n_{ij} is observed is, however, problematic. There is then no way to assess the validity of condition (ii) following (7.1), without further assumptions that are essentially uncheckable, or data where some individuals' inspection times are prespecified. In particular, even though we can model $\Pr(b_{ij}, n_{ij}|\tilde{H}_{i,j-1})$ or $\Pr(n_{ij}|b_{ij}, \tilde{H}_{i,j-1})$, where here $\tilde{H}_{i,j-1}$ contains the history of events, covariates, and inspection times up to $b_{i,j-1}$, there is no way to assess whether $\Pr(n_{ij}|b_{ij}, \tilde{H}_{i,j-1})$ satisfies condition (ii).

Models in which the censoring or inspection process shares a common random effect with the event process have been suggested for settings involving event-dependent censoring or inspection. For example, we might assume that conditional on an unobservable random effect u_i, the recurrent event process $\{N_i(t), 0 \leq t\}$ is Poisson with rate function $u_i \rho_i(t)$, that the inspection times follow a renewal process with hazard function $u_i^{\alpha} h_i(w)$ for the gaps between inspections, and that the inspection and event processes are conditionally independent. The joint density function of the b_{ij} and n_{ij} is then

$$\int_0^\infty \left\{ \prod_{j=1}^{k_i} e^{-u_i \mu_{ij}} \frac{(u_i \mu_{ij})^{n_{ij}}}{n_{ij!}} \right\} \left\{ \prod_{j=1}^{k_{ij}} u_i^\alpha h_i(w_{ij}) e^{-u_i^\alpha H_i(w_{ij})} \right\} dG(u_i), \quad (7.33)$$

where $\mu_{ij} = \int_{b_{i,j-1}}^{b_{ij}} \rho_i(t)dt$, $w_{ij} = b_{ij} - b_{i,j-1}$, $G(u)$ is the distribution function of u_i, and it is assumed that followup finishes at b_{ik_i}, which is a stopping time. Independence of the event and inspection process, conditional only on covariates, could be assessed by considering whether α is zero. It is, however, impossible to assess the model giving (7.33) without further assumptions that are uncheckable with the data given.

Shared random effects models such as (7.33) are often unappealing because of one key feature: the intensity function for the inspection process at time t depends on event history after t. In particular, consider the inspection time process for individual i as a point process with intensity function $\lambda_I(t|H_i^c(t))$, where $H_i^c(t)$ is the complete history of previous inspection times and recurrent events. Integration with respect to the random effect u_i in models such as (7.33) produces a result where $\Pr\{\text{inspection in } (t, t+\Delta t)|\tilde{H}_i(t), N_i(s), s \geq t\}$ depends on $N_i(s), s \geq t$ and so is not of the form $\lambda_I(t|\tilde{H}_i(t))$. In many settings this goes against our understanding of the inspection process. This type of situation is specifically excluded in the method of dealing with dependent censoring based on inverse probability of censoring weights; see (7.26) and the preceding remarks.

In practice, we usually try to render the inspection and event processes conditionally independent by building sufficient event history and covariate effects into our models for n_{ij} $(j = 1, 2, \ldots)$, given b_{ij} and $\tilde{H}_{i,j-1}$. Sometimes, however, it seems plausible that in spite of this, inspection times may be event-dependent. Without obtaining full event history data on some individuals, or having some individuals with prespecified inspection times, it is difficult to confirm this or assess potential effects. In such cases we should be cautious in inferring properties of the pure event process, based on our models for $\Pr(n_{ij}|b_{ij}, \tilde{H}_{i,j-1})$.

Loss to followup is potentially more serious than moderate event-dependent inspection times, especially in settings described above where individuals can become lost to followup long before an administrative end-of-followup time. In particular, suppose that individual i has potential followup time C_i, determined by an administrative end-of-study date, but that she is last seen at inspection time b_{ik_i} which is smaller than C_i. Normally we might take b_{ik_i} as the end-of-followup time, but if the inspection process is event-dependent and C_i is much larger than b_{ik_i}, this may bias an analysis. If we are able to trace some individuals lost to followup and determine their event histories, then we can assess whether loss to followup is event-dependent. If this cannot be done it is prudent to run analyses with alternate choices for end of followup.

Finally, if we wish to consider parametric models for marginal features such as $\mu(t) = E\{N_i(t)\}$, we can utilize the first approach discussed in Section 7.2.2, that is, marginalization from an intensity-based model. IPCW methods

can also be developed by treating the inspection times as a point process and modeling the inspection intensity (Lin et al., 2004). Inverse weights are based on the estimated intensity, so some form of smoothing is needed to provide an estimate at all event times. Given the variability in such estimates and the possibility of misspecification, it is not yet clear how this approach performs.

7.3 Event-Dependent Selection

In some studies the inclusion of an individual is dependent on her having experienced some outcome. If this outcome is related to an event history process which is a focus of the study, then this should be accounted for in the analysis of the process. Let us phrase this more specifically, as follows. Suppose that the population from which study individuals are drawn has M members with event history processes $N_i^{(\infty)} = \{N_i(t), 0 \leq t\}$, $i = 1, \ldots, M$. Let $S \subset \{1, 2 \ldots, M\}$ denote the set of individuals sampled, or selected for the study. If S is selected independently of $\{N_i^{(\infty)}, i = 1, \ldots, M\}$ then the sampling plan is ignorable, and inferences can be based on the probability of the observed event histories of the individuals in S. For settings where this is not the case, and selection is event-dependent, two situations can be distinguished. The first is where the events on which selection is based are part of the initial conditions for the process rather than the "response". No problems arise if we condition on such events in our analyses. The second situation is where S depends on $\{N_i^{(\infty)}, i = 1, \ldots, M\}$ given covariates and initial conditions. In that case also, we must consider the selection conditions in developing an analysis. We discuss event-dependent selection via a number of examples.

7.3.1 Some Examples

Event-dependent selection can arise in various ways, but in most cases either (a) the individual must have experienced some event or prior events over a time period, or (b) the individual must not have experienced a terminating event prior to a given time. These conditions often occur in retrospective studies where an individual is selected after at least a portion of her event history has occurred, but they may also arise in prospective studies. The following examples illustrate this, and indicate how the selection process may be handled.

Example 7.3.1 Selection of patients for a psychiatric study

In a study on hospitalizations for psychiatric disorders (Kessing et al.,1999), retrospective data were available on individuals who had experienced an initial event (hospitalization and discharge with a specific diagnosis) during a

previous time period. If we consider the first hospital discharge to be an initiating event that defines a time origin following which later events may occur, then the study population consists of persons who experience a first psychiatric admission and subsequent discharge, and there is no selection effect if we include time and details of the admission and discharge as covariates. If, however, we wish to assess the rate of psychiatric admissions in the general population, then there is a selection effect. To simplify the discussion, consider just the admission events and not the discharges from hospital. In the former case, the likelihood function for an individual with $n \geq 0$ subsequent admissions at ages $t_1 < \cdots < t_n$ following a first admission at age τ_0, would be

$$\prod_{j=1}^{n} \lambda\left(t_j | H(t_j)\right) \exp\left\{-\int_{\tau_0}^{\tau_1} \lambda\left(t | H(t)\right) dt\right\}, \tag{7.34}$$

where the individual is followed to age τ_1, and where $\lambda(t|H(t))$ is the intensity function for an admission at age t. The history $H(t)$ includes information on time in hospital and discharges; $\lambda(t|H(t))$ equals zero while an individual is in hospital. In the latter case, the likelihood function has to be conditional on the individual having a first admission by age τ_1. This gives

$$\frac{\lambda_0(\tau_0) \exp\left\{-\Lambda_0(\tau_0)\right\}}{1 - \exp\left\{-\Lambda_0(\tau_1)\right\}} \prod_{j=1}^{n} \lambda\left(t_j | H(t_j)\right) \exp\left\{-\int_{\tau_0}^{\tau_1} \lambda\left(t | H(t)\right) dt\right\}, \tag{7.35}$$

where we assume that first admissions follow a Poisson process, and use $\lambda_0(t)$ to denote the intensity for a first admission at age t. The second term in (7.35) is as in (7.34), but the term involving $\lambda_0(t)$ is the conditional probability of a first admission at age τ_0, given that it must have occurred before age τ_1.

Other selection conditions may also apply in studies of this type. For example, an individual may need to be alive at some point in time in order to be included in the sample; if death is not independent of a person's psychiatric history, there is a selection effect. This is considered below in Example 7.3.4. In another study (Kessing et al., 2004), individuals had to experience a discharge from hospital during a specific time period; retrospective data on hospitalization were then collected for these individuals. This is a rather complicated type of selection, and it is difficult to specify models for which the likelihood function is easy to handle. In particular, let $Y(t)$ denote the process of admissions and discharges for an individual and note that the conditional density

$$\Pr(Y(t), \tau_0 < t < \tau_1 | A),$$

where $A = \{\text{discharge during } (\tau_1 - \Delta, \tau_1), \text{ initial discharge at } \tau_0\}$, gives the likelihood function. This is very complicated for processes such as the two-state processes of Section 6.5, which describe repeated hospitalizations and discharges.

Example 7.3.2 Entry conditions in clinical trials

Clinical trials frequently feature selection criteria which are designed to optimize information regarding treatment effects. This is often achieved by restricting attention to patients with features putting them at higher than average risk of events. It is increasingly common, for example, for entry criteria to stipulate that subjects must have experienced some minimum number of events in a period of time prior to study entry.

The Asymptomatic Cardiac Ischemia Pilot study (ACIP Investigators, 1992) is an example of such a trial. The purpose of the study was to evaluate a treatment effect on the number of episodes of transient myocardial ischemia (TMI) detected over a 48-hour heart monitoring period by ambulatory electrocardiogram (ECG). Patients were first monitored for 48 hours during a baseline period of assessment and in order to avoid recruiting patients with particularly low rates of TMI, the investigators explored the impact of introducing a selection threshold C, such that only subjects with a baseline count above C would be eligible for the study.

Let N_{i1} denote the baseline count for subjects screened for selection, labeled $i = 1, 2, \ldots, m + m_2$, and N_{i2} denote the count following the randomization to treatment among eligible subjects. Without loss of generality we label the randomized subjects $i = 1, 2, \ldots, m$ and let x_i denote the treatment received.

Mixed Poisson models with patient-specific random effects provide a convenient framework for analysis. They give a joint distribution of N_{i1} and N_{i2} for study subjects, and within this framework there are several likelihoods that one could construct. If the baseline counts from subjects not satisfying the screening requirements are not available, then a likelihood can be based on the joint distribution of the available (truncated) data,

$$L(\theta) \propto \prod_{i=1}^{m} \Pr\left(N_{i1} = n_{i1}, N_{i2} = n_{i2} | N_{i1} \geq C, x_i; \theta\right).$$

If the underlying assumption is that $N_{ij}|u_i \sim \text{Poisson}(u_i \mu_j \exp(I(j = 2)x_i\beta))$, where u_i is gamma distributed with mean one and variance ϕ, then $L(\theta)$ has a closed form and can be maximized using general-purpose software for optimization. If the data from all screened subjects are available, a likelihood based on all patients would be

$$\prod_{i=1}^{m+m_2} \Pr(N_{i1} = n_{i1}) \prod_{i=1}^{m} \Pr(N_{i2} = n_{i2} | N_{i1} = n_{i1}, x_i).$$

In terms of estimating the treatment coefficient β, one could just maximize the second term in this expression, but the first term does contain information on ϕ and the baseline means. Similar considerations apply when the event times, rather than just the counts, are available in the followup periods.

In some cases, selection criteria based on prior event counts are more complex. In an epilepsy study reported in Fuchs et al. (1994), patients were screened for two successive one-month periods and deemed eligible for recruitment if they had at least four seizures in one period and at least one seizure in the other period. If N_{i1A} and N_{i1B} denote the counts for the first and second screening periods and $N_{i1} = (N_{i1A}, N_{i1B})'$, then the likelihood based on the full data is proportional to

$$\prod_{i=1}^{m} \Pr\left(N_{i1A} = n_{i1A}, N_{i1B} = n_{i1B}, N_{i2} = n_{i2} | N_{i1} \in \mathcal{C}, x_i; \theta\right) ,$$

where $\mathcal{C} = \{(n_{i1A}, n_{i1B}) | n_{i1A} \geq 4, n_{i1B} \geq 1 \quad \text{or} \quad n_{i1A} \geq 1, n_{i1B} \geq 4\}$. For the purposes of estimating covariate effects on the mean of N_{i2}, the model for $\Pr(N_{i2} = n_{i2} | N_{i1} = n_{i1})$ can also be used, but the likelihood based on all screened subjects is simpler to work with if the data are available.

Model and likelihood formulation for more general processes with selection criteria can be challenging when random effects are used to address heterogeneity and to induce association in the event data for the screening and followup periods. This was discussed in Section 4.5.2 in the context of gap time analyses. Let $(t_{i1}, \ldots, t_{in_i}, n_i)$ represent the data (total event count and event times) observed over $[0, \tau_i]$, the treatment period for subject i, and $H_i(-\tau_0, 0)$ the data during a common screening period $[-\tau_0, 0)$. If conditions are imposed on $H_i(-\tau_0, 0)$ for entry to the study then the distribution for the prospective data is

$$\int \Pr(t_{i1}, \ldots, t_{in_i}, n_i | u_i, H_i(-\tau_0, 0)) dG(u_i | H_i(-\tau_0, 0)) ,$$

where $G(u_i | H_i(-\tau_0, 0))$ is the distribution function for the random effect given the prior history. For the mixed Poisson models discussed earlier the required conditional distribution $dG(u_i | H_i(-\tau_0, 0))$ is easy to obtain, but this is not the case for other types of models. Intensity-based models without random effects are preferable when Poisson models are not appropriate; information in $[-\tau_0, 0)$ is then included in the initial conditions $H_i(0)$ for each individual, which we condition on in forming likelihoods.

Example 7.3.3 Failure information from product warranty claims

Warranty claims databases provide information about the occurrence of certain types of failures or problems for manufactured products. Once a product unit is sold, we let t denote the "age" of the unit (time since it was sold) and let $N_i(t)$ denote the number of warranty claims (perhaps of a particular type) up to age t. Products are usually covered by a warranty for some fixed period after the date of sale. For some products there are also other conditions, such as distance limits in the case of motor vehicles. When a warranty claim is made,

the date of sale is determined and other information is collected. However, for most types of products, the data on sale and other information is missing for units that have not experienced a claim. This generates event-dependent selection effects in terms of the data available for analysis.

Suppose that a database contains information on all claims up to calendar time X. If unit i was sold at time $x_i \leq X$, then there is a censoring time $\tau_i = \min(X - x_i, w_i)$ on the age scale, where w_i is the age limit on warranty coverage for that unit. With products such as appliances, w_i is usually fixed, for example, one or two years, but for some products, it may vary across units. A car may have warranty coverage for either three years or 60,000 km, for example, so that some cars are covered for less than three years. It is very common that either w_i or the date of sale x_i is unknown for most product units until a warranty claim is made. For example, manufacturers generally do not know the date of sale of most product units, unless notified at the time of a warranty claim. This means that at calendar time X, the followup time τ_i for many units is unknown.

To illustrate the effect of this, suppose that claims for unit i follow a Poisson process with intensity $\lambda_i(t)$; let $[0, \tau_i]$ be the followup period for the unit and let z_i represent fixed covariates. If τ_i and z_i are unknown until the first claim occurs, at which time x_i, w_i (and τ_i), and z_i are determined, then the observed data are event-dependent in the sense that they depend on whether $N_i(\tau_i) \geq 1$. Thus, if units $i = 1, \ldots, m$ had $n_i \geq 1$ claims, at times t_{ij} $(j = 1, \ldots, n_i)$, the likelihood function for the parameters θ specifying the $\lambda_i(t)$ is

$$L(\theta) = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda_i(t_{ij}) \right\} \left\{ \frac{e^{-\Lambda_i(\tau_i)}}{1 - e^{-\Lambda_i(\tau_i)}} \right\}. \tag{7.36}$$

This is similar to (7.35), and is sometimes called a zero-truncated distribution or likelihood.

If the manufacturer knew that an additional m_1 units $i = m+1, \ldots, m+m_1$ had been sold, and if their dates of sale and covariates were known, then $L(\theta)$ could be supplemented by the information that these units had not experienced a claim. This would give the likelihood

$$\begin{aligned} L_1(\theta) &= \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda_i(t_{ij}) \right\} e^{-\Lambda_i(\tau_i)} \cdot \prod_{i=m+1}^{m+m_1} e^{-\Lambda_i(\tau_i)} \\ &= L(\theta) \left\{ \prod_{i=1}^{m} \left(1 - e^{-\Lambda_i(\tau_i)}\right) \right\} \left\{ \prod_{i=m+1}^{m+m_1} e^{-\Lambda_i(\tau_i)} \right\}. \end{aligned} \tag{7.37}$$

However, if τ_i $(i = m+1, \ldots, m+m_1)$ and any necessary covariates are missing, then (7.37) is not available. This may be viewed as an event-dependent missing data problem; we return to it in Section 7.3.2.

If a full probability model for the process under study is specified, then when selection or observation is event-dependent, we simply have to incorporate the selection condition in the probability calculation giving the likelihood function, as in (7.35) and (7.36). It is difficult to avoid probability models for the event process here. For example, if we wished only to estimate a common mean function $\mu(t) = E\{N_i(t)\}$, we might seek an unbiased estimating function with weights based on inverse probability of selection, analogous to those in Section 7.2. This does not work in the setting described above, however: given $N_i^{(\infty)}$, the probability the unit is selected is either 0 or 1. Section 7.3.2 considers settings where some auxiliary information may be obtained, in which case more options are available.

Example 7.3.4 Studies with selection dependent on survival

Retrospective studies, in which past data are collected at some point in time, often have selection conditions that are in some way event-dependent; Example 7.3.1 is one such situation. A condition that is often involved is that an individual must be alive and at risk of events at the time of selection. In particular, suppose that an individual is selected at calendar time R and past information is collected on an event history process $\{N(t), t \leq R\}$ for that individual, only if the individual has not experienced a terminating event prior to R. In the case of psychiatric admissions to hospital discussed in Example 7.3.1, this would mean that the individual had to be alive at the time of selection.

Such a condition is a potential source of bias, and must not be ignored if the terminating event is associated with the recurrent events under study, as in Section 6.6. Probability models such as those considered in Sections 6.6.1, 6.6.2, and 6.6.4 can be used in such settings, but the likelihood functions tend to be complicated. In the terminology of Section 6.6, with T representing the time of the terminating event, an individual selected at time R has the conditional probability density $\Pr(N^{(R)}|T \geq R)$ as the basis for the likelihood function, where $N^{(R)} = \{N(t), t \leq R\}$. This is difficult to calculate for models with general intensities, such as those in Section 6.6.1, but can be obtained for Markov models using the multistate model formulation of Section 6.6.4. As in the setting of Example 7.3.3, it is difficult to avoid full probability models here, even when features such as $E\{dN_i(t)|T_i \geq t\}$ or $E\{N_i(t)\}$ might be the main interest.

In order to be recruited into prospective clinical trials, subjects must naturally survive to the time of accrual. Depending on the scientific question at hand, it may be important to incorporate this selection criterion into the analysis. In Sections 6.7.3 and 6.7.4 we discussed issues in the analysis of data from a trial of patients with bone metastases. Figure 6.2 contains a multistate model which reflects the possible course of skeletal complications and death following the development of bone metastases. The process begins with the

occurrence of bone metastases whereupon subjects begin in state 0, representing no skeletal complications. Progressions to more advanced states happen with the occurrence of each skeletal event or death.

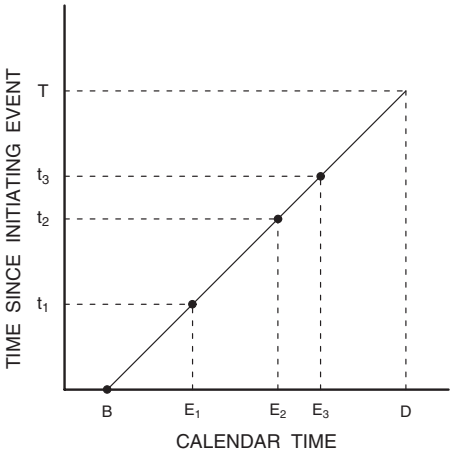


Fig. 7.4. Lexis diagram of event occurrence and death for a hypothetical subject.

Consider the Lexis diagram in Figure 7.4 for a hypothetical subject. The horizontal axis represents calendar time and the vertical axis indicates the time measured from the origin of the event process (B). For this subject, the bone metastases occurred at calendar time B and skeletal events were experienced at calendar times E_1, E_2, and E_3 respectively, followed by death at calendar time D. Here $t_k = E_k - B$ is the time of the kth event and $T = D - B$ is the time of death measured from the development of bone metastases.

Complications arising from selection criteria are often ignored when clinical trial data are analyzed, and simple treatment comparisons are of primary interest. This may be permissible under random treatment allocation, where the initial conditions may be considered balanced across treatment groups. If the aim, however, is to use data from patients in a placebo control arm of a clinical trial to make statements about, say, the patterns or expected number of lifetime skeletal complications in patients with bone metastases, effects due to selection criteria must be taken into account.

The accrual of patients in prospective studies typically takes place over a period of time, but for simplicity here we consider a study with a single accrual date which we denote R in calendar time. Figure 7.5 is a Lexis dia-

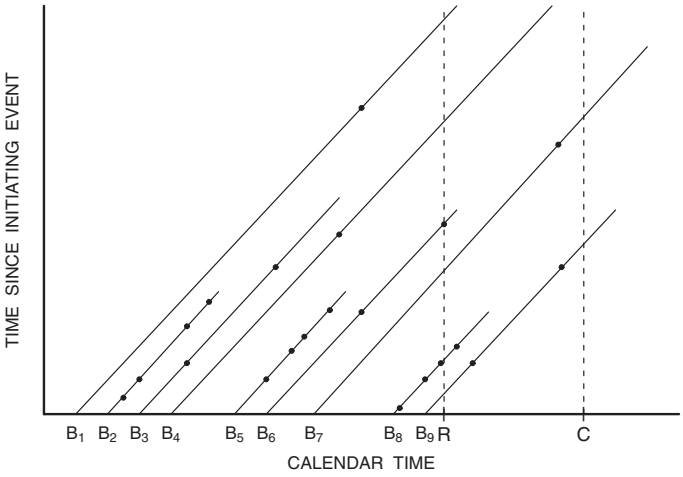


Fig. 7.5. Lexis diagram of initiating event, recurrent events, and death for a hypothetical sample of subjects.

gram for a hypothetical population of nine subjects, where we indicate the date of the first bone metastasis for subject i by B_i and we omit the times of the recurrent events $(E_{i1}, E_{i2}, \ldots)$ and death (D_i) on the horizontal axis. We consider the ith individual's event process to start at the time B_i of her metastases and denote the times of the recurrent events with dots on the 45-degree lines, which terminate upon death. For illustration, we suppose in Figure 7.5 that subjects with more skeletal events tend to die sooner. Only the subjects whose timelines cross the vertical line at R are alive at this recruitment time and eligible for prospective followup, so subjects 2, 3, and 5 would be excluded from the study. Furthermore, suppose the study was completed at calendar time C so that event and survival times are administratively censored for an arbitrary subject at $\tau = C - B$. In the absence of covariates, we then prospectively observe for an arbitrary eligible subject the data $(\{N(s), R - B \le s \le X\}, X = \min(C, D) - B = \min(\tau, T), \delta = I(T \le \tau))$. The probability model for the prospective data observed from the subjects satisfying the selection conditions is then

$$\Pr(\{N_i(s), R - B_i \le s \le X_i\}, X_i, \delta_i | D_i > R - B_i)\,.$$

Sometimes information may be retrospectively collected on $\{N_i(s), 0 \le s < R - B_i\}$ but this may be less reliable than the prospectively collected data. Some studies may further restrict attention to patients with documented skeletal events because these patients are often at greater risk of future skeletal

events. In this case there is the added condition that $N_i(R - B_i) \geq 1$ for patients to be included.

7.3.2 Supplementary Information on Selection

Sometimes there is supplementary information on individuals in the population who are not selected for full followup. When this occurs, the additional data can often be used to enhance inference about the event process, and it may allow simpler procedures than those based on the complicated likelihood from the selected sample alone. We consider two ways in which this may be done, in settings where it is possible to collect information on the event histories of some individuals who do not satisfy the selection criterion.

Suppose as in the preceding section that there are M independent individuals $i = 1, \ldots, M$ from which the study sample is selected. Suppose that at the time the sample is selected, individuals have followup times τ_i $(i = 1, \ldots, M)$, which are assumed independent of the event processes. Suppose also that individuals whose event histories $N_i^{(\tau_i)}$ satisfy a condition C are always selected, but that a random fraction $p > 0$ of those who do not satisfy C are also selected. Let $R_{i1} = I(N_i^{(\tau_i)}$ satisfies $C)$ and $R_i = I$ (unit i is selected); the foregoing indicates that $\Pr(R_i = 1|R_{i1} = 1) = 1$ and $\Pr(R_i = 1|R_{i1} = 0) = p$. This gives what is often termed a response-selective sample (e.g. Kalbfleisch and Lawless, 1988; Lawless et al., 1999), and we indicate two ways that inference about the event process may be approached:

(i) *Conditional likelihood.* If each individual $i = 1, \ldots, M$ is selected independently of the others, so that the R_i are mutually independent, then we can consider

$$L_C(\theta) = \prod_{i=1}^{M} \Pr(N_i^{(\tau_i)}|R_i = 1)^{R_i} . \tag{7.38}$$

This is a proper likelihood and standard methods can be applied; the primary difficulty is in calculating the probabilities in (7.38).

(ii) *Weighted pseudo-score estimating functions.* Let $\pi_i = \Pr(R_i = 1|N_i^{(\tau_i)})$ for $i = 1, \ldots, M$, and consider the estimating function

$$U_W(\theta) = \sum_{i=1}^{M} \frac{R_i}{\pi_i} \frac{\partial \log \Pr(N_i^{(\tau_i)})}{\partial \theta} . \tag{7.39}$$

This estimating function is unbiased with respect to expectation over R_i and $N_i^{(\tau_i)}$, $i = 1, \ldots, M$: to see this take the expectation of the ith term in the order $E_{N_i^{(\tau_i)}}[E_{R_i|N_i^{(\tau_i)}}(\cdot)]$.

The estimating function involves only the selected individuals, and by solving $U_W(\theta) = 0$ we obtain what is under mild conditions a consistent estimator $\widehat{\theta}$. An asymptotic variance estimate for $\widehat{\theta}$ can be obtained from the

large sample theory for estimating functions (Appendix A). An advantage of this approach is that it uses weighted versions of the usual prospective log-likelihood for the process in question.

To illustrate the two approaches, we consider the warranty data setting of Example 7.3.3.

Example 7.3.5

Suppose that a manufacturer randomly selects a small fraction p of product units that have not experienced a warranty claim; assuming this can be done and the customers contacted, they can obtain the date of sale, the followup time τ_i, and any covariates that are relevant. Suppose also as in Example 7.3.3 that claims for unit i occur according to a Poisson process with intensity function $\lambda_i(t;\theta)$, and that τ_i is independent of the claim process for unit i. The conditional likelihood (7.38) in this case is

$$L_C(\theta) \propto \prod_{i=1}^{M} \frac{\left\{ \prod_{j=1}^{n_i} \lambda_i\left(t_{ij};\theta\right) e^{-\Lambda_i(\tau_i;\theta)} \right\}^{R_{i1}} \left\{ p e^{-\Lambda_i(\tau_i;\theta)} \right\}^{R_{i2}}}{\left\{ 1 - e^{-\Lambda_i(\tau_i;\theta)} + p e^{-\Lambda_i(\tau_i;\theta)} \right\}^{R_i}}, \tag{7.40}$$

where $R_{i1} = I(N_i(\tau_i) \geq 1)$, $R_{i2} = (1 - R_{i1})R_i = I(N_i(\tau_i) = 0, R_i = 1)$, and $R_i = R_{i1} + R_{i2}$. It is actually preferable (e.g. Lawless et al., 1999) to replace p in (7.40) with $\widehat{p} = m_s/(M - m)$, where $m = \sum_{i=1}^{M} R_{i1}$ is the number of units with at least one claim, and $m_s = \sum_{i=1}^{M} R_{i2}$ is the number of units chosen for the supplementary sample. It turns out that (7.40) with $p = \widehat{p}$ can be used for inference about θ either when the R_{i2} are independent, or when a supplementary sample of fixed size m_s is chosen, in which case the R_{i2} are mildly dependent.

The pseudo-score estimating function $U_W(\theta)$ of (7.39) in this setting is

$$\sum_{i=1}^{M} \left[R_{i1} \frac{\partial \log}{\partial \theta} \left\{ \prod_{j=1}^{n_i} \lambda_i\left(t_{ij};\theta\right) e^{-\Lambda_i(\tau_i;\theta)} \right\} + \frac{R_{i2}}{p} \frac{\partial \log}{\partial \theta} \left\{ e^{-\Lambda_i(\tau_i;\theta)} \right\} \right]. \tag{7.41}$$

It has the advantage that it is simply the likelihood estimating function for data observed on independent Poisson processes, with the addition of case weights p^{-1} for units for which $N_i(\tau_i) = 0$. Therefore, if software for handling the Poisson process allows case weights, it can be used to obtain $\widehat{\theta}$. Variance estimates should, however, be obtained by estimating function theory. As with the conditional likelihood approach, it is preferable to replace p with $\widehat{p} = m_s/(M - m)$ in (7.41).

The addition of a supplementary sample can greatly increase efficiency of estimation of θ. In settings where the fraction of units with $N_i(\tau_i) > 0$ tends to

be small, estimation based on (7.36) is imprecise relative to estimation based on (7.40) or (7.41), with even a value of p as small as .01 or .05.

The relative efficiencies of estimation based on (7.40) and (7.41) depend on the setting. The weighted estimating function is typically less efficient, but easier to implement. This approach can also be used in conjunction with estimating functions in the absence of a full probability model. In particular, suppose that $U_i(N_i^{(\tau_i)};\theta)$ is an unbiased estimating function: $E\{U_i(N_i^{(\tau_i)};\theta)\} = 0$. Then the weighted estimating function

$$U_W(\theta) = \sum_{i=1}^{M} \frac{R_i}{\pi_i} \cdot U_i(N_i^{(\tau_i)};\theta) \tag{7.42}$$

is also unbiased and can be used to estimate θ.

Suppose, for example, we wish simply to estimate the rate and mean functions for the warranty claims process. Letting $\mu(t) = E\{N_i(t)\}$ and writing $d\mu(t) = \rho(t)dt$, we could use the estimating functions

$$U(t) = \sum_{i=1}^{M} U_i\left(dN_i(t); d\mu(t)\right) = \sum_{i=1}^{M} I\left(\tau_i \geq t\right)\left(dN_i(t) - d\mu(t)\right),$$

in the case where all units $i = 1, \ldots, M$ are observed, so that their followup times τ_i are known. This gives the Nelson–Aalen estimate (3.17) for $\mu(t)$. In the present setting, we can replace the estimating functions $U(t)$ with the functions

$$U_W(t) = \sum_{i=1}^{M} \frac{R_i}{\pi_i} I\left(\tau_i \geq t\right)\left(dN_i(t) - d\mu(t)\right). \tag{7.43}$$

The equations $U_W(t) = 0$ give

$$d\widehat{\mu}(t) = \frac{\sum_{i=1}^{M} \frac{R_i}{\pi_i} I\left(\tau_i \geq t\right) dN_i(t)}{\sum_{i=1}^{M} \frac{R_i}{\pi_i} I\left(\tau_i \geq t\right)}. \tag{7.44}$$

As earlier, it is preferable to replace π_i with $\widehat{\pi}_i = \widehat{p}$ for units for which $R_{i1} = 0$. Note that because $dN_i(t) = 0$ for units for which $R_{i2} = 0$, (7.44) reduces to

$$d\widehat{\mu}(t) = \frac{\sum_{i=1}^{M} R_{i1} I\left(\tau_i \geq t\right) dN_i(t)}{\sum_{i=1}^{M} \left\{R_{i1} + \frac{R_{i2}}{p}\right\} I\left(\tau_i \geq t\right)}, \tag{7.45}$$

where $R_{i2} = R_i(1 - R_{i1})$. An approach that is sometimes useful in practice is to consider the τ_i as random variables, independent of the event processes, and to replace the denominator of (7.45) with its expectation. Simple calculation

shows this to be $M \cdot G(t) = M \cdot \Pr\{\tau_i \geq t\}$, so that if an estimate $\widehat{G}(t)$ is available from auxiliary data, the estimate

$$d\widehat{\mu}(t) = \frac{\sum_{i=1}^{M} R_{i1} I(\tau_i \geq t)\, dN_i(t)}{M \cdot \widehat{G}(t)} \tag{7.46}$$

can be used. Hu and Lawless (1996a) consider this in connection with warranty data; see also Hu and Lawless (1996b) concerning supplementary sampling.

Finally, we note that the supplementary sampling scheme discussed here can be viewed as analogous to case-control sampling in epidemiological contexts. In the present setting, individuals with at least one event over given followup periods $[0, \tau_i]$ are the cases and individuals with no events are the controls. Clearly, other similar study designs are possible, depending on the setting; for example, one might select as "cases" all individuals who experience at least one event over a specified calendar time period. A condition that should be kept in mind, however, is that we assume the population from which individuals are selected is of a known size (at least approximately), so that the sampling probabilities π_i for "controls" (individuals with $(R_{i1} = 0)$) are known.

7.4 Bibliographic Notes

Intermittent followup of event history processes is discussed by Grüger et al. (1991), Lawless and Yan (1992), and others. Thall (1988), Thall and Lachin (1988), Thall and Vail (1990), Staniswalis et al. (1997), and Lawless and Zhan (1998) consider the case of regression for recurrent events. Sun and Kalbfleisch (1993, 1995) and Wellner and Zhang (2000) consider nonparametric estimation of rate and mean functions in Poisson processes from interval-count data, whereas Lawless and Zhan (1998) and Balshaw and Dean (2002) consider piecewise-constant rate functions. Nielsen and Dean (2005) consider regression splines for baseline rate functions and apply these to interval-count data. Stukel (1993) compares several methods when individuals have the same inspection times; methods for the longitudinal analysis of counts (e.g. Diggle et al., 2002) apply here. Cameron and Trivedi (1998) extensively discuss the regression analysis of count data, and Karlis and Xekalaki (2005) survey mixed Poisson distributions. Sun and Matthews (1997) and Sun and Wei (2000) consider estimation of regression parameters without estimation of baseline rate or mean functions. Hu et al. (2002) consider estimates of both regression parameters and baseline functions, and discuss the data in Section 7.1.3. Zhang (2002), Zhang and Jamshidian (2003), and Wellner et al. (2004) also consider semiparametric methods. Additional references and discussion of semiparametric methods are given by Sun (2006, Ch. 9).

Chen et al. (2005) consider interval-count data for multiple types of events. Chen and Cook (2003) consider a case where one event process, which is used as a covariate for another process, is observed only intermittently. Satten and Sternberg (1999) and Sternberg and Satten (1999) discuss the difficult case of semi-Markov models with intermittent observation.

Event-dependent censoring is reviewed by Miloslavsky et al. (2004), who stress the use of inverse probability of censoring weighting due to Robins and Rotnitzky (1992, 1995). This idea has received considerable development for event history processes; see, for example, Satten et al. (2001), Datta and Satten (2001, 2002), and Cook et al. (2003). Ghosh and Lin (2002) discuss inverse probability of censoring weights for semiparametric regression of the marginal rate function, and Ghosh and Lin (2003) consider a bivariate time transform model. Methods discussed in Chapter 6 for dependent terminating events can also be adapted for event-dependent censoring.

When observation of individuals is intermittent, event-dependent inspection and loss to followup have received rather little discussion. Sun et al. (2005) consider a nonanticipatory, or conditionally independent, model for intermittent inspection as in Section 7.1. References on longitudinal data analysis (e.g. Diggle and Kenward, 1994; Little, 1995; Diggle et al., 2007) provide additional discussion, and discuss anticipatory or nonignorable mechanisms. Farewell et al. (2003) consider such nonignorable loss to followup in disease clinic databases, and emphasize the need to trace individuals who are lost-to-followup. Baker et al. (1993), Frangakis and Rubin (2001), Lee and Wolfe (1998), and Lee and Tsai (2005) also consider this issue. Lin et al. (2004) develop IPCW methods for marginal estimation with event-dependent inspection times.

Event-dependent selection of individuals is considered in many specific contexts, as illustrated in the examples of Section 7.3. For other examples, not involving recurrent events, see, for example, Kalbfleisch and Lawless (1988, 1989), and Brookmeyer and Gail (1994), when individuals must experience some event in order to be selected. For examples of retrospective data, see Allison (1985), Andersen and Green (1985), Trussell et al. (1992), and Diamond and McDonald (1992). Sternberg and Satten (1999) consider event-dependent selection and intermittent inspection together.

The use of time since the last event (see Problem 7.11), or the times of the last two or three events before selection has often been considered in demography and the social sciences (e.g. Allison, 1985; Baydar and White, 1988; Keiding et al., 2002; Keiding, 2006). However, the conditions needed to obtain information on a process of recurrent events are extremely restrictive, particularly when gap times are involved (e.g. Allison, 1985; Heckman and Singer, 1986).

7.5 Problems and Supplements

7.1. Renewal processes are difficult to fit when only interval-count data are available, and alternatives to maximum likelihood may be sought.

a. Write down the likelihood for a renewal process (2.37) where the gap times have distribution function $F(w)$, in the case of the following data: one event in $[0, b_1]$, zero events in $(b_1, b_2]$, and two events in $(b_2, b_3]$, where b_1, b_2, b_3 are prespecified inspection times. Consider the generalization to an arbitrary set of interval counts, as in Section 7.1.

b. Problem 2.7 gives an integral equation involving the mean function $\mu(t)$ and $F(w)$, for a renewal process. Show that the equation

$$F(t) = \mu(t) - \int_0^t F(t-x)d\mu(x) \tag{7.47}$$

also holds for $t > 0$. If you have a smooth parametric estimate of $\widehat{\mu}(t)$, consider how you might obtain an estimate of $F(t)$ from (7.47). What concerns would you have about this method?

[Section 7.1; Erto, 1989; Baxter, 1994; Nelson, 2003, Section 8.4]

7.2. Consider the following artificial extreme case where inspection process conditions (i) and (ii) in Section 7.1 are violated: when an event occurs, an individual arrives for an inspection soon after. Suppose that an individual is inspected at times $b_1 < \cdots < b_k$ but that the analyst is unaware that events have occurred just prior to each of these times. Under a Poisson process with rate function $\rho(t)$, the analyst might then assume conditions (i) and (ii) hold, and take the likelihood contribution for the individual to be

$$\prod_{j=1}^{k} \left(\int_{b_{j-1}}^{b_j} \rho(t)dt \right) \exp\left(- \int_{b_{j-1}}^{b_j} \rho(t)dt \right) \tag{7.48}$$

whereas it should be approximately

$$\prod_{j=1}^{k} \rho(b_j) \exp\left(- \int_{b_{j-1}}^{b_j} \rho(t)dt \right) . \tag{7.49}$$

Show that no bias in estimation is incurred in using (7.48) when $\rho(t) = \rho$ is constant. Investigate the bias in a case where $\rho(t)$ is not constant.

[Section 7.1]

7.3. Consider the covariance structure for interval counts n_{ij} $(i = 1, \ldots, m; j = 1, \ldots, k_i)$ under the mixed Poisson process framework (7.3). This is (see also (7.17) and (7.18) in the case where $J = 1$)

$$\mathrm{var}(n_{ij}) = \mu_{ij} + \phi\mu_{ij}^2, \quad \text{and} \quad \mathrm{cov}(n_{ij}, n_{i\ell}) = \phi\mu_{ij}\mu_{i\ell} \quad (j \neq \ell) .$$

Let $n_i = (n_{i1}, \ldots, n_{ik_i})'$, $\mu_i = (\mu_{i1}, \ldots, \mu_{ik_i})'$, $D_i = \partial\mu_i/\partial\theta'$, and $V_i = \text{cov}(n_i)$, with entries given above. Consider the estimating equations

$$U_1(\theta, \phi) = \sum_{i=1}^{m} D_i' V_i^{-1}(n_i - \mu_i) = 0 \tag{7.50}$$

which, given a value for ϕ, can be solved to provide an estimate $\widehat{\theta}$. As a supplementary estimating equation for ϕ, consider

$$U_2(\theta, \phi) = \sum_{i=1}^{m} \frac{\mu_{i\cdot}^2}{\sigma_{i\cdot}^4} \left\{ (n_{i\cdot} - \mu_{i\cdot})^2 - \sigma_{i\cdot}^2 \right\} = 0\,, \tag{7.51}$$

where $n_{i\cdot} = \sum_{j=1}^{k_i} n_{ij}$, $\mu_{i\cdot} = \sum_{j=1}^{k_i} \mu_{ij}$ and $\sigma_{i\cdot}^2 = \text{var}(n_{i\cdot}) = \mu_{i\cdot} + \phi\mu_{i\cdot}^2$. To estimate θ and ϕ we may proceed iteratively by (i) choosing an initial value $\widehat{\phi}_0$ for ϕ ($\phi_0 = 0$ is suitable) and solving (7.50) to get a value $\widehat{\theta}_1$ for θ; (ii) solving (7.51) with $\theta = \widehat{\theta}_1$ fixed, to get a value $\widehat{\phi}$; and (iii) repeating steps (i) and (ii) with $\widehat{\phi}_\ell$ replacing $\widehat{\phi}_{\ell-1}$ at stages $\ell = 2, 3, \ldots$. Under suitable conditions this iteration scheme will converge to give the estimate $(\widehat{\theta}, \widehat{\phi})$. The estimating equation (7.51) is an alternative to (7.20) in the case where $J = 1$.

a. Show that (7.50) is an unbiased estimating equation, assuming the specification for $E(n_{ij})$ is correct, and that (7.51) is an unbiased estimating equation if the specification for $V_i = \text{cov}(n_i)$ is correct.
b. Use asymptotic results for estimating functions (Appendix A) to obtain an estimated asymptotic covariance matrix for $\sqrt{m}(\widehat{\theta}' - \theta', \widehat{\phi} - \phi)'$.
c. Show that if the specification for V_i is incorrect but that for $E(n_{ij})$ is correct, then $\widehat{\theta}$ is still estimated consistently by the approach above. Show that an estimated asymptotic covariance matrix for $\sqrt{m}(\widehat{\theta} - \theta)$ is given by $\widehat{A}^{-1}\widehat{B}\widehat{A}^{-1}$, where

$$\widehat{A} = \frac{1}{m} \sum_{i=1}^{m} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i$$

$$\widehat{B} = \frac{1}{m} \sum_{i=1}^{m} \widehat{D}_i' \widehat{V}_i^{-1} (n_i - \widehat{\mu}_i)(n_i - \widehat{\mu}_i)' \widehat{V}_i^{-1} \widehat{D}_i$$

and $\widehat{D}_i, \widehat{V}_i$ are D_i, V_i with $\widehat{\theta}, \widehat{\phi}$ replacing θ, ϕ.
d. Apply the approach here to estimate the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$, and ϕ for the model with $\mu_0(t) = \alpha_1 t^{\alpha_2}$, applied to the bladder cancer data of Section 7.1.3.

[Section 7.1; Lawless and Zhan, 1998]

7.4. Consider the case of $J = 2$ types of events in the framework of Section 7.1.4.

a. Examine the covariance structure for the $n_{ij\ell}$ when $u_{i1} = v_1 + v_3$, $u_{i2} = v_2 + v_3$, where v_1, v_2, v_3 are independent gamma random variables with means 1 and variances ϕ_1, ϕ_{12}, ϕ_3. Consider the likelihood function based on the interval counts $n_{ij\ell}$ ($\ell = 1, \ldots, k_{ij}$; $j = 1, \ldots, J$; $i = 1, \ldots, m$).
b. Also consider the covariance structure and the likelihood function based on the $n_{ij\ell}$ when $(\log u_{i1}, \log u_{i2})$ have a bivariate normal distribution with mean $(0, 0)$, variances σ_1^2, σ_2^2, and correlation ρ.

[Section 7.1.4; Chen et al., 2005]

7.5. The data in Section 1.2.1 on the occurrence of tumors in rats exposed to a carcinogen were analyzed in Section 3.8.1 by treating the event times as exact. In fact, the rats were examined at intervals of two to six days, and the number of new tumors since the last examination was recorded. The examinations were on days 3, 5, 8, 11, 14, 17, 21, 24, 28, 31, 35, 38, 42, 45, 48, 52, 56, 59, 61, 63, 66, 70, 74, 77, 80, 85, 88, 90, 92, 97, 101, 107, 112, 114, 119, 122. Fit the parametric models (3.54) considered in Section 3.8.1 to the treatment and control groups by treating the data as interval counts. Consider both the Poisson process model and the negative binomial process model in each case.

[Section 7.1.2]

7.6. When there are no covariates as in Section 7.2.2, the probabilities $p_r(t) = \Pr\{N_i(t) = r\}$ can be estimated by noting that

$$p_r(t) = \Pr\{T_r \le t\} - \Pr\{T_{r+1} \le t\} \qquad r = 1, 2, \ldots , \tag{7.52}$$

where T_r is the time of the rth event ($r = 1, 2, \ldots$) and $T_0 = 0$. Estimates of the $F_r(t) = \Pr(T_r \le t)$ thus provide estimates of the $p_r(t)$.

a. If censoring times C_i are completely independent of the recurrent event process, then $F_r(t)$ can be estimated using the Kaplan–Meier estimate based on the data $t_i = \min(T_{ir}, C_i), \delta_i = I(T_{ir} \le C_i)$. Show that if there is no censoring at all, then (7.52) produces the same estimate as the Markov (Aalen–Johansen) estimate in Section 5.3.2.
b. Note that if $H_r(t)$ is the cumulative hazard function for T_r, then the Kaplan–Meier estimate of T_{ir} is based on the estimates

$$d\widehat{H}_r(s) = \frac{\sum_{i=1}^{m} I\left(T_{ir} = s, C_i \ge s\right)}{\sum_{i=1}^{m} I\left(T_{ir} \ge s, C_i \ge s\right)} . \tag{7.53}$$

Show how (7.53) can be inconsistent for $dH_r(s)$ when T_{ir} and C_i are not independent. Consider the case where the recurrent event process is Markov as in Section 5.3.2, and censoring is event-dependent as in (7.31). What happens when the recurrent event process is a Poisson process?

[Sections 5.3.2, 7.2.2]

7.7. Show that (7.32) is consistent for $p_r(t)$ provided $\widehat{G}_i(t)$ is based on (7.27) and (7.26) holds.

[Sections 7.2.1, 7.2.2]

7.8. Consider the model in (7.33) for the case of a single observation period $(b_{i0}, b_{i1}]$, with $w_{i1} = b_{i1} - b_{i0}$. Suppose that the event process is Poisson, with $\mu_{i1} = \rho w_{i1}$ and that $h_i(w_{i1}) = \lambda$, so $H_i(w_{i1}) = \lambda w_{i1}$. Obtain $\Pr(n_{i1}|w_{i1})$ under the model (7.33) when $\alpha = 1$ and z_i has a gamma distribution with density (2.28). Compare this with $\Pr(n_{i1}|w_{i1})$ obtained under the assumption that $\alpha = 0$, that is, that w_{i1} is independent of z_i. What does the estimate $\widehat{\rho} = n_{\cdot 1}/w_{\cdot 1}$, based on observation periods for m independent individuals, converge to in probability?

[Section 7.2.3]

7.9. Use the multistate framework of Sections 5.3 and 6.6.4 to deal with estimation of marginal features when there is a terminating event present, and event-dependent censoring. In particular, consider the model in Figure 6.2 with states added to represent censoring. Consider ways to estimate $\rho(t)dt = E\{dN_i(t)|T_i \geq t\}$ and $E\{N_i(t)\}$ as in Section 6.6.3, either through the multistate framework or otherwise.

[Section 7.2.2]

7.10. For the bladder tumor data discussed in Section 7.1.3 (see also Appendix D.1), check to see whether there is any evidence that the censoring time $\tau_i = b_{ik_i}$ is related to previous event history by considering binary response models for

$$\Pr\{b_{ij} \text{ is the last inspection time for individual } i|H_{ij}\} \qquad j = 1, 2 \ldots,$$

where H_{ij} is the event, inspection time, and covariate history up to time b_{ij}. Consider in particular whether the number of events in $(b_{i,j-1}, b_{ij}]$ is related to the probability of dropout at b_{ij}.

[Sections 7.1, 7.2]

7.11. Researchers in certain disciplines frequently try to use retrospective data on the time U_i since the last event in a process, obtained from a cross-sectional survey of individuals at some calendar time. This is difficult to do in many situations but for certain processes, such data can be used for estimation.

a. Suppose the event process for individual i is Poisson with rate function $\rho_i(t)$ and that the cross-sectional survey occurs at $t = a_i$ for individual i. Show that in this case

$$\Pr(U_i > u) = \exp\left\{-\int_{a_i-u}^{a_i} \rho_i(t)dt\right\}$$

and thus obtain the density function for U_i. If $\rho_i(t) = \rho_0(t)\exp(x_i'\beta)$, where x_i is a vector of covariates, consider maximum likelihood estimation based on the U_i. Note that an individual with no events by a_i contributes a term $\Pr(U_i > a_i)$ to the likelihood.

b. Consider the corresponding situation when events follow a renewal process, as described in Section 4.5. What additional information is needed here that is not needed in the case of a Poisson process?

c. In the renewal process case, the distribution of the backward recurrence time $U_i = -t_{i0}$ is given in Problem 4.10. If the event rate before time a_i is constant, show that U_i has the equilibrium backward recurrence time distribution

$$f_0^*(u_i) = \frac{1}{\mu}S_0(u_i)\,,$$

where μ is the mean gap time (between events) and $S_0(w)$ is the survivor function for the gap time. Investigate maximum likelihood estimation when $S_0(w) = (1+\lambda w)^{-\alpha}$, and compare the information in U_i versus that in a prospectively observed gap time W_i.

[Section 7.3; Allison, 1985; Keiding, 2006]

7.12. Consider the case of equipment, such as a motor vehicle, in which usage (e.g. distance driven) accumulates linearly with rate $z_i > 0$ for the ith unit. Let $N_i(t)$ denote the number of "failures" experienced by the equipment up to time t after the unit is placed in service. Suppose that, given z_i, the process $\{N_i(t), 0 \le t\}$ is Poisson with mean function of the form

$$\mu_i(t) = \mu_0(tz_i^{\beta})\,,$$

where $\mu_0(t) = \mu_0(t;\alpha)$ is a baseline rate function and α, β are unknown parameters. Note that the accumulated usage to time t is $w_i = z_i t$ and thus that if $\beta = 0$ the event process is independent of the usage rate, and if $\beta = 1$ the mean function depends only on the accumulated usage.

a. Consider data where τ_i is a fixed time, and the failure times and z_i are observed if and only if $N_i(\tau_i) > 0$. If $N_i(\tau_i) = 0$ the value of z_i is unobserved. For convenience, label the units for which $N_i(\tau_i) > 0$ as $i = 1, \ldots, m$ and those for which $N_i(\tau_i) = 0$ as $i = m+1, \ldots, M$. Write down a likelihood function based on the data.

b. Suppose the distribution of z_i in the population is known or estimated from external sources to have density function $g(z)$, $z > 0$. Use this additional information to give a likelihood function that incorporates the knowledge that $N_i(\tau_i) = 0$ for $i = m+1, \ldots, M$.

[Section 7.3; Lawless et al., 1995]

7.13. Assessment of the effect of external factors from case-only data. Suppose that, subject to an external time-varying covariate proces $\{x_i(t), 0 \leq t\}$, an individual experiences events according to a Poisson process with intensity function

$$\lambda_i(t) = \lambda_0(t) \exp\{\alpha_i + x_i'(t)\beta\},$$

where α_i represents individual-specific factors, which may be unobservable. Assume that data are available from m individuals who experienced at least one event over some calendar time period that corresponds to the time intervals $[a_i, b_i]$ for individuals $i = 1, \ldots, m$.

a. Base a conditional likelihood on the distribution of the event times t_{ij} $(j = 1, \ldots, n_i)$ for individual i, given that $N_i(a_i, b_i) = n_i$. Show that this involves $\lambda_0(t)$ and β, but not α_i.
b. Consider estimation of the parameters γ and β, assuming a parametric model $\lambda_0(t; \gamma)$. Examine the special case where $\lambda_0(t; \gamma) = \gamma$ is constant.
c. Derive a score test of the hypothesis $H_0 : \beta = 0$.

[Section 7.3, Farrington and Whitaker, 2006]

8
Other Topics

8.1 Event Processes with Marks

In many settings there are auxiliary data associated with an event that reflect the severity, importance, or implications of its occurrence. The associated random variable is called the *mark* of the event and the process as a whole is called a *marked point process*. Settings in which marked point processes arise include casualty insurance, where claims have an associated size; warranty claims, where interest lies both in the occurrence of claims and the cost of repairs; and medicine, where one may be interested in the costs of treatment associated with specific medical events.

Let $\{N_i(t), 0 \leq t\}$ denote a counting process with $N_i(\tau_i) = n_i$ events occurring over $[0, \tau_i]$ at times $t_{i1}, \ldots, t_{in_i}$ for individual i. In some settings, it may be desirable to allow various forms of dependence between event occurrence and the associated marks (e.g. Cox and Isham, 1980, Section 5.5), but we mostly assume here that the marks at different event times are independent and identically distributed. The extension to allow marks and event occurrence to depend on covariates, but remain conditionally independent, is straightforward. Let C_{ij} be a possibly vector-valued random variable that denotes the mark associated with the event at t_{ij}, with cumulative distribution function $G_i(c)$. If C_{ij} is a vector, its components needn't be independent. The history of the full process is then $H_i(t) = \{N_i(s), \tilde{C}_i(s), 0 \leq s < t\}$, where $\tilde{C}_i(s) = \{C_{i1}, \ldots, C_{iN_i(s)}\}$.

The cumulative sum

$$C_i(t) = \sum_{j=1}^{N_i(t)} C_{ij} \tag{8.1}$$

is often of central importance, especially when the marks represent costs of some type; then $C_i(t)$ is the cumulative cost up to time t. We focus on the case where the C_{ij} $(j = 1, 2, \ldots)$ are scalar, in addition to being independent and identically distributed with mean μ_c and variance σ_c^2. It then follows easily from (8.1) that

$$E\{C_i(t)\} = \mu_c E\{N_i(t)\}, \quad \text{var}\{C_i(t)\} = \sigma_c^2 \, E\{N_i(t)\} + \mu_c^2 \text{var}\{N_i(t)\} . \quad (8.2)$$

It can also be seen that if $M_Y(s) = E\{\exp(Ys)\}$ denotes the moment generating function of the random variable Y, then

$$M_{C_i(t)}(s) = M_{N_i(t)}\left(\log M_{C_{ij}}(s)\right) ,$$

provided the generating functions exist.

A very important special case is when the event process is Poisson, in which case the process $\{C_i(t), 0 \le t\}$ is called a *compound Poisson process*; see Problem 2.5. Under the assumption that the mark values C_{ij} $(j = 1, 2, \ldots)$ are independent, it is easily seen that the process $\{C_i(t), 0 \le t\}$ then has *independent increments*. That is, for $0 \le s_1 < t_1 < s_2 < t_2$, the increments $C(s_1, t_1) = C(t_1) - C(s_1)$ and $C(s_2, t_2) = C(t_2) - C(s_2)$ are independent.

We remark that multitype recurrent event data discussed in Chapter 6 may be viewed as a marked point process in which the mark indicates the type of the event. In that case the random variable J_{ik} associated with a kth event at t_{ik} takes values on the set $\{1, \ldots, J\}$. The J_{ik} in Chapter 6 may depend on $H(t_{ik})$ and the marked point process models here can also be extended to allow this. We do not pursue this specifically, but allow for the possibility in writing down a likelihood function for observed data.

Define the history $H_i(t) = \{N_i(s), C_i(s), 0 \le s < t\}$ and let $\Delta N_i(u_r)$ and $\Delta C_i(u_r)$ contain the information on events and marks in a short interval $[u_r, u_{r+1})$. Then if we consider events over $[0, \tau]$ and the partition $0 = u_0 < u_1 < \cdots < u_R = \tau$, the probability of the data for individual i is

$$\prod_{r=0}^{R} \Pr(\Delta C_i(u_r), \Delta N_i(u_r) | H_i(u_r)) =$$

$$\prod_{r=0}^{R} \Pr(\Delta N_i(u_r) | H_i(u_r)) \Pr(\Delta C_i(u_r) | \Delta N_i(u_r), H_i(u_r)) .$$

As $R \to \infty$ and the $\Delta u_r = u_{r+1} - u_r \to 0$, the likelihood contribution from individual i becomes, by the line of argument leading to (2.7),

$$\left\{ \prod_{j=1}^{n_i} \lambda_i \left(t_{ij} | H_i(t_{ij})\right) \Pr\left(C_{ij} | H_i(t_{ij}), dN_i(t_{ij}) = 1\right) \right\}$$

$$\times \exp\left\{ - \int_0^\tau \lambda_i(u | H_i(u)) du \right\} . \quad (8.3)$$

This factors into a component for the event generating process and a component for the conditional distribution of the marks given the history of the process at each t_{ij}. In analyzing data on a marked point process, we typically consider the event process and then the distribution of marks, which as shown here, can be allowed to depend on time and on previous event or cost history.

This is conveniently done using regression models and standard exploratory techniques.

When the two factors in (8.3) involve separate parameters, estimation may be carried out separately for the two aspects of the process. Often processes are modeled in such a way that this is the case. For the compound Poisson processes described above, for example, the parameters in $G(c)$ are usually distinct from those in the event process intensity. However, for multitype events we often adopt models where a different type of factorization applies; see Problem 8.4. We next consider cumulative processes such as (8.1), and extend the framework to allow for terminating events and more general accumulation processes.

8.2 Models for Cumulative Costs

8.2.1 Introduction

Costs or benefits that accumulate over time for individuals are of interest in many life history processes. Familiar examples include the cost of health care for persons with chronic medical conditions, the payments to insured persons during periods of disability, the payments associated with property insurance claims, and cumulative quality of life measures which are sometimes used in the evaluation of treatments for terminally ill patients. Costs or benefits may be multivariate and may accrue for a variety of reasons. For example, in studies of persons with chronic obstructive pulmonary disease, costs were incurred by prescription of prophylactic or therapeutic medications, by hospitalizations, by time off work, and so on. Note that this setting is more general than that for marked point processes discussed in Section 8.1, where costs only accrue at the occurrence of events.

For convenience we often use the term costs to refer to cost or other cumulative measures such as utility, profit, or quality of life, and let $C(t)$ denote a cumulative (univariate) cost for an individual over the time period $[0, t]$. There is often also a random variable T that represents the duration of the cumulative process, so the objects of interest are T and $\{C(t),\ 0 \le t \le T\}$. In some contexts it is sufficient to base analyses directly on the cost data accumulating over time, but it is usually more informative to consider models for the underlying processes that generate costs as well. Advantages of analyzing and modeling the processes generating the costs include increased understanding; the ability to deal with observation schemes involving censoring, intermittent observation, or truncation; better methods for predicting costs; and a convenient separation of the underlying processes from costs, which may be subjective or vary across locations.

There are thus two main approaches for the analysis of accumulating cost data. The first is to directly model, for each individual, the cumulative cost process $\{C_i(t),\ 0 \le t\}$, and time T_i at which the process terminates. The

time T_i may, for example, represent the time of death in a study of health services utilization among patients with terminal medical conditions, or the time treatment for a transient condition ends. In many settings T_i is subject to right censoring at some time τ_i, in which case the cost process is unobserved for $t > \tau_i$. Interest may lie in estimation of the distribution of "total lifetime cost" $C_i = C_i(T_i)$, or just the expected lifetime cost $E\{C_i\}$. More generally, we may wish to model the accumulation of cost right up to T_i. In most realistic situations T_i is not independent of the cost process; more specifically, if $C_i^{(t)} = \{C_i(u),\ 0 \le u < t\}$ is the cost history to time t, then the termination time hazard function,

$$\lim_{\Delta t \to 0} \frac{\Pr(T_i < t + \Delta t | T_i \ge t, C_i^{(t)})}{\Delta t}\,, \tag{8.4}$$

depends on $C_i^{(t)}$. This implies that, even if the censoring time τ_i and $(T_i, C_i(T_i))$ are independent, the censoring value $C_i^* = C_i(\tau_i)$ and the total lifetime cost $C_i = C_i(T_i)$ are not in general independent.

The second approach involves modeling the underlying process that generates the cost, as well as the distribution of costs. We need to distinguish between situations where costs are incurred only in connection with events that occur at points in time and situations where costs may accrue on a continuous basis, according to which of various states an individual occupies over time. For the former, marked point processes provide a convenient framework, with the cumulative cost process given by (8.1); we discuss statistical methods associated with this approach in the next section. For the second type of situation, the multistate model formulation below is useful. More generally, we may formulate models that combine both continuously accruing costs and costs associated with specific events.

The following framework is useful for costs that accrue continuously. Suppose that at time t an individual occupies one of K life states $1, \ldots, K$, where all individuals begin in state 1 at $t = 0$, states $1, \ldots, K-1$ are transient, and state K is an absorbing state in which there is no further accumulation of cost. Any type of multistate model may be used to characterize the underlying process; some have been discussed in Chapters 5 and 7. The states, in particular, may indicate numbers of events that have occurred, as well as an individual's "condition". If $Z(t)$ represents the state occupied by an individual at time t, assume that the incremental cost over the short interval $[t, t+dt)$ is $V\{Z(t), t\}dt$, which could be deterministic or stochastic. The total cumulative cost up to time t is then

$$C(t) = \int_0^t V\{Z(u), u\}du\,. \tag{8.5}$$

The process terminates upon entry to state K at time T, and $V\{K, u\} = 0$ for all $u > 0$.

We restrict consideration here to cases where

$$V\{Z(u), u\} = v_j(u) \qquad \text{if } Z(u) = j\,, \tag{8.6}$$

where $v_j(u)$ is a known (deterministic) function, $j = 1, 2, \ldots, K-1$. In this case (8.5) gives

$$C(t) = \sum_{j=1}^{K-1} \int_0^t v_j(u) \cdot I[Z(u) = j] du \tag{8.7}$$

and

$$E\{C(t)\} = \sum_{j=1}^{K-1} \int_0^t v_j(u) \cdot p_j(u) du\,, \tag{8.8}$$

where

$$p_j(u) = \Pr[Z(u) = j | Z(0) = 1] \qquad j = 1, \ldots, K\,, \tag{8.9}$$

are state occupancy probability functions. Note that in this framework $C(T) = C(\infty)$.

Assumptions regarding the process $\{Z(t), 0 \le t\}$ may be made to estimate $p_j(t)$. In a completely general setting, transition intensities might depend on prior cost history, but in the case of deterministic cost rates (8.6) we have

$$\Pr\left\{Z(t + \Delta t) = j | H(t), C^{(t)}\right\} = \Pr\left\{Z(t + \Delta t) = j | H(t)\right\}\,,$$

where $H(t) = \{Z(u), 0 \le u < t\}$ and $C^{(t)} = \{C(u), 0 \le u < t\}$, so we merely need to model the multistate process.

8.2.2 Estimation for Cost Processes

We focus on the case of cost processes generated by point events as in Section 8.1, and only discuss briefly the case of continuously accruing costs. In the case of cost processes of the type (8.1), we typically model the event process which generates the costs, along with the distributions $G_i(c)$ for the costs C_{ij} $(j = 1, 2, \ldots)$. Models for the recurrent events may involve covariates, individual level random effects, or stratification, as discussed in earlier chapters. If there is a terminating event at time T_i, this may be incorporated as described in Section 6.6. The cost distributions $G_i(c)$ are modeled and fitted separately, and may also involve covariates, random effects, or stratification. The mean and variance functions for the cost process, or the moment generating function (8.2), are then readily estimated. For example, when costs are independent of the event times, $E\{C_i(t)\}$ is estimated by $\widehat{\mu}_c\,\widehat{\mu}(t)$, where $\mu_c = E(C_{ij})$ and $\mu(t) = E\{N_i(t)\}$. Variance estimates or confidence intervals can be obtained by standard methods.

Many modeling strategies are possible, and we cannot give an exhaustive discussion. Complex models in which costs may depend on previous events and costs can be fitted using (8.3), but the cost process (8.1) is generally complicated and may have to be examined by simulation. Often a simple

stratified approach is useful, in which events are split into types $j = 1, \ldots, J$, with events of type j generating a fixed cost C_j. If $N_{ij}(t)$ is the number of type j events for individual i over $[0, t]$, then

$$C_i(t) = \sum_{j=1}^{J} C_j N_{ij}(t)\,.$$

Models for multitype events, discussed in Chapter 6, can be used in this setting and the properties of the cumulative cost process are readily determined.

It should be stressed that the assumption of costs being independent of previous event or cost history is strong, and should be checked in any specific setting. This can be done by fitting models for C_{ij} $(j = 1, 2, \ldots)$ in which covariates represent aspects of previous history. Temporal trends in costs can similarly be examined.

The multistate setting where individuals accrue cost continuously is also straightforward to apply, provided the cost rate functions are deterministic, as in (8.6). A discussion of this area is beyond our present scope, but an array of methods which parallel those in this book are available for general multistate models; Andersen et al. (1993) is a comprehensive source. The developments in Sections 5.3 and 6.6 provide estimation methodology for the special case of a "progressive" Markov process, in which an individual may only make transitions from states j to $j + 1$. In particular, methods of estimating the occupancy probabilities (8.9) are given, and they may be used to estimate expected cumulative costs via (8.8).

In some situations we may be willing to settle for estimation of expected cumulative costs,

$$\mu_{ci}(t) = E\{C_i(t)\}\,, \tag{8.10}$$

where it is understood that if there is a termination time T_i, no further costs can accrue for $t > T_i$. The modeling approaches discussed above may be used to estimate the $\mu_{ci}(t)$, but robust methods that do not rely too heavily on assumptions are also attractive. An approach that can be used when there are no termination times, and where the observation period $[0, \tau_i]$ for each individual is independent of the event and cost history, is to extend the rate and mean function estimation methods of Section 3.6. In particular, let us consider $\{C_i(t), 0 \le t\}$ for independent individuals $i = 1, \ldots, m$ as non-decreasing processes with mean function $\mu_c(t)$, and consider a set of estimating equations analogous to (3.16),

$$\sum_{i=1}^{m} Y_i(s)\,\{dC_i(s) - d\mu_c(s)\} = 0 \qquad s > 0\,, \tag{8.11}$$

where $Y_i(s) = I(s \le \tau_i)$ and $dC_i(s)$ is the increment in $C_i(s)$ over $[s, s + ds)$. This unbiased estimating equation gives the estimates $d\widehat{\mu}_c(s) = d\bar{C}.(s)/Y.(s)$, where $d\bar{C}.(s) = \sum_{i=1}^{m} Y_i(s) dC_i(s)$ and $Y.(s) = \sum_{i=1}^{m} Y_i(s)$. This gives the estimate

$$\widehat{\mu}_c(t) = \int_0^t \frac{d\bar{C}_\cdot(s)}{Y_\cdot(s)} \tag{8.12}$$

and it is readily shown by a derivation analogous to that leading to (3.34) that

$$\widehat{\text{var}}\left\{\sqrt{m}\,(\widehat{\mu}_c(t) - \mu_c(t))\right\} = m\sum_{i=1}^{m}\left\{\int_0^t \frac{Y_i(s)}{Y_\cdot(s)}\left[dC_i(s) - \frac{d\bar{C}_\cdot(s)}{Y_\cdot(s)}\right]\right\}^2 \tag{8.13}$$

is a consistent estimator of the asymptotic variance for $\sqrt{m}(\widehat{\mu}_c(t) - \mu_c(t))$. In cases where the cost processes are pure jump processes (i.e. costs only accrue at discrete time points), then (8.13) reduces to a sum, but processes where costs accrue continuously over time are also allowed.

The approach just described can be extended to include covariates, in a way that parallels the methods in Section 3.6.3. It can also be extended to deal with termination times T_i which may depend on the cost process, by using the methods of Section 6.6.

More insight can generally be gained by modeling the cost generating process separately from the costs themselves, as discussed previously. In this case it is possible to apply ideas in Sections 3.6, 5.3, 6.6, and 7.1, in which we consider robust methods of estimating the mean function $\mu_i(t) = E\{N_i(t)\}$ for recurrent event processes or the occupancy functions (8.9) for multistate processes. These can be combined with estimates of expected costs or cost rates via (8.2), (8.8), and similar expressions, to yield robust estimates of $\mu_c(t)$. Adjustments using inverse probability weights can be used to deal with termination or censoring times that depend on prior event or cost history.

8.2.3 Examples

We consider below a pair of examples that illustrate how mean cost functions can be estimated. Neither example involves a terminating event but followup times vary across individuals.

Example 8.1: Field repair data

This dataset (see Appendix D) gives simulated data on unscheduled repairs for a fleet of $m = 134$ large utility vehicles operated by a city. The data were collected over a three-year period on new vehicles which were purchased and placed in service over the first two years of the study. Time is measured in years from the start of the study, and costs are in hundreds of dollars.

We consider two estimates of the mean cost function $\mu_c(t) = E\{C_i(t)\}$ per vehicle. The first uses the marked point process approach of Section 8.1; (8.2) suggests the estimate

$$\widehat{\mu}_c(t) = \widehat{\mu}_c\,\widehat{\mu}_{NA}(t)\,, \tag{8.14}$$

where $\widehat{\mu}_c$ is the average cost of all the observed repairs and $\widehat{\mu}_{NA}(t)$ is the Nelson–Aalen estimate (3.17) for the mean function $\mu(t) = E\{N_i(t)\}$ for the number of repairs. Let us also define the sample variance σ_c^2 for repair cost, so that with $n_\cdot = \sum_{i=1}^m n_i$, we have

$$\widehat{\mu}_c = \frac{1}{n_\cdot} \sum_{i=1}^{m} \sum_{j=1}^{n_i} C_{ij} \qquad \widehat{\sigma}_c^2 = \frac{1}{n_\cdot - 1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (C_{ij} - \widehat{\mu}_c)^2 \ .$$

Assuming that the repair costs do not display a time trend and are not related to previous event history, and that censoring is not event- or cost-dependent, (8.14) is valid, and a variance estimate for $\widehat{\mu}_c(t)$ is readily obtained. In particular, we get by the delta theorem that

$$\widehat{\text{asvar}}\,\{\widehat{\mu}_c(t)\} = \widehat{\mu}_c^2\ \widehat{\text{asvar}}\,\{\widehat{\mu}_{NA}(t)\} + \widehat{\mu}_{NA}(t)^2\ \widehat{\text{asvar}}(\widehat{\mu}_c)\,. \tag{8.15}$$

In (8.15) we can use $\widehat{\text{asvar}}(\widehat{\mu}_c) = \widehat{\sigma}_c^2/n_\cdot$ and either (3.19) or, preferably, the robust estimate (3.34) for $\widehat{\text{asvar}}\,\{\widehat{\mu}_{NA}(t)\}$.

A second approach is to use (8.12) for $\widehat{\mu}_c(t)$, and the variance estimate (8.13). This estimate will handle settings where the costs may depend on t or on previous event history and so makes fewer assumptions than the estimate (8.14). It also requires, however, that censoring times be completely independent of the event and cost process. Note that the variance estimate (8.13) can be written as a sum, giving

$$\widehat{\text{var}}\,\{\widehat{\mu}_c(t)\} = \sum_{i=1}^{m} \left\{ \sum_{\ell=1}^{m} \sum_{j:t_{\ell j} \le t} \frac{Y_i(t_{\ell j})}{Y_\cdot(t_{\ell j})} \left[C_{ij} I(\ell = i) - \frac{C_{\ell j}}{Y_\cdot(t_{\ell j})} \right] \right\}^2 .$$

Table 8.1 shows estimates of $\mu_c(t)$ and associated standard errors at the five values $t = 0.5$, 1.0, 1.5, 2.0, and 2.5, for (8.14) and (8.12). For (8.15), the robust variance estimate (3.34) for $\widehat{\mu}_{NA}(t)$ was used. We see that the two estimates and their standard errors are in close agreement.

Table 8.1. Estimates of the mean cost function $\mu_c(t)$ from a sample of 134 individuals with varying followup times.

t	EST. (8.14)	S.E. (8.15)	EST. (8.12)	S.E. (8.13)
0.5	11.17	0.90	11.46	0.95
1.0	22.04	1.45	21.77	1.46
1.5	31.66	1.92	31.24	1.93
2.0	40.25	2.37	40.14	2.37
2.5	51.59	3.27	51.60	3.30

These data were in fact generated so as to have no time trend in the costs of repairs. If there had been a time trend, then the estimate (8.12) would be

preferred. In addition, the censoring times here were purely administrative, that is, due to the end date of the data accumulation. Both the estimates (8.12) and (8.14) require that censoring be independent of the event and cost processes and so are suitable here. When there is a possibility that censoring is event- or cost-related we can modify the estimates through the use of IPC weights, described in Section 7.3. Finally, we note that both estimates are robust as far as the underlying recurrent event process is concerned.

Example 8.2: Heat pump repair data

Nelson (2003, Sections 1.2 and 3.3) has presented data on fan repairs for residential heat pumps covered by a service contract. The data contain information on $m = 119$ heat pumps, with varying lengths of followup; the maximum followup time was a little over 3650 days. There were a total of 26 repairs, but the dataset does not identify the pumps which had failures. It does, however, indicate the number of pumps at risk at the time of each failure and so it is possible to estimate the cumulative mean functions $\mu(t)$ for repairs and $\mu_c(t)$ for costs of repairs, as a function of pump age. Such "aggregate" data are sometimes encountered in field reliability settings where the number of equipment units in service is known at different times or ages, and where the occurrence of repairs is recorded. This limits the possibilities for analysis, though, because we are unable to study unit-to-unit variation and thus to provide probability models for failures or robust variance estimates for mean functions.

Table 8.2, adapted from Table 3.3 of Nelson (2003), shows the times $t_1 < t_2 < \cdots < t_{26}$ of repairs and at each t_i, the cost C_i of the repair and the number n_i of heat pumps at risk of failure. That is, n_i is the number of pumps with followup times greater than or equal to t_i. As in Example 8.1, we can give two estimates of the mean cost function $\mu_c(t)$, given by (8.14) and (8.12), respectively. Figure 8.1 shows the two estimates, which are somewhat different. A plot of the cost C_i versus the time t_i of each repair clearly shows that the repair costs are tending to increase with time. In this case we should trust estimate (8.12), and expect that (8.14) will be biased upwards. Variance estimation is problematic because pumps experiencing failures are not identified. For the estimate (8.12), the variance estimate (8.13) is therefore unavailable. If failures followed identical Poisson processes, then the variance estimate (3.18) for the Nelson–Aalen estimate $\widehat{\mu}_{NA}(t)$ of $\mu(t)$ could be used,

$$\widehat{\text{asvar}}\,\{\widehat{\mu}_{NA}(t)\} = \sum_{i:t_i \le t} \frac{1}{n_i^2}\,.$$

This could be combined with the variance estimate $\widehat{\text{asvar}}(\widehat{\mu}_c) = \widehat{\sigma}_c^2/26$ via (8.15), to obtain a variance estimate for $\widehat{\mu}_c(t)$ given by (8.14). Because there is typically substantial heterogeneity across units in the type of setting discussed

here, however, this variance estimate is likely to underestimate substantially the variability of $\widehat{\mu}_c(t)$. In any case, we prefer not to use (8.14) here.

Table 8.2. Times (in days) and costs (in dollars) of 26 fan failures, with numbers of fans at risk at each time.

Failure Time (t_i)	Repair Cost (C_i)	Number at Risk (n_i)	Failure Time (t_i)	Repair Cost (C_i)	Number at Risk (n_i)
141	44.20	119	2593	184.00	92
843	110.20	107	2674	167.20	91
1269	130.20	100	2710	149.00	91
1381	150.40	99	2838	42.00	89
1471	113.40	99	2946	255.70	89
1567	151.90	99	2951	243.30	89
1642	191.20	99	3296	145.00	84
1646	36.00	99	3307	208.00	84
1908	158.50	96	3368	248.70	84
2261	243.80	94	3391	256.90	84
2273	189.50	94	3440	305.50	83
2363	225.40	94	3489	202.80	83
2440	197.80	93	3635	242.00	76

8.3 Prediction

8.3.1 Introduction

In some settings the prediction of time to the next event, or the prediction of the number of events in some future time period, may be of interest. For example, in the software testing context of Sections 1.2.2. and 3.8.3, prediction of the number of additional faults that would be found by continuing testing for a prescribed period can help to guide decisions about when to cease testing. Prediction is especially of interest when there are costs associated with the events, in which case we want to predict both numbers of events and related costs. This is a common problem in the warranty claims context of Section 3.8.4, where we may want to predict the eventual total number and cost of warranty claims across a group of automobiles, based on data currently available. The prediction of events requiring medical treatment, and their associated costs, is a similar frequently occurring problem.

Given a probability model for the recurrent event process and any associated costs, we can generate predictions. Occasionally a point prediction of the time or number of future events is needed, but in most cases we wish

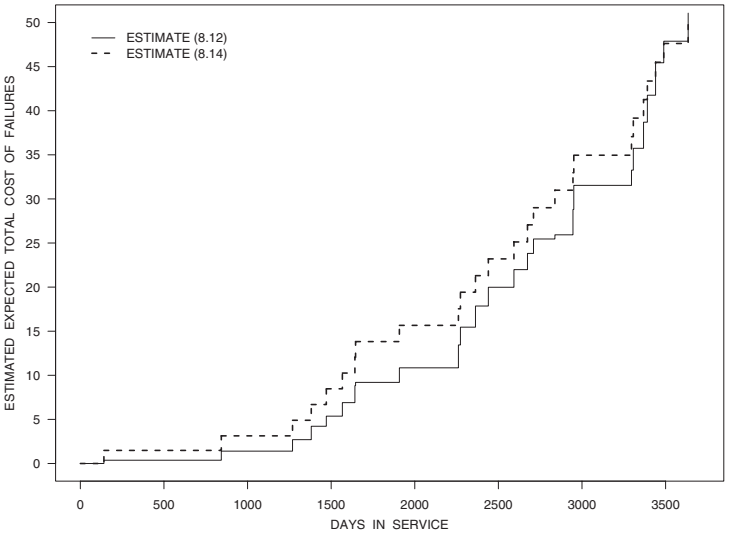


Fig. 8.1. Two estimates of mean cumulative cost function for heat pump fan motors: (1) estimate (8.12) and (2) estimate (8.14).

to account for uncertainty by giving a probability-based prediction interval. If the probability model for the process is known, we can use models discussed throughout this book for prediction. If we are simply interested in the time to the next event from some current time, conditional on the event history up to that time, then (2.8) and (2.9) are relevant under conditions in Section 2.1:

$$\Pr\{N(s, s+t) = 0 | H(s^+)\} = \exp\left\{ - \int_s^{s+t} \lambda(u|H(u)) du \right\}, \tag{8.16}$$

where $\lambda(u|H(u))$ is the process intensity, gives the probability that the next event does not occur until after time $s+t$, given the history to time s. The probability distribution of the number of events or total costs over a future time interval is in general complicated, although it is of a simple form for Poisson and some related processes. Event processes are readily simulated via their intensity functions, however, and predictive distributions can be estimated to a desired degree of accuracy that way.

Of course, exact probability distributions for a process are never known in practice, but rather must be estimated or specified from existing data. There are two main sources of error or uncertainty in predictions; one is the inherent variability in the event process, and the other is the imperfectness of the probability model adopted. Sometimes the latter source of uncertainty is broken into uncertainty about the family or type of process, and uncertainty

about the values of parameters. To consider this in a little detail, suppose that we wish to predict the residual time T to the next event, given the event history up to the current time s. By (8.16) we have

$$\Pr(T > t|H(s^+)) = \exp\left\{-\int_s^{s+t} \lambda(u|H(u))du\right\}. \tag{8.17}$$

If we specify $\lambda(u|H(u))$ through some estimation and diagnostic process as $\lambda(u|H(u);\theta)$, then we can use (8.17) with $\lambda(u|H(u);\widehat{\theta})$ to give predictive probabilities or quantiles. However, there are two additional sources of uncertainty: (i) the fact that $\widehat{\theta}$, and indeed the model $\lambda(u|H(u);\theta)$ itself, has been based on data, so is subject to sampling variation and other sources of uncertainty in the data, and (ii) the fact that the model family at best approximates the true process. This may be of special concern if the prediction involves significant extrapolation into a future time period. The procedure whereby (8.17) with $\lambda(u|H(u);\widehat{\theta})$ is used is often called "plug-in" prediction. When there is substantial uncertainty about the adequacy of the model as, for example, when extrapolation into the future is involved, one should compare predictions from a variety of plausible models to get a reasonable assessment of uncertainty. It is also possible to adjust the probabilities assigned to plug-in prediction to reflect uncertainty due to estimation. This is called *calibration*, and we consider it next, before looking at some specific examples of prediction.

Finally, we remark that if $\lambda(t|H(t))$ involves time-varying covariates, then values at future times have to be considered. This is more difficult and requires a model for the covariate process and some way to extrapolate forward in time. This issue has received little discussion in the literature.

8.3.2 Predictive Probabilities and Calibration

To simplify the discussion suppose that we wish to predict a scalar random variable Y representing a future observation, given some observed data X. Assume that the distribution function of Y given $X = x$ is determined up to a parameter θ, as $F(y|x;\theta)$; in some cases θ may be high-dimensional. Assume as well that the distribution of (Y, X) is specified by θ and that θ is estimable from X; that is, we can determine an estimate $\widehat{\theta}(x)$ from observed data $X = x$. The plug-in prediction approach simply uses $F(y|x;\widehat{\theta}(x))$; we call this the *plug-in predictive cumulative distribution function*. The predictive distribution density $f(y|x;\widehat{\theta}(x))$ is similarly called the *plug-in predictive density*. We can then find intervals $(L(x), U(x))$ such that

$$\widehat{\Pr}(L(x) \le Y \le R(x)|x) = F(R(x)|x;\widehat{\theta}(x)) - F(L(x)|x;\widehat{\theta}(x)) \tag{8.18}$$

equals a specified value such as .90 or .95.

From a frequentist viewpoint, we may wish to consider the relative frequency with which intervals $(L(x), R(x))$ determined by (8.18) bracket Y in

repeated realizations of the pair (X, Y). That is, we consider instead of (8.18) the probability

$$\Pr\left(L(X) \le Y \le R(X)\right) = \alpha(\theta)\,. \tag{8.19}$$

In a few special special settings it is possible to find functions $L(X), R(X)$ so that the probabilities (8.19) do not depend on θ. The case where Y is independent of X, given θ, has been widely considered; in that case x merely provides the estimate $\widehat{\theta}(x)$ upon which $L(x)$ and $R(x)$ are based. Here, we are interested in settings where Y and X are usually not independent. In any case, the process of determining, exactly or approximately, the function $\alpha(\theta)$ is referred to as *calibration*. This can be done using simulation, as discussed by Lawless and Fredette (2005) and references therein, and one procedure is as follows.

Suppose to start that Y is continuous, and define the random variable and associated distribution function

$$U = F(Y|X;\widehat{\theta}(X))\,, \qquad G(u;\theta) = \Pr(U \le u;\theta)\,. \tag{8.20}$$

If $\widehat{\theta}(X)$ were replaced by the true value of θ in (8.20) then U would be a pivotal quantity with a uniform distribution on (0,1). The plug-in approach to prediction in fact assumes that U is uniform on (0,1). The process of calibration (e.g. Beran, 1990; Lawless and Fredette, 2005) is to determine or approximate the true distribution of U; by doing so we hope to obtain prediction intervals $(L(X), R(X))$ for which the coverage probability (8.19) is known. If $\alpha(\theta)$ does not depend on θ in (8.18), this is possible. However, if it does depend on θ then the best we can do is determine the distribution of U, and similarly values $\alpha(\theta)$ in (8.19), for specified θ-values. In practice, the procedure that is usually followed is to use the value $\theta = \widehat{\theta}(x)$ for calibration.

We can carry out the calibration procedure as follows:

(i) Specify the approximate pivotal quantity U in (8.20), based on the estimated $\widehat{\theta}(X)$ and the model $F(Y|X;\theta)$.
(ii) Simulate B realizations (Y^*, X^*) of (Y, X) using the value $\theta = \widehat{\theta}(x)$ from the observed data; this gives $\widehat{\theta}(X^*)$ and realizations $U_1^*, \ldots, U_B^*$ of U.
(iii) Use the empirical distribution function based on $U_1^*, \ldots, U_B^*$ as an estimate $\tilde{G}(u)$ of the distribution function $G(u;\theta)$.

The distribution $\tilde{G}(u)$ can now be used to give prediction intervals or probabilities for Y, given $X = x$, as follows: associate with the prediction interval $(-\infty, y)$ for Y the probability $\tilde{G}[F(y|x;\widehat{\theta}(x)]$. This is equivalent to defining a *predictive distribution function* for Y, given $X = x$, as

$$\tilde{F}(y|x) = \tilde{G}[F(y|x;\widehat{\theta}(x))]\,, \tag{8.21}$$

with a corresponding *predictive density function* $\tilde{f}(y|x) = d\tilde{F}(y|x)/dy$. The distribution (8.21) is actually a confidence distribution; it has the mathematical properties of a probability distribution but is not the exact distribution of

any random variable. However, it produces prediction intervals $(L(X), R(X))$ that are calibrated in the sense that we approximately know $\alpha(\theta)$ in (8.19). In addition, as the amount of information about θ in X becomes arbitrarily large, both the predictive distribution (8.21) and the plug-in predictive distribution $F(y|x; \widehat{\theta}(X))$ converge in probability to the true distribution $F(y|x; \theta)$. However, the distribution (8.21) is preferred for smaller amounts of data because it accounts for uncertainty due to the estimation of θ, whereas the plug-in distribution does not.

Further discussion and properties of (8.21) are given by Lawless and Fredette (2005), but it may be used quite generally for prediction; examples are provided in the next section. When the amount of data in X is large, there is little difference between (8.21) and the plug-in distribution, so the calibration process can be skipped and the simple plug-in distribution used. Some calibration is often advisable, however, because it may not be intuitively clear how much the uncertainty in a vector of parameters affects prediction.

When the variable Y to be predicted is a count of events, the procedure above is only approximate, because $F(y|x; \theta)$ is not continuous in y, and $F(Y|X; \theta)$ is not uniform on (0,1). If Y is likely to be large, then the process above is accurate enough used as is. In settings where Y is likely to be small, there is in general no completely satisfactory approach when there is substantial uncertainty about θ. The exception is for the very special case of a homogeneous Poisson process, where exact discrete coverage probabilities are available; see Problem 8.9.

Finally, we reiterate that the procedures here assume that the true distribution of Y given X is a member of a family $F(y|x; \theta)$. Models should of course be checked for adequacy, but when a prediction involves substantial extrapolation beyond the domain of the data X, this cannot be done satisfactorily. In that case it is prudent to consider predictions based on a range of plausible models.

8.3.3 Some Examples of Prediction

We consider several examples, all of which involve extended Poisson models.

Example 8.3: Prediction in homogeneous Poisson processes

To illustrate the calculation of prediction intervals and probabilities, we consider the very special case of homogeneous Poisson processes. Subsequent examples consider more complex models.

Suppose first that a single process $\{N(t), 0 \le t\}$ is under study and that it is assumed to be a Poisson process with constant rate function, $\rho(t) = \lambda$. Let us consider the prediction of two random variables, based on data $\{n \ge 0$ events, at times $t_1 < \cdots < t_n$ in $[0, \tau]\}$: (i) $Y =$ time from τ to the next event and (ii) $Y = N(\tau, \tau + s)$, the number of events over the time period $(\tau, \tau + s]$.

For Y = time to the next event following τ, the distribution function is

$$F(y;\lambda) = \Pr(Y \le y;\lambda) = 1 - \exp(-\lambda y) \qquad y > 0\,.$$

The plug-in approach to prediction simply uses the predictive distribution $F(y;\widehat{\lambda})$, where $\widehat{\lambda}$ is the maximum likelihood estimate of λ based on the data $H(\tau)$. We require here that $n \ge 1$, in which case we find from Problem 2.4 or a special case of (3.4) that $\widehat{\lambda} = n/\tau$. To account for the estimation of λ, we may obtain a calibrated predictive distribution, as described in the preceding section. To do this, we define $U = 1 - \exp(-\widehat{\lambda}Y)$ as in (8.20), with $G(u;\lambda) = \Pr(U \le u;\lambda)$. If $\tilde{G}(u)$ accurately estimates $G(u;\lambda)$, then the calibrated predictive distribution is given by (8.21). In general, we can obtain an estimate $\tilde{G}(u)$ as described in Section 8.3.2, but in the case where observation of $N(t)$ was prearranged to continue to t_n (so that $\tau = t_n$), the distribution of U can be shown to have the closed form

$$\tilde{G}(u) = 1 - \left\{1 - \frac{1}{n}\log(1-u)\right\}^{-n} \qquad 0 < u < 1\,. \tag{8.22}$$

The derivation is sketched in Problem 8.8. Thus, U is in this case a pivotal quantity and prediction intervals are calibrated exactly. That is, the predictive distribution (8.21) in this case becomes

$$\tilde{F}_p(y) = 1 - \left(1 + \frac{\widehat{\lambda}y}{n}\right)^{-n}\,, \tag{8.23}$$

and prediction intervals $(L(\widehat{\lambda}), R(\widehat{\lambda}))$ obtained from (8.23) have exactly the nominal coverage probability.

As $n \to \infty$, the predictive distribution (8.23) converges to the plug-in distribution $F(y;\widehat{\lambda}) = 1 - \exp(-\widehat{\lambda}y)$, reflecting the fact that sampling variation in $\widehat{\lambda}$ goes to zero as $n \to \infty$, and both $F(y;\widehat{\lambda})$ and (8.23) converge in probability to the true distribution $F(y;\lambda)$ for Y. When n is small, (8.23) is preferred over the plug-in distribution. For example, when $n = 10$, (8.23) gives $\tilde{F}_p(5.85\widehat{\lambda}^{-1}) = 0.99$, so that a one-sided 0.99 prediction interval for Y is $(0, 5.85\widehat{\lambda}^{-1})$. The plug-in distribution $F(y;\widehat{\lambda})$, on the other hand, gives $F(4.61\widehat{\lambda}^{-1};\widehat{\lambda}) = 0.99$, and the 0.99 prediction interval is $(0, 4.61\widehat{\lambda}^{-1})$. The difference between this and the well-calibrated interval is quite substantial, and the actual coverage of the prediction intervals $(0 \le Y \le 4.61\widehat{\lambda}^{-1})$ is less than 0.99. If $n = 30$, the 0.99 prediction limit obtained from (8.23) is 4.98, in closer agreement with the plug-in limit. Of course, these limits rely on the assumption that a homogeneous Poisson process is the true process.

Turning to the prediction of $Y = N(\tau, \tau + s)$, we know that $N(\tau, \tau + s)$ has a Poisson distribution with mean λs, so that

$$F(y;\lambda) = \sum_{r=0}^{y} e^{-\lambda s}\frac{(\lambda s)^r}{r!} \qquad y = 0, 1, 2, \ldots\,.$$

In this case $F(Y;\lambda)$ is not uniform on (0,1) because Y is discrete, and plug-in prediction limits based on $F(y;\widehat{\lambda})$ take on values in a discrete set. We can produce "calibrated" prediction intervals through the function $U = F(Y;\widehat{\lambda})$, as described in Section 8.3.2. This is done by generating new event data $X^* = H^*(\tau^+)$ giving a new estimate $\widehat{\lambda}^*$, and independently, a value $Y^* = N(\tau, \tau+s)$; then $U^* = F(Y^*;\widehat{\lambda}^*)$. The data X^* and Y^* are generated using a Poisson process with rate $\widehat{\lambda}$. By repeating this B times, we can estimate $G(u;\lambda)$ and a predictive distribution (8.21) for Y; Lawless and Fredette (2005, p. 537–538) provide an illustration.

Similar methods apply when m identical Poisson processes are observed, and we wish to make predictions either for a single process or for an aggregate process obtained by combining individual processes. Once the total number of events observed reaches 30 or 40, there is little gained by using calibrated intervals rather than plug-in intervals; differences between calibrated and plug-in limits or probabilities are ultimately dominated by the effects of model uncertainty.

Example 8.4: Prediction for Poisson processes with random effects

A more interesting problem arises in settings where individual processes are Poisson, but there is heterogeneity across processes. Let us return to the multiplicative random effects model introduced in Section 2.2.3, where the ith process $\{N_i(t), 0 \le t\}$ is conditionally Poisson with rate function $u_i\rho(t)$, given a random variable u_i which has a gamma density $g(u)$ given by (2.28), with mean 1 and variance ϕ. Suppose that data $H_i(\tau_i) = \{N_i(t), 0 \le t < \tau_i\}$ are available for independent processes $i = 1, \ldots, m$ and that we wish to predict some future number of events $N_j(\tau_j, \tau_j + s)$ for the jth process.

If ϕ and $\rho(t)$ are known then we simply use the conditional distribution of $N_j(\tau_j, \tau_j + s)$, given $H_j(\tau_j)$, for prediction. This can be obtained as

$$\Pr\{N_j(\tau_j,\tau_j+s)=s|H_j(\tau_j)\} = \frac{\int_0^\infty \Pr\{N_j(\tau_j,\tau_j+s)=r, H_j(\tau_j)|u_j\}g(u_j)du_j}{\int_0^\infty \Pr\{H_j(\tau_j)|u_j\}g(u_j)du_j},$$

and calculations similar to those in Problem 2.6 give

$$\begin{aligned}\Pr\{N_j(\tau_j,\tau_j+s) = r|H_j(\tau_j)\} \qquad & (8.24)\\ = \frac{\Gamma(r+\phi^{-1}+n)}{\Gamma(\phi^{-1}+n)r!} & \frac{\mu(\tau_j,\tau_j+s)^r(\phi^{-1}+\mu(0,\tau_j))^{\phi^{-1}+n}}{(\phi^{-1}+\mu(0,\tau_j)+\mu(\tau_j,\tau_j+s))^{r+\phi^{-1}+n}},\end{aligned}$$

where $N_j(\tau_j) = n$ and $\mu(v,w) = \int_v^w \rho(s)ds$. This is a negative binomial distribution.

Typically ϕ and $\rho(t)$ are unknown and are estimated from the data $H_i(\tau_i)$, $i = 1, \ldots, m$, as described in Section 3.5.2. In this case insertion of $\widehat{\phi}$ and $\widehat{\rho}(t)$ for ϕ and $\rho(t)$ in (8.24) provides plug-in predictive probabilities or

prediction intervals. Calibrated prediction is also possible, following the prescription given in Section 8.3.2. In Example 8.6, we describe the calculation of calibrated intervals for an application involving car warranty claims.

We remark that prediction within random effects models such as the one here is sometimes called empirical Bayes prediction. Our approach is non-Bayesian, however, and we simply refer to the problem as one involving random effects.

Example 8.5: Prediction of fault detection in software testing

In Section 3.8.3 we fitted a model to data on the testing and debugging of a large software system. The time variable t was the cumulative number of person-days of testing, and data on the numbers of faults detected up to various times t_j $(j = 1, \ldots, k)$ were recorded. The number of lines of code modification C_j over the interval (t_{j-1}, t_j) was also recorded. Because the introduction of new code may introduce new faults, the C_j were considered as covariates in the model. Based on a fit of the model (3.57), the number of software faults N_{k+1} that would be detected if testing were continued indefinitely beyond the final testing time $t_k = 1336.7$ (person-days) was predicted. This was done using a plug-in method: conditional on the testing and code modification results up to t_k, N_{k+1} has a Poisson distribution with mean μ_{k+1} given by (3.62). The plug-in method simply uses (3.62) with maximum likelihood estimates inserted for the three parameters α, β, and θ, and gave a .95 prediction interval for N_{k+1} of (170, 225).

A calibrated prediction interval can also be obtained. We ignore the discreteness in N_{k+1}; this should have a small effect because the likely values for N_{k+1} are large. The steps in obtaining the calibrated interval are as follows.

(i) Using the values $\alpha = \widehat{\alpha}, \beta = \widehat{\beta}, \theta = \widehat{\theta}$ in the model (3.57), use (3.59) and (3.60) to give pseudo-data N_j^* $(j = 1, \ldots, k)$; the C_j are assumed to be the same as in the observed data giving $\widehat{\alpha}, \widehat{\beta}$, and $\widehat{\theta}$.
(ii) From the N_j^*, obtain new estimates $\widehat{\alpha}^*, \widehat{\beta}^*, \widehat{\theta}^*$ by maximizing the likelihood (3.61). From these obtain $\widehat{\mu}_{k+1}^*$ using (3.62).
(iii) Generate $y^* \sim$ Poisson $(\widehat{\mu}_{k+1})$, where $\widehat{\mu}_{k+1}$ is given by (3.62) with $\alpha = \widehat{\alpha}, \beta = \widehat{\beta}, \theta = \widehat{\theta}$.
(iv) Compute $U^* = \sum_{y=0}^{y^*} (\widehat{\mu}_{k+1}^*)^y \exp(-\widehat{\mu}_{k+1}^*)/y!$.
(v) Repeat steps (i)–(iv) B times, to get values $U_1^*, \ldots, U_B^*$.
(vi) To get an upper α prediction limit y_u, take the $[\alpha B]$ quantile $U_{[\alpha B]}^*$ of $U_1^*, \ldots, U_B^*$. Then, y_u is chosen to satisfy

$$\sum_{y=0}^{y_u} \frac{\widehat{\mu}_{k+1}^y \exp(-\widehat{\mu}_{k+1})}{y!} = U_{[\alpha B]}^* \,.$$

A two-sided .95 prediction interval can be obtained by using (vi) with $\alpha = .025$ and $\alpha = .975$, respectively. In the present example, the amount of

data on which $\widehat{\alpha}, \widehat{\beta}$, and $\widehat{\theta}$ are based is quite large, and it is not expected that calibration will give an interval much different from the plug-in interval (170, 225). A more serious source of additional uncertainty in the prediction interval is extrapolation of the model (3.57) much beyond t_k. Although (3.57) fits the data well up to time t_k, extrapolation is done on faith, and cannot be checked.

Example 8.6: Prediction of warranty claims

Data on automobile warranty claims were discussed in Sections 1.2.4 and 3.8.4, where a group of over 38,000 vehicles was considered. A common problem is to predict the eventual number of claims across a population of cars, based on warranty data that has accumulated up to some calendar time. The problem can be described as follows. Suppose that the warranty coverage lasts for a period of T days after a car is sold, and let $N_i(t)$ be the number of warranty claims up to t days after sale, for vehicles $i = 1, \ldots, M$. The objective is to predict the total number of claims, $W = \sum N_i(T)$. Cars are sold at different times, and so we consider a calendar time point at which $m \leq M$ cars have been sold. Without loss of generality we denote the cars sold as $i = 1, \ldots, m$, and we let τ_i denote min(T, number of days since car i was sold). Because the manufacturer knows $N_i(\tau_i)$, $i = 1, \ldots, m$, in order to predict W we actually need only to predict

$$S = \sum_{i=1}^{M} N(\tau_i, T)\,, \tag{8.25}$$

given the claim histories $H_i(\tau_i^+)$, $i = 1, \ldots, m$. Note that some cars may have $\tau_i = T$, so that $N(\tau_i, T) = 0$ in (8.23), and that if car i is not yet sold, then $\tau_i = 0$ and there is no claims history.

Fredette and Lawless (2007) address the prediction problem by assuming that claims for car i are conditionally Poisson with a rate function $u_i \rho(t; \theta)$, given a random variable u_i that is gamma-distributed, as in Example 8.2. The terms $N(\tau_i, T)$ in (8.25) with $0 < \tau_i < T$ are, conditional on $H_i(\tau_i)$, negative binomial random variables with probability functions of the form (8.24). Usually M is large, and Fredette and Lawless (2007) consider efficient ways to compute plug-in or calibrated prediction intervals for S by simulation.

Figure 8.2 shows 0.95 prediction intervals for 15,775 cars that are a subset of those considered in Section 3.8.4. Predictions of W are shown at calendar times that are $100, 150, 200, \ldots, 500$ days after the first vehicle was sold. These are shown in Figure 8.2 against the accumulating number of warranty claims for the population of cars, culminating in a total of $W = 2620$ claims by day 571, at which time all 15,775 cars had reached the end of their one-year warranties. Both plug-in and calibrated intervals are shown; the models used have five parameters and there is not a lot of information about some aspects of the parameter vector until about 250–300 days of data have accumulated, so calibration is desirable. We see from Figure 8.2 that early predictions up to

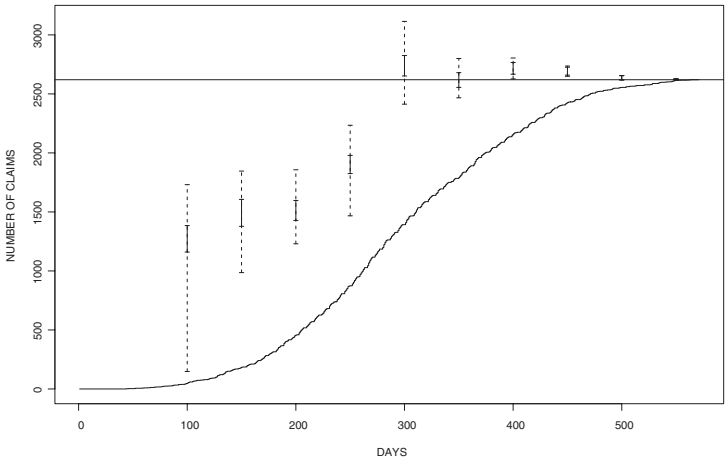


Fig. 8.2. 95% prediction intervals for numbers of claims based on car warranty data; solid lines give plug-in and dotted lines give calibrated intervals.

about 250 days, which are based on data for which τ_i is much less than 365 days for most cars, turn out to be too low. A considerable degree of extrapolation is needed for the rate function $\rho(t;\theta)$ for these early predictions, which are all made at time points where under 40% of the eventual claims have occurred. This illustrates the inherent difficulty in predicting events very far into the future, but it is noteworthy that once about 40% of the claims have occurred, predictions of W are quite accurate. The widths of the prediction intervals necessarily decrease to zero as time goes by, because the portion S of W that has to be predicted approaches zero as all the τ_i approach T. Fredette and Lawless (2007) discuss ways to monitor the accumulating claims data and to check models used for prediction, and are able to improve the earlier predictions slightly.

8.4 Recurrent Events in Randomized Trials

8.4.1 Specification and Testing of Treatment Effects

In many areas of health research the aim is to assess whether a new treatment or intervention has an effect on the occurrence of undesirable clinical events. The preferred design for making inferences about causal effects of this sort is the randomized clinical trial. Three widely discussed consequences of randomization are i) the ability to carry out tests based on the randomization distribution, ii) the creation of balance across treatment groups in the

distribution of baseline covariates, and iii) mitigation of the effect of confounding factors in marginal comparisons between treatment groups. Consequence i) is useful but not essential given the many robust procedures available for analyses, but ii) and iii), which are closely related, are of central importance because they support causal inferences about treatment effects.

The specification and testing of treatment effects in randomized studies have been discussed and illustrated in several chapters. Here we recapitulate some of the important points and provide a more extensive discussion.

As indicated in Section 1.3.5, it is particularly important in clinical trials that models be formulated so that treatment effects are easily interpreted and understood. This ensures that the implications of using the experimental treatment instead of the control treatment will be clear, and it facilitates comparisons across similar studies. In settings involving very few events per subject it may be reasonable to focus simply on time to the first event. When events occur more frequently, however, the preferred approach is to express treatment effects in terms of marginal features of the full event processes, such as rates of events or expected numbers of events.

Methods based on rate and mean functions, discussed extensively in Chapter 3, generally offer the simplest specification of treatment effects for recurrent events. The mixed Poisson model is a natural framework for rate or mean function analyses and we consider it here. For simplicity let x_i be a binary treatment indicator, with $x_i = 1$ denoting an experimental treatment and $x_i = 0$ a control treatment for the ith subject. Let u_i denote a subject-specific random effect which represents heterogeneity across subjects; we restrict u_i to have a finite mean, and without loss of generality assume $E(u_i) = 1$. Multiplicative models (3.28), with

$$E\{dN_i(t)|u_i, x_i\} = u_i \rho_0(t) \exp(\beta x_i) dt\,,$$

express treatment effects as relative rates $\exp(\beta)$ of events for experimental versus control subjects. This formulation has a number of convenient features:

i. Relative rates are easily understood, and their interpretation does not depend on the Poisson assumption, or any specific type of event process.

ii. The treatment effect also applies to expected numbers of events, because

$$E\{N_i(t)|u_i, x_i\} = u_i \mu_0(t) \exp(\beta x_i)\,.$$

iii. Subject-specific and population-average relative rates are the same. That is,

$$E\{dN_i(t)|x_i\} = \rho_0(t) \exp(\beta x_i) dt\,,$$

so $\exp(\beta)$ represents both subject-specific and population-average effects of treatment.

iv. Simple robust methods of estimating β or testing that $\beta = 0$ are available (see Section 3.6 and 3.7.5).

v. Randomization of treatment implies that u_i is independent of x_i, but the model does assume that the rate function has a multiplicative form. This can readily be checked (see Section 3.7.2).

vi. The multiplicative model is easily extended to allow for a time-varying treatment effect through the introduction of interaction terms involving treatment and specified functions of time, or simply by replacing $\rho_0(t)\exp(\beta x)$ with $\rho_{0x}(t)$ in the specifications above. The treatment effect, although time-varying, is still easily interpreted as a ratio of event rate functions.

The key point in the preceding discussion is the specification of treatment effects in terms of rate functions. Additive models, discussed briefly in Section 3.4.4, can similarly provide robust and easily interpreted effects. We also note that event-dependent dropouts or event-dependent terminating events, which arise in some clinical studies, make interpretation of treatment effects more complex but that rate and mean function specifications remain convenient, as discussed in Section 6.6 and Section 7.2.

Although rate and mean functions offer significant advantages for the specification and testing of treatment effects in many settings, there are nevertheless situations where analyses based on gap times seem most natural. These include processes where a physical renewal occurs after the occurrence of an event, or where an event such as an infection or hospitalization triggers a temporary period of treatment during which a subject is not at risk for a new event. Examples of such processes in clinical trials have been discussed in Sections 4.3.2, 4.4.3, 5.5.1, and 6.7.2. As discussed in Section 4.2, it is important to allow for heterogeneity of subjects or association between their gap times. Cox proportional hazards models conditional on random effects are readily handled as in Section 4.2, and are attractive in being semiparametric. However, unlike the multiplicative rate function models, the subject-specific and population-average effects differ, and the population-average effects are not in general of proportional hazards form. Accelerated failure time or log-location-scale models, considered in Sections 4.2.2 and 4.2.3, do give the same subject-specific and population-average effects, but rely slightly more on a specific joint gap time distribution. Semiparametric multivariate accelerated failure time methods are more difficult to apply. It is easy to allow changes in successive marginal gap time distributions, but harder to allow for continuous time-on-study effects than in rate-based models. Event-dependent terminating events or loss to followup are also harder to accommodate.

It is also important to note that randomization at study entry is not sufficient to facilitate causal inferences about treatment effects on the second and subsequent gaps, when individuals may be differentially selected for inclusion in analyses of second and subsequent gap times. That is, the covariate distributions will be similar between groups for the first gap analysis due to randomization, but if not all subjects experience a first event this will not be the

case for second and subsequent gaps. Unless the gap time models capture the effects of all covariates, estimated treatment effects will be difficult to interpret. It has therefore been suggested in some studies where gap time analyses are scientifically most relevant, that patients be rerandomized to treatment at the occurrence of each event. This provides a basis for valid tests of treatment effects for second and subsequent gaps, but it remains difficult to characterize clearly the target population necessary to interpret findings.

In some settings, the effect of treatment may change with each passing event. This may be due to the underlying nature of the process, or it may occur if physicians treating patients intervene upon the occurrence of an event. Intensity-based models, which condition at any given time on previous event history, may be useful to gain insight into process dynamics, but they are much less suitable for the primary analysis of treatment effects because the previous event history is potentially responsive to treatment (Yusuf et al., 1991). Treatment effects expressed in such models are difficult to interpret and heavily dependent on model assumptions.

There may not be a single superior way to express a treatment effect. The key requirements are that specification and analysis of an effect be scientifically relevant and statistically sound. It is often found that quite different models provide adequate descriptions of the data, especially when the number of events per subject is small. Treatment effects, and their estimates depend, however, on the model specified. There has often been confusion about this point in the literature. A few authors have considered the effects of model misspecification on certain definitions of effects (e.g. Box-Steffensmeier and De Boef, 2006; Boher and Cook, 2006; Metcalfe and Thompson, 2006).

Often we wish initially to test a null hypothesis of no treatment effect, which means the problem is one of testing that two recurrent event processes are the same. Tests based on different models can be considered and their efficiency and robustness properties assessed. Even though estimated treatment effects under alternative models may differ substantially, it is often found that the evidence against a null hypothesis of no effect is similar across models which fit the data well. Robust tests based on rate function models discussed in Section 3.7.5 perform well under a variety of scenarios (e.g. Cook et al., 1996; Boher and Cook, 2006) and we consider such models as a basis for sample size selection in the following section. Robust tests based on marginal survival models for the times $T_1, T_2, \ldots$ of successive events (Wei et al., 1989; Section 3.6.5) also give valid tests of the null hypothesis of no treatment differences for a wide class of processes. It is important, however, for tests to be based on plausible underlying models for the event process and in the recurrent event setting this is not the case for these methods. In this framework, individuals are considered at risk for their kth event from the start of followup (i.e. even before their $(k-1)$st event). This counterintuitive risk set definition leads to uninterpretable regression coefficients in the recurrent event setting (Boher and Cook, 2006), which makes it difficult to quantify the impact of treatment.

8.4.2 Trial Design for Mixed Poisson Processes

Consider a clinical trial in which m individuals are randomly assigned with probability 0.5 to either a group receiving an experimental treatment or a control group. If individual i is in the treatment group $x_i = 1$, and $x_i = 0$ otherwise. Each individual is at risk of a recurrent event, and $\{N_i(t), 0 \le t\}$ denotes the counting process of events for individual i. Individuals are to be followed over the interval $[0, \tau]$, but some may withdraw from the study early. Let C_i denote the withdrawal time for individual i, so that $\tau_i = \min(C_i, \tau)$ is their right censoring time. Let $N_i(\tau_i) = n_i$ represent the total number of events observed for individual i, and $t_{i1}, \ldots t_{in_i}$ the respective event times over $[0, \tau_i]$.

Given u_i, assume $\{N_i(t), 0 \le t\}$ is a Poisson process with rate function $u_i \rho_0(s) \exp(\beta_1 x_i)$, where $\rho_0(s)$ is a baseline rate function, β_1 is the log relative rate reflecting the treatment effect, and u_i is a subject-specific effect. As in Section 3.5 we assume the u_i are independent and identically distributed gamma random variables with $E(u_i) = 1$ and $\text{var}(u_i) = \phi$. In many clinical studies the event rate is approximately constant and we consider this important setting here. Under this time-homogeneous model where $\rho_0(s) = \rho_0$, the log-likelihood from the resulting marginal negative binomial model is $\ell(\theta) = \sum_{i=1}^{m} \ell_i(\theta)$, where $\ell_i(\theta)$ is

$$\sum_{k=0}^{n_i - 1} \log(1 + k\phi) + n_i(\beta_0 + \beta_1 x_i) - (\phi^{-1} + n_i) \log(1 + \phi \exp(\beta_0 + \beta_1 x_i)\tau_i)\,,$$

with $\beta_0 = \log \rho_0$, $\beta = (\beta_0, \beta_1)'$, and $\theta = (\beta', \phi)'$. Notice that this log-likelihood does not depend on the actual event times under the time-homogeneous model, but only the event counts, the duration of followup, and the treatment indicator. The score vector is denoted $U_\theta(\theta) = \partial \ell(\theta)/\partial \theta = (U'_\beta(\theta), U_\phi(\theta))'$, where $U_\beta(\theta) = \partial \ell(\theta)/\partial \beta$ and $U_\phi(\theta) = \partial \ell(\theta)/\partial \phi$, and we let $I(\theta) = -\partial^2 \ell(\theta)/\partial \theta \partial \theta'$ and $\mathcal{I}(\theta) = E\{I(\theta)\}$ denote the observed and expected information matrices, respectively. From Section 3.5.2, one can show that β is orthogonal to ϕ (i.e. $\mathcal{I}_{\beta\phi}(\theta) = 0$) and hence an asymptotic covariance matrix for $\widehat{\beta}$ is $\mathcal{I}^{-1}_{\beta\beta}(\theta)$, where $\mathcal{I}_{\beta\beta}(\theta)$ is the 2×2 submatrix of $\mathcal{I}(\theta)$ conformable with β. This gives

$$\text{asvar}\{\sqrt{m}(\widehat{\beta}_0 - \beta_0)\} = \left\{ \frac{\exp(\beta_0) E(\tau | X = 0)}{1 + \phi \exp(\beta_0) E(\tau | X = 0)} \right\}^{-1}, \tag{8.26}$$

$$\text{asvar}\{\sqrt{m}(\widehat{\beta}_1 - \beta_1)\} = \sum_{X=0}^{1} \left\{ \frac{\exp(\beta_0 + \beta_1 X) E(\tau | X)}{1 + \phi \exp(\beta_0 + \beta_1 X) E(\tau | X)} \right\}^{-1}, \tag{8.27}$$

where we view $\tau_1, \ldots, \tau_m$ as i.i.d. random variables. When planning studies we typically assume independent nondifferential (i.e. independent of treatment) censoring and adopt convenient distributions for C (e.g. exponential), to facilitate calculation of the expectations in (8.26) and (8.27).

If α_1 and α_2 denote the desired type I and type II error rates of a two-sided test of $H_0: \beta_1 = \beta_{10}$, where the alternative value is $\beta_1 = \beta_{1A}$, then one needs to find the minimum sample size m that satisfies

$$m > \frac{[\text{asvar}_0(\sqrt{m}(\widehat{\beta}_1 - \beta_{10}))Z_{\alpha_1/2} + \text{asvar}_A(\sqrt{m}(\widehat{\beta}_1 - \beta_{1A}))Z_{\alpha_2}]^2}{(\beta_{10} - \beta_{1A})^2}, \quad (8.28)$$

where $\text{asvar}_0(\sqrt{m}(\widehat{\beta}_1 - \beta_{10}))$ and $\text{asvar}_A(\sqrt{m}(\widehat{\beta}_1 - \beta_{1A}))$ denote the variance given by (8.27) under the null and alternative hypotheses, respectively, and Z_p represents the $p-$quantile for a standard normal distribution. Values for β_{10} and β_{1A} are needed, as well as provisional values for the other parameters in the model and the censoring distribution. Often $\beta_{10} = 0$ so that the null hypothesis is that the events occur at the same rate in the treatment and control groups.

If the model is appropriate and the parameter values specified in (8.28) are close to the true values, then the objective of power $1 - \alpha_2$ at $\beta_1 = \beta_{1A}$ will be met. Reports on previous studies, however, may not provide adequate information to approximate ϕ well when mixed Poisson models are appropriate. If ϕ is larger than the value specified in (8.28), then the actual power will be less than the desired level. When there is considerable uncertainty in the value of ϕ or other parameters, it is advisable to specify a range of values and then select the largest sample size resulting from the parameter value configurations. Alternatively, one can consider adaptive designs which carry out sample size re-estimation periodically throughout the course of the study. Much work has been done for binary and continuous reponses, but this warrants study for recurrent event responses.

Finally, although sample size calculations are often based on Wald-type test statistics, as here, actual analyses may be based on other approaches such as the robust pseudo-score tests of Section 3.7.5. The frequency properties of different tests are generally comparable for large samples, so sample size choices based on (8.28) can still be employed.

8.4.3 Use of Baseline Count Data

Clinical trials often incorporate a period of observation in which subjects are monitored prior to randomization to treatment, in order to collect information on their baseline level of disease activity. This gives a baseline response which in the context of recurrent event data may represent the number of times the clinical event occurred over a specified period preceding randomization. Examples include premature ventricular contractions in cardiology trials, epileptic seizures in epilepsy trials, and respiratory studies in which the baseline counts represent the number of asthma attacks in the preceding year. In some cases (see Section 7.3) subject screening and selection may be related to a baseline response.

Suppose the study consists of m subjects. Let τ_1 denote the common duration of observation prior to randomization and N_{i1} the baseline counts for subjects $i = 1, \ldots, m$. Suppose τ_2 is the common duration of observation after randomization and N_{i2} the corresponding count. As in the preceding section, we focus on settings where mixed Poisson processes are applicable, and where event rate functions are constant. To allow heterogeneity of subjects and association between N_{i1} and N_{i2} we thus suppose that given a subject-specific random effect u_i, N_{i1} and N_{i2} are independent Poisson variables with

$$\Pr(N_{i1} = n_{i1}|u_i; \rho_1) = \frac{(u_i\rho_1\tau_1)^{n_{i1}} \exp(-u_i\rho_1\tau_1)}{n_{i1}!}, \tag{8.29}$$

and

$$\Pr(N_{i2} = n_{i2}|u_i; \rho_2, \beta) = \frac{(u_i\rho_2 \exp(\beta x_i)\tau_2)^{n_{i2}} \exp(-u_i\rho_2 e^{\beta x_i}\tau_2)}{n_{i2}!}, \tag{8.30}$$

where ρ_1 is the rate of events before randomization, ρ_2 is the rate of events among control patients after randomization, and $\rho_2 \exp(\beta)$ is the rate of events in the treated group after randomization. Without loss of generality we take $\tau_1 = \tau_2$, $\mu_1 = \rho_1\tau$, $\mu_2 = \rho_2\tau$, and $\mu_{i2} = \mu_2 \exp(x_i\beta)$, $i = 1, \ldots, m$. We further assume that the u_i are independently gamma distributed with mean 1 and variance ϕ. Marginalizing over the random effect then gives a negative binomial model for the count after randomization,

$$\begin{aligned}\Pr(N_{i2} = n_{i2}; \rho_2, \beta, \phi) = {} & \frac{\Gamma(\phi^{-1} + n_{i2})}{\Gamma(\phi^{-1})n_{i2}!} \\ & \times \left(\frac{1}{1+\mu_{i2}\phi}\right)^{\phi^{-1}} \left(\frac{\mu_{i2}\phi}{1+\mu_{i2}\phi}\right)^{n_{i2}},\end{aligned} \tag{8.31}$$

where $n_{i2} = 0, 1, \ldots$. A negative trinomial model is obtained for (N_{i1}, N_{i2}), with

$$\begin{aligned}\Pr(N_{i1} = n_{i1}, N_{i2} = n_{i2}; \theta) = {} & \frac{\Gamma(\phi^{-1} + n_{i1} + n_{i2})}{\Gamma(\phi^{-1})n_{i1}!n_{i2}!} \\ & \times \frac{\mu_1^{n_{i1}}\mu_{i2}^{n_{i2}}\phi^{n_{i1}+n_{i2}}}{(1+(\mu_1+\mu_{i2})\phi)^{\phi^{-1}+n_{i1}+n_{i2}}},\end{aligned} \tag{8.32}$$

where $\theta = (\rho_1, \rho_2, \beta, \phi)'$. From (8.31) and (8.32) we obtain

$$\begin{aligned}\Pr(N_{i2} = n_{i2}|n_{i1}; \theta) = {} & \frac{\Gamma(\phi^{-1} + n_{i1} + n_{i2})}{\Gamma(\phi^{-1} + n_{i1})n_{i2}!} \\ & \times \frac{(1+\mu_1\phi)^{\phi^{-1}+n_{i1}}(\mu_{i2}\phi)^{n_{i2}}}{(1+(\mu_1+\mu_{i2})\phi)^{\phi^{-1}+n_{i1}+n_{i2}}},\end{aligned} \tag{8.33}$$

and it can be shown that a likelihood based on (8.33) contains the same information on β as a likelihood based on (8.32).

Another distribution of interest is obtained by conditioning on $n_{i1} + n_{i2}$ to eliminate u_i, $i = 1, \ldots, m$. Straightforward calculations give

$$\Pr(n_{i1}, n_{i2} | n_{i1} + n_{i2}; \alpha, \beta) = \binom{n_{i1} + n_{i2}}{n_{i2}} (1 - \pi_i)^{n_{i1}} \pi_i^{n_{i2}}, \tag{8.34}$$

where $\pi_i = \exp(\alpha + \beta x_i)/(1 + \exp(\alpha + \beta x_i))$ and $\alpha = \log(\mu_2/\mu_1)$. This suggests using a logistic regression analysis based on the binomial distributions (8.34). The fact that u_i is eliminated means that there is no need to make distributional assumptions regarding the random effect.

We can compare the use of different distributions above for inferences about β. Note that (8.31) is indexed by ρ_2, β, and ϕ, (8.32) by ρ_1, ρ_2, β, and ϕ, and (8.34) by α and β. We refer to these respective models as marginal, joint, and conditional models, and we let $\widehat{\beta}^M$, $\widehat{\beta}^J$, and $\widehat{\beta}^C$ denote the estimators obtained by maximizing the respective likelihood functions. The expected information matrices arising from these likelihoods lead to the following expressions for asymptotic variances:

$$\text{asvar}(\sqrt{m}(\widehat{\beta}^M - \beta)) = \left\{ \frac{1}{\rho_2} + \frac{1}{\rho_2 \exp(\beta)} + 2\phi \right\} \tag{8.35}$$

$$\text{asvar}(\sqrt{m}(\widehat{\beta}^J - \beta)) = \left\{ \frac{1}{\rho_2} + \frac{1}{\rho_2 \exp(\beta)} + \frac{2\phi}{1 + \phi\rho_1} \right\} \tag{8.36}$$

$$\text{asvar}(\sqrt{m}(\widehat{\beta}^C - \beta)) = \left\{ \frac{1}{\rho_2} + \frac{1}{\rho_2 \exp(\beta)} + \frac{2}{\rho_1} \right\}. \tag{8.37}$$

Based on these asymptotic variances, the relative efficiency of the marginal to joint analysis is

$$\begin{aligned} RE^{M:J} &= \frac{\text{asvar}(\sqrt{m}(\widehat{\beta}^J - \beta))}{\text{asvar}(\sqrt{m}(\widehat{\beta}^M - \beta))} \\ &= \frac{(1 + \exp(\beta)) + 2\phi\rho_2 \exp(\beta)/(1 + \phi\rho_1)}{1 + \exp(\beta) + 2\phi\rho_2 \exp(\beta)}, \end{aligned} \tag{8.38}$$

and for the conditional to joint analysis is

$$\begin{aligned} RE^{C:J} &= \frac{\text{asvar}(\sqrt{m}(\widehat{\beta}^J - \beta))}{\text{asvar}(\sqrt{m}(\widehat{\beta}^C - \beta))} \\ &= \frac{\rho_1(1 + \exp(\beta))(1 + \phi\rho_1) + 2\phi\rho_1\rho_2 \exp(\beta)}{\rho_1(1 + \exp(\beta))(1 + \phi\rho_1) + 2\phi\rho_1\rho_2 \exp(\beta) + 2\rho_2 \exp(\beta)} \end{aligned} \tag{8.39}$$

which are both always ≤ 1. Thus a joint analysis of the baseline and postrandomization counts is asymptotically more efficient than either a strictly marginal analysis based only on the postrandomization count, or a conditional

analysis based on (8.34). If there is evidence to suggest the joint model is correct it is therefore recommended for use on the grounds of efficiency. If more robust methods are desired, then one can entertain either the robust methods of Section 3.6 for inferences, or the conditional analysis of (8.34), which does not require a random effect distribution, but does rely on the Poisson assumption.

To provide further guidance between these two options, note that

$$RE^{M:C} = \frac{\text{asvar}(\sqrt{m}(\widehat{\beta}^C - \beta))}{\text{asvar}(\sqrt{m}(\widehat{\beta}^M - \beta))} = \frac{\rho_1(1+\exp(\beta)) + 2\rho_2\exp(\beta)}{\rho_1(1+\exp(\beta) + 2\phi\rho_2\exp(\beta))} \quad (8.40)$$

which reveals that whenever $\phi\rho > 1$ the conditional analysis yields a more efficient estimator of β than the marginal analysis, within the class of negative binomial models. Thus in settings with considerable extra-Poisson variation (i.e. $\phi > \rho^{-1}$) it may be preferable to condition on the total event count, eliminate the random effect, and carry out analyses based on a binomial model.

For illustration, suppose we set $\exp(\beta) = 0.75$ to represent a moderate treatment effect, $\tau = 1$, and let $\mu_1 + \mu_2 = 4$ represent a moderate mean for the total number of events among control patients over the baseline and followup observation periods. Figure 8.3 displays a plot of the 60%, 80%, 100%, and 120% relative efficiency contours for β as a function of $\exp(\alpha)/(1+\exp(\alpha)) = \mu_2/(\mu_1+\mu_2)$ and ϕ, for this scenario. The points for which the asymptotic variance under the conditional model is 20% lower than the asymptotic variance under the marginal model are denoted by the 80% relative efficiency contour. The asymptotic variance under the conditional model is 20% greater than the marginal model for points on the 120% contour. From Figure 8.3 it is clear that there is a large region in the parameter space in which the conditional analysis is more efficient than the marginal analysis and this region represents scenarios which one might encounter in many biomedical settings. As the baseline mean becomes small (i.e. $\rho_1 \to 0$) the marginal analysis leads to more efficient estimates than the conditional analysis, even when ϕ is large. Moreover, with very small ϕ, the marginal analysis is generally preferred. When the mean number of events in the followup period is comparable to or smaller than the mean number of events in the baseline period (i.e. $\mu_2/(\mu_1+\mu_2) < 0.50$), however, even when extra-Poisson variation is moderate (i.e. $\phi < 1.0$) there can be as much as a 20% lower asymptotic variance under the conditional analysis. As one might expect, for any given α the gains from the conditional analysis become more substantial as ϕ increases.

Note that although (8.34) was derived under the assumption of a mixed Poisson model, estimating functions based on it are valid more generally. The score equations for α and β from (8.34) are

$$U_\alpha(\theta) = \sum_{i=1}^{m}(n_{i2} - (n_{i1}+n_{i2})\pi_i) \quad (8.41)$$

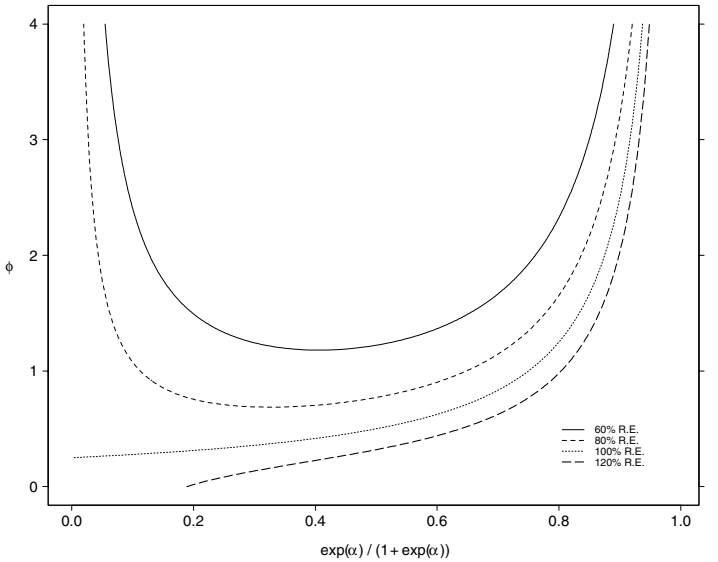


Fig. 8.3. Contour plots of the relative efficiency of conditional versus marginal analyses under a mixed Poisson model ($\mu_1 + \mu_2 = 4$, $\exp(\beta) = 0.75$).

$$U_\beta(\theta) = \sum_{i=1}^{m} (n_{i2} - (n_{i1} + n_{i2})\pi_i) x_i \,, \tag{8.42}$$

and because $E(N_{i1}|x_i) = \mu_1$ and $E(N_{i2}|x_i) = \mu_2 \exp(\beta x_i)$, we see that $E\{U_\alpha(\theta)\} = 0$ and $E\{U_\beta(\theta)\} = 0$ as long as the mean specifications hold. Thus (8.41) and (8.42) are unbiased and the estimate $\widehat{\beta}$ can be interpreted beyond the mixed Poisson model. Estimates of α and β may be obtained by using logistic regression software, but robust variance estimates should be obtained as described in Appendix A.

Finally we remark that data during a baseline period are often retrospectively determined in order to shorten the length of the study, but in such settings the baseline data may be inaccurate, particularly if subjects are recruited from diverse environments. It is best to prospectively observe patients over a baseline period of observation that commences after a patient has first been contacted. Then the baseline data are as reliable as the data recorded on treatment, because the same event definitions and conventions for recording data can be employed.

8.4.4 Interim Monitoring with Recurrent Events

Long-term clinical trials warrant careful monitoring to ensure patient accrual is on schedule, compliance rates are acceptable, and there are no major concerns regarding adverse effects. Periodic assessments of the treatment effect

on the primary outcome are also often carried out and formal stopping rules may be employed, based on interim results, to facilitate study termination as soon as one treatment can be declared superior. The decision to stop or continue a trial is typically influenced by whether the test statistics cross a "stopping boundary" defined by critical values for the interim tests.

Group sequential methods are a class of procedures for repeated significance testing based on accumulating data. A key feature of these methods is that an overall type I error is preserved by using suitable critical values for the interim and final tests. Consider a study with $G-1$ interim analyses scheduled in addition to the final analysis. Let c_g denote the critical value at stage g so that if the test statistic exceeds c_g, the study will be stopped and a treatment difference declared; otherwise, the trial is continued. If the study is not stopped before the final analysis, the null hypothesis is not rejected unless the final test statistic exceeds c_G. It is necessary to derive the critical values $c_1, \ldots, c_G$ such that under the null hypothesis the probability of rejection over the course of the study is equal to the nominal type I error rate.

Procedures due to Pocock (1977) and O'Brien and Fleming (1979) are often used, but the flexible Lan and DeMets (1983) procedure is perhaps the most widely adopted for event time data. Jennison and Turnbull (2000) give comprehensive coverage of the issues and methods for designing and analyzing group sequential trials. Here we briefly outline how interim monitoring can be carried out for trials with recurrent event outcomes, based on the robust two-sample test statistics of Section 3.7.5.

Figure 8.4 is a Lexis diagram in which the $x-$axis represents calendar time; the left side of the axis represents the calendar time at the start of the study and the right side represents the end of the study. There are ten lines with slope one, with each line corresponding to one hypothetical individual recruited to the study at the time of his $x-$intercept. The length of the projection of these lines onto the axes reflects his total time on study. The projection of the dots (representing the occurrence of clinical events of interest) onto the $x-$axis represents the calendar time of their occurrence, and the projection onto the $y-$axis represents the time since study entry of their occurrence. For the sake of simplicity we let v_1, v_2, and v_3 denote three calendar times at which analyses are to be carried out: v_1 and v_2 represent calendar times of "interim" analyses and v_3 the end of study. Slightly more general notation is defined in what follows so that we can indicate the data and associated test statistics available midway through the trial.

Consider a trial with two treatment groups as in Section 3.7.5, and let $\{N_{ki}(t), 0 \leq t\}$ denote the counting process for individual i in group k, where $E\{N_{ki}(t)\} = \mu_k(t)$, $i = 1, 2, \ldots, m_k$; $k = 1, 2$. When data are to be analyzed at multiple points in calendar time, it is convenient to let $Y_{ki}(t; v) = 1$ if individual i in treatment group k is at risk of events at study time t and calendar time v. Let $\tau_{ki}(v)$ denote the length of followup for individual i in group k at calendar time v. Note that $Y_{ki}(t; v)$ could be zero because individual i had not yet been recruited, or because $t > \tau_{ki}(v)$. The test statistic (3.52)

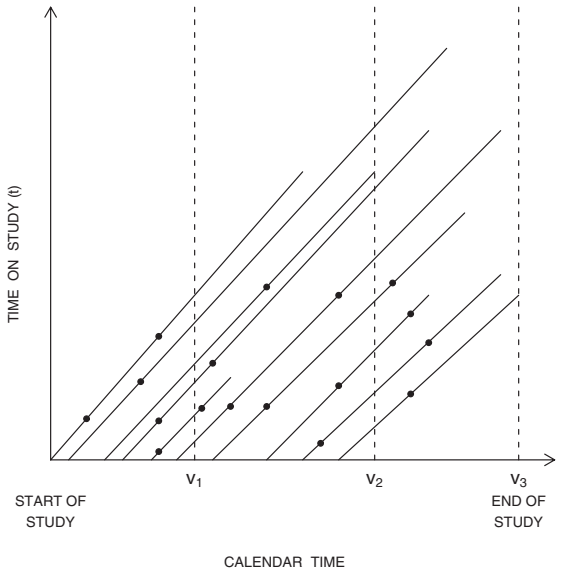


Fig. 8.4. Calendar time and study time for trial with recurrent events, variable recruitment times, and interim monitoring at times v_1 and v_2.

computed at calendar time v is then written as

$$U(\tau(v)) = \int_0^{\tau(v)} w(u;v)\left\{d\widehat{\mu}_2(u;v) - d\widehat{\mu}_1(u;v)\right\},$$

where

$$d\widehat{\mu}_k(u;v) = \frac{\sum_{i=1}^{m_k} Y_{ki}(u;v)dN_{ki}(u)}{\sum_{i=1}^{m_k} Y_{ki}(u;v)},$$

$w(u;v) = Y_{1\cdot}(u;v)Y_{2\cdot}(u;v)/Y_{\cdot\cdot}(u;v)$, and $\tau(v) = \max(\tau_{ki}(v))$, $i = 1, 2, \ldots, m_k$; $k = 1, 2$. The analogous expression for the variance estimate given in (3.55) is

$$\widehat{\mathrm{var}}_R\left\{\sqrt{m}^{-1}U(\tau(v))\right\} = \frac{1}{m}\sum_{k=1}^{2}\sum_{i=1}^{m_k}\left\{\int_0^{\tau(v)} w(u;v)\frac{Y_{ki}(u;v)}{Y_{k\cdot}(u;v)}d\widehat{M}_{ki}(u;v)\right\}^2, \quad (8.43)$$

where $d\widehat{M}_{ki}(u;v) = dN_{ki}(u) - d\widehat{\mu}_k(u;v)$ and $m = m_1 + m_2$. A robust covariance $m^{-1}\mathrm{cov}_R\left\{U(\tau(v_1)), U(\tau(v_2))\right\}$ of these pseudo-score statistics at times v_1 and v_2 is given by

$$\frac{1}{m}\sum_{k=1}^{2}\sum_{i=1}^{m_k}\int_0^{\tau(v_1)}\int_0^{\tau(v_2)}\frac{w(u_1;v_1)w(u_2;v_2)Y_{ki}(u_1;v_1)Y_{ki}(u_2;v_2)}{Y_{k\cdot}(u_1;v_1)Y_{k\cdot}(u_2;v_2)} \times \operatorname{cov}\{dN_{ki}(u_1), dN_{ki}(u_2)\}$$

with estimate $m^{-1}\widehat{\operatorname{cov}}_R\{U(\tau(v_1)), U(\tau(v_2))\}$

$$\frac{1}{m}\sum_{k=1}^{2}\sum_{i=1}^{m_k}\int_0^{\tau(v_1)}\int_0^{\tau(v_2)}\frac{w(u_1;v_1)w(u_2;v_2)Y_{ki}(u_1;v_1)Y_{ki}(u_2;v_2)}{Y_{k\cdot}(u_1;v_1)Y_{k\cdot}(u_2;v_2)} \times d\widehat{M}_{ki}(u_1;v_1)d\widehat{M}_{ki}(u_2;v_2)\,. \tag{8.44}$$

Alternative variance and covariance estimates are obtained by replacing $d\widehat{\mu}_k(u;v)$ in (8.43) and (8.44) with the pooled estimate under the null hypothesis,

$$d\widehat{\mu}(u;v) = \frac{\sum_{k=1}^{2}\sum_{i=1}^{m_k} Y_{ki}(u;v)dN_{ki}(u)}{\sum_{k=1}^{2}\sum_{i=1}^{m_k} Y_{ki}(u;v)}\,.$$

The standardized test statistic at a particular interim or final analysis is

$$\bar{U}(\tau(v)) = \sqrt{m}^{-1}U(\tau(v))/\sqrt{\widehat{\operatorname{var}}_R\left\{\sqrt{m}^{-1}U(\tau(v))\right\}}$$

and under the null hypothesis $(\bar{U}(\tau(v_1)),\ldots,\bar{U}(\tau(v_G)))'$ is asymptotically multivariate normal with zero mean, unit variance, and correlation matrix obtained from (8.44).

Suppose the null and alternative hypotheses are $H_0 : \mu_1(t) = \mu_2(t)$ $(0 \le t)$ versus $H_A : \mu_1(t) \ne \mu_2(t)$ $(0 \le t)$, and the study is to have a type I error rate of α. Interim monitoring is facilitated by specifying an "error spending function" (Lan and DeMets, 1983). Error spending functions are monotonically increasing functions, which we denote $\{\alpha^*(s),\ 0 \le s \le 1\}$, such that $\alpha^*(0) = 0$ and $\alpha^*(1) = \alpha$ where α is a desired Type I error rate. Here s indexes how far through the trial you are on some scale. When analyses are planned based on equal increments of "information" (Lan and DeMets, 1989), $\alpha^*(s) = \alpha\log(1-(\mathrm{e}-1)s)$ and $\alpha^*(s) = 2\{1-\Phi(z_{\alpha/2}/s)\}$ generate boundaries similar to Pocock (1977) and the more conservative O'Brien–Fleming (1979) boundaries, respectively. Error spending functions of the form $\alpha^*(s) = \alpha s^p$ $(0 < p < \infty)$ constitute a family with the index p reflecting the conservative nature of the design for early stopping.

Let $A_g = [\bar{U}(\tau(v_g)) < c_g]$ be the event that the test does not lead to rejection of the null hypothesis at stage g, $g = 1, 2, \ldots, G$. The critical values c_g are found recursively such that they satisfy

$$\Pr\left(\bigcap_{\ell=1}^{g-1} A_\ell \cap \bar{A}_g; H_0\right) = \alpha^*(s_g) - \alpha^*(s_{g-1})\,.$$

This probability is evaluated based on the assumption that under H_0, $\bar{U}$ has a mean zero multivariate normal distribution with covariance estimate obtained from (8.44). See Cook and Lawless (1996) for an illustration.

The critical value at the final analysis will be larger than the critical value of a nonsequential design. As a result, there is typically a larger maximum sample size required for trials incorporating interim monitoring than for trials not incorporating stopping rules. For planning purposes, it is convenient to adopt a particular parametric model which can lead to an expression for the covariance matrix of the test statistics at interim analyses. Jiang (1999) considers a mixed Poisson model as a useful framework for planning a study and indicates how to derive stopping boundaries and how to carry out sample size estimation. Jiang (1999) also considers monitoring based on repeated confidence intervals for a regression coefficient in a multiplicative model, which is particularly appealing for use in sequential noninferiority studies. Finally, we note that given the many external factors that may influence the decision to terminate or continue a trial, these stopping rules are really guidelines (DeMets, 1984) that are put in place to improve the efficiency of the study and address ethical and financial concerns.

8.5 Clustered Data

Sometimes individuals experiencing recurrent events fall naturally into clusters, and if they are heterogeneous, it is desirable to account for this in the analysis. From another point of view, individuals in this setting may display within-cluster association. It should perhaps be noted that the distinction between cluster sampling, where clusters are the primary sampling units and individuals are sampled or observed within clusters, and stratified sampling, where individuals are sampled from each of a specified set of strata, is sometimes blurred in the literature on event history analysis. For example, in multicenter studies the centers are sometimes treated as clusters, but they are often more properly thought of as strata. In any case, methods for stratified data have been considered in previous chapters, and here we discuss clustered data, in which observed clusters may be considered a random sample from some population of clusters. In some settings the clusters selected for observation may be the entire population of physical clusters, but it may still be convenient to consider these as a sample from some conceptual population. An example of clustering is where vehicles under warranty are selected from a number of geographical regions, which are selected from a larger number of regions; vehicles in the same region may display association in certain types of warranty claims because of common environmental factors.

We assume that independent clusters $i = 1, \ldots, m^c$ are selected and that cluster i consists of m_i individuals, each of whom may experience recurrent events. There are two main approaches, one being to use cluster-specific random effects and the other being to use marginal models while accounting

for within-cluster association in event occurrences. We consider the random effects approach first, and for illustration suppose that we are primarily interested in individual event rates and the possible effects of external covariates $x_{ij}(t)$ for individual j in cluster i.

A convenient approach is to define individual random effects

$$u_{ij} = v_i w_{ij} \qquad j = 1, \ldots, m_i;\ i = 1, \ldots, m^c\,,$$

where the v_i and w_{ij} are mutually independent random variables with $v_i \sim G_v(\cdot)$, $w_{ij} \sim G_w(\cdot)$, and $E(v_i) = E(w_{ij}) = 1$, $\mathrm{var}(v_i) = \phi_v$ and $\mathrm{var}(w_{ij}) = \phi_w$. In addition, we assume that given the u_{ij}, the individual event processes are independent Poisson processes with rate functions $u_{ij}\rho_{ij}(t)$, where the $\rho_{ij}(t)$ may include covariates. If individual j in cluster i has n_{ij} events at times t_{ijk} $(k = 1, \ldots, n_{ij})$ over $[0, \tau_{ij}]$, then the conditional likelihood contribution from the individuals in cluster i, given the u_{ij}, is

$$\prod_{j=1}^{m_i} \left\{ \prod_{k=1}^{n_{ij}} v_i w_{ij} \rho_{ij}(t_{ijk}) \right\} \exp\left\{ -v_i w_{ij} \mu_{ij}(\tau_{ij}) \right\}\,, \tag{8.45}$$

where $\mu_{ij}(t) = \int_0^t \rho_{ij}(s)ds$. The observable marginal likelihood is obtained by integrating (8.45) with respect to v_i and $w_{i1}, \ldots, w_{im_i}$. A convenient assumption is that v_i and the w_{ij} have gamma distributions, in which case the densities are of the form (2.28). We then find the likelihood contribution for cluster i, L_i, to be

$$\int_0^\infty \left\{ \prod_{j=1}^{m_i} \prod_{k=1}^{n_{ij}} \rho_{ij}(t_{ijk}) \frac{\Gamma(n_{ij} + \phi_w^{-1}) \phi_w^{n_{ij}} v_i^{n_{ij}}}{\Gamma(\phi_w^{-1})(1 + \phi_w v_i \mu_{ij})^{n_{ij} + \phi_w^{-1}}} \right\} g(v_i; \phi_v) dv_i\,, \tag{8.46}$$

where $g(v_i; \phi_v)$ is given by (2.28) with $\phi = \phi_v$. The integral in (8.46) can be evaluated numerically and the log-likelihood, which is the sum of terms $\log L_i$ $(i = 1, \ldots, m^c)$ can be maximized using optimization software.

A second convenient approach is to extend the robust rate function methods of Section 3.6. For individual j in cluster i we consider the semiparametric rate function model

$$E\left\{ dN_{ij}(t) | x_i^{(\infty)} \right\} = \rho_0(t) dt \exp(x_i'(t)\beta)\,, \tag{8.47}$$

where $\rho_0(t)$ is an unspecified baseline rate function. The Poisson process estimating functions given by (3.21) and (3.23), and used for robust estimation in Section 3.6.3, are readily adapted here. As previously, we require for this that censoring times τ_{ij} be completely independent of the event processes. Then, β may be estimated by solving the equations

$$U(\beta) = \sum_{i=1}^{m^c} \sum_{j=1}^{m_i} \int_0^\tau Y_{ij}(s) W_{ij}(s; \beta) dN_{ij}(s) = 0\,, \tag{8.48}$$

where $\tau = \max(\tau_{ij})$, $Y_{ij}(s) = I(\tau_{ij} \geq s)$, and

$$W_{ij}(s;\beta) = x_{ij}(s) - \frac{\sum_{\ell=1}^{m^c} \sum_{r=1}^{m_\ell} Y_{\ell r}(s) x_{\ell r}(s) \exp(x'_{\ell r}(s)\beta)}{\sum_{\ell=1}^{m^c} \sum_{r=1}^{m_\ell} Y_{\ell r}(s) \exp(x'_{\ell r}(s)\beta)} . \tag{8.49}$$

In addition, $\mu_0(t) = \int_0^t \rho_0(s)ds$ is estimated by

$$\widehat{\mu}_0(t) = \int_0^t \frac{d\bar{N}_{\cdot\cdot}(s)}{\sum_{i=1}^{m^c} \sum_{j=1}^{m_i} Y_{ij}(s) \exp(x'_{ij}(s)\widehat{\beta})} , \tag{8.50}$$

where $d\bar{N}_{\cdot\cdot}(s) = \sum_i \sum_j Y_{ij}(s) dN_{ij}(s)$.

The estimates $\widehat{\beta}$ and $\widehat{\mu}_0(t)$ are consistent provided censoring is independent and the specification (8.47) is correct. When censoring is not independent, inverse probability of censoring weights can be incorporated, as described in Section 7.2. As in Section 3.6.3, robust variance estimates must be obtained for $\widehat{\beta}$ and $\widehat{\mu}_0(t)$; this has been done by Schaubel and Cai (2005a). The estimate for $\widehat{\beta}$ is an extension of (3.38):

$$\widehat{\text{asvar}} \left\{ \sqrt{m^c}(\widehat{\beta} - \beta) \right\} = A^{-1}(\widehat{\beta}) \widehat{B} A^{-1}(\widehat{\beta}) , \tag{8.51}$$

where $A(\beta)$ is given by

$$\sum_{i=1}^{m^c} \sum_{j=1}^{m_i} \int_0^\tau Y_{ij}(s) \left[\frac{S^{(2)}(\beta;s)}{S^{(0)}(\beta;s)} - \frac{S^{(1)}(\beta;s)}{S^{(0)}(\beta;s)} \left\{ \frac{S^{(1)}(\beta;s)}{S^{(0)}(\beta;s)} \right\}' \right] dN_{ij}(s) , \tag{8.52}$$

with $S^{(r)}(\beta;s) = \sum_i \sum_j Y_{ij}(s) x_{ij}^r(s) \exp(x'_{ij}(s)\beta)$ for $r = 0, 1$; $S^{(2)}(\beta;s) = \sum_i \sum_j Y_{ij}(s) x_{ij}(s) x'_{ij}(s) \exp(x'_{ij}(s)\beta)$; and

$$\widehat{B} = \frac{1}{m^c} \sum_{i=1}^{m^c} \widehat{B}_i \widehat{B}'_i , \tag{8.53}$$

where

$$\widehat{B}_i = \sum_{j=1}^{m_i} \int_0^\tau Y_{ij}(s) W_{ij}(s;\widehat{\beta}) \left\{ dN_{ij}(s) - Y_{ij}(s) d\widehat{\mu}_{ij}(s) \right\} ,$$

with $d\widehat{\mu}_{ij}(s) = \widehat{\rho}_{ij}(s)ds = d\widehat{\mu}_0(s) \exp(x'_{ij}(s)\widehat{\beta})$.

Schaubel and Cai (2005a) provide a robust variance estimate for $\widehat{\mu}_0(t)$. They also allow the baseline rate functions to vary across individuals in a cluster, so that $\rho_{0j}(t)$ replaces $\rho_0(t)$ in (8.47). In fact, the function `coxph` in S-PLUS and R gives the robust estimates and variance estimates above, provided we use the `cluster(clusterid)` option, where `clusterid` is a variable indicating the cluster to which an individual belongs. It can also deal with differing $\rho_{0j}(t)$ through the use of the `strata` option.

8.6 Missing Covariate Values

Sometimes the values of certain covariates may be missing for some of the individuals in a study. This may occur by happenstance, but there are also situations where this occurs by design. For example, in two-phase studies (e.g. Lawless et al., 1999) values for covariates that are expensive to measure are often obtained for only a subset of individuals.

The analysis of data with missing values has a large literature and it is not feasible to provide a detailed treatment here. We do, however, outline one approach and give a few key references. Little and Rubin (2002) provide broad coverage and many references. Ibrahim et al. (2005) give a thorough review of missing covariates in generalized linear models.

A variety of methods exists for dealing with missing data. We focus here on maximum likelihood, which is quite general but requires assumptions about the distribution of covariates. Pseudo-likelihood estimation (e.g. Pepe and Fleming, 1991) and estimating functions that use inverse probability weighting (Robins et al., 1994) have also been widely discussed. McLeish and Struthers (2006) survey estimating function methods, and Kalbfleisch and Prentice (2002, Section 11.5) provide some references related to the Cox model in survival analysis. Approaches involving imputation of missing values have also been considered; see Little and Rubin (2002) and the review article by Ibrahim et al. (2005).

Consider fixed covariates and let x_i denote the covariate vector for individual i. Following frequently used terminology, we denote the set of individuals for whom no values in x_i are missing as V (sometimes referred to as the *validation set*) and the set of individuals for whom one or more components of x_i are missing as $\bar{V}$. With a slight abuse of notation, let $x_i = (x_{1i}', x_{2i}')'$ for an individual $i \in \bar{V}$, where x_{1i} is observed and x_{2i} is missing. In some applications all individuals in $\bar{V}$ have the same set of covariate values missing but more generally this may vary, so that (abusing notation) x_{1i} does not necessarily represent the same covariates for different $i \in \bar{V}$. In what follows the missing data mechanism is assumed to be missing at random (MAR) in the sense of Little and Rubin (2002).

Let D_i represent the event history data for individual i, and let $\Pr(D_i|x_i;\theta)$ denote the model of interest, where θ is a parameter vector. For $i \in V$, the contribution to the likelihood function for θ is $\Pr(D_i|x_i;\theta)$, but for $i \in \bar{V}$ it is

$$\Pr\left(D_i|x_{1i};\theta,\psi\right) = E_{X_2}\left\{\Pr\left(D_i|x_{1i},X_2\right)|x_{1i}\right\} . \tag{8.54}$$

This involves the distribution of X_2 given $X_{1i} = x_{1i}$, and ψ in (8.54) denotes parameters that index this distribution. Thus, we cannot avoid considering the distribution of the covariates, as we normally like to do in regression analysis.

To proceed, let $g_i(x_2|x_{1i};\psi)$ denote the probability density or mass function for the missing covariates for individual i, given the observed covariates, and let $f(x;\psi)$ denote the unconditional distribution for the full covariate vector. We can exclude any covariates that are always observed from $f(x;\psi)$

and condition on them instead, but for simplicity we do not indicate this in the notation. The likelihood function takes the form

$$L(\theta,\psi)=\prod_{i\in V}\Pr\left(D_i|x_i;\theta\right)f(x_i;\psi)\cdot\prod_{i\in\bar{V}}\Pr\left(D_i|x_{1i};\theta,\psi\right)f_i(x_{1i};\psi)\,, \quad (8.55)$$

where $f_i(x_{1i};\psi)$ is the marginal density for x_{1i} and where

$$\Pr\left(D_i|x_{1i};\theta,\psi\right)=\int_{-\infty}^{\infty}\Pr\left(D_i|x_{1i},x_2;\theta\right)g_i\left(x_2|x_{1i};\psi\right)dx_2\,.$$

In this situation we consider joint estimation of θ and ψ. The likelihood estimating functions for θ and ψ are

$$U_1(\theta,\psi)=\partial\log L(\theta,\psi)/\partial\theta,\qquad U_2(\theta,\psi)=\partial\log L(\theta,\psi)/\partial\psi$$

and after a little algebra we can write these in the form

$$U_1(\theta,\psi)=\sum_{i\in V}\frac{\partial\log\Pr\left(D_i|x_i;\theta\right)}{\partial\theta} \quad (8.56)$$
$$+\sum_{i\in\bar{V}}\int_{-\infty}^{\infty}w_i\left(x_2;\theta,\psi\right)\frac{\partial\log\Pr\left(D_i|x_{1i},x_2;\theta\right)}{\partial\theta}dx_2$$
$$U_2(\theta,\psi)=\sum_{i\in V}\frac{\partial\log f\left(x_i;\psi\right)}{\partial\psi}+\sum_{i\in\bar{V}}\frac{\partial\log f_i\left(x_{1i};\psi\right)}{\partial\psi} \quad (8.57)$$
$$+\sum_{i\in\bar{V}}\int_{-\infty}^{\infty}w_i\left(x_2;\theta,\psi\right)\frac{\partial\log g_i\left(x_2|x_{1i};\psi\right)}{\partial\psi}dx_2\,,$$

where

$$w_i\left(x_2;\theta,\psi\right)=\frac{\Pr\left(D_i|x_{1i},x_2;\theta\right)g_i(x_2|x_{1i};\psi)}{\int_{-\infty}^{\infty}\Pr(D_i|x_{1i},u;\theta)g_i(u|x_{1i};\psi)du} \quad (8.58)$$

is the conditional density of X_2 given D_i and x_{1i}. Maximum likelihood estimates $\widehat{\theta}$ (and $\widehat{\psi}$) can be obtained using an EM algorithm by starting with initial estimates $\tilde{\theta}_0$ and $\tilde{\psi}_0$, and obtaining $w_i(x_2;\tilde{\theta}_0,\tilde{\psi}_0)$ from (8.58). Then, we solve the equations obtained by setting (8.56) and (8.57) equal to zero, using the fixed values $w_i(x_2;\tilde{\theta}_0,\tilde{\psi}_0)$ in place of $w_i(x_2;\theta,\psi)$; this produces values $\tilde{\theta}_1$ and $\tilde{\psi}_1$. This process is then iterated, with $\tilde{\theta}_1$ and $\tilde{\psi}_1$ used as the new $\tilde{\theta}_0$ and $\tilde{\psi}_0$ at each stage.

The advantage of the procedure above is that (8.56) is a weighted version of the maximum likelihood estimating function for θ based on complete data. To make it feasible, however, we require a tractable and adequate model for the missing covariates. When all covariates x_{1i} in the expressions above are categorical, it is possible to treat the covariate distributions nonparametrically

and to replace the integrals in (8.56)–(8.58) with sums over a finite set of values for x_2. More generally, parametric models such as discrete normal mixtures are tractable in some contexts. Interval estimation and tests for parameters in θ are most easily approached via profile likelihood or likelihood ratio statistics; this merely involves using the EM algorithm with one or more θ parameters held constant. Zhang and Rockette (2006, 2007) and Zhao et al. (2006) provide illustrations in the context of generalized linear models. Chen and Little (1999) consider survival data.

The tractability of inference procedures for missing data depends to some extent on the nature of the data D_i, although the more crucial issue concerns the patterns of covariates. The development above has been parametric; semiparametric models can be treated as special cases but other approaches can also be attractive. There has been little discussion of methods for missing covariates in the context of recurrent events, and research is warranted.

8.7 Covariate Measurement Error

Covariates may in some settings be measured with a significant degree of error. A large literature and a variety of methods have evolved for the following problem. Suppose that a response variable Y has probability density or mass function $f(y|x;\beta)$, where $x = (x_1', x_2')'$ is a vector of covariates in which x_1 is measured accurately but x_2 is not, for at least some individuals. Instead, for certain individuals an imprecise or surrogate measurement Z of x_2 is available.

If there is good reason to trust the model $f(y|x;\beta)$ then it is of interest to estimate β on the basis of data on y, x_1, and x_2 and, if x_2 is unavailable, on data for y, x_1, and z. It is typically assumed that $\Pr(Y|x_1, x_2, z) = \Pr(Y|x_1, x_2)$, and we assume this here. How one may proceed depends on what type of data is available. If observations $(y_i, x_{1i}, x_{2i}, z_i)$ are available for individuals i in some set V (again, often called the *validation set*), with observations (y_i, x_{1i}, z_i) for individuals $i \in \bar{V}$, then methods for missing data as discussed in the preceding section can be applied, by identifying (x_1', z') here with x_1 in Section 8.6 and restricting the regression coefficient for z to be zero. (This assumption can be checked by including the coefficient in the model.) In this case, assumptions about the distribution of X_2 given X_1 and Z are required.

Carroll et al. (2006) provide extensive discussion of measurement error problems and present a number of alternative approaches. There has been very little discussion of measurement error methodology for problems involving recurrent events, exceptions being Turnbull et al. (1997) and Jiang et al. (1999). There has been a good deal of development for survival analysis, and in particular for the semiparametric Cox model. Kalbfleisch and Prentice (2002, Section 11.6) provide a short overview and references; see also Zucker (2005) and for additive models, Song and Huang (2006). See Andersen and Liestol (2003) and Song and Huang (2006) for recent work on time-varying covariates.

In many settings there is no validation set; that is, there are no individuals for whom both x_2 and z are available. This is the case, for example, where the only available measurements are of x_1 and z. When there are repeat measurements of z for some set of individuals, progress is possible under certain assumptions about measurement errors. The most common model is where, given x_1, Z takes the form $Z = x_1 + \epsilon$, with the measurement error ϵ having a distribution with mean 0 and variance σ^2. In this case, σ can be estimated from repeated $z-$measurements. If no such repeated measurements are available, then we must instead proceed by assuming a value for σ.

One set of methods that has proven useful for survival models is called corrected score methods (e.g. Nakamura, 1992). Recent applications to the Cox survival model are discussed by Hu and Lin (2002), Song and Huang (2005), and Yi and Lawless (2007). This approach can also be rather easily applied to Poisson (Andersen–Gill) regression models or rate function models, discussed in Chapter 3, and to other multiplicative models. This has not been discussed in the literature, so we outline how this can be done. For simplicity we consider x_{2i} to be a scalar, but extension to the multidimensional case is straightforward.

Suppose that the process for individual i has intensity function

$$\lambda_i(t) = \lambda_0(t|H_i(t)) \exp(x'_{1i}\beta_1 + x_{2i}\beta_2)\,, \tag{8.59}$$

where $\lambda_0(t|H_i(t))$ may depend on previous event history but not on covariates. Suppose that x_{2i} is not observed, but rather

$$Z_i = x_{2i} + \epsilon_i\,, \tag{8.60}$$

where ϵ_i has mean 0, variance σ^2, and a moment generating function $m(t) = E(\exp(\epsilon_i t))$. Consider the log-likelihood contribution, based on (2.55), that is obtained from observation of a single individual. If (8.59) could be used, that is, if x_{2i} as well as x_{1i} were observed, we have

$$\ell_i(\theta) = \sum_{j=1}^{n_i} \log \lambda_i(t_{ij}|H_i(t_{ij})) - \int_0^\infty Y_i(u)\lambda_i(u)du\,, \tag{8.61}$$

where θ includes β_1, β_2, and parameters specifying $\lambda_0(t|H(t))$. Now suppose x_{2i} is not available, but consider the "corrected log-likelihood",

$$\begin{aligned}\ell_i^*(\theta) = &\sum_{j=1}^{n_i}\{\log \lambda_0(t_{ij}|H_i(t_{ij})) + x'_{1i}\beta_1 + z_i\beta_2\} \\ &-m(\beta_2)^{-1}\exp(x'_{1i}\beta_1 + z_i\beta_2)\int_0^\infty Y_i(u)\lambda_0(u|H_i(u))du\,.\end{aligned} \tag{8.62}$$

Under the measurement error model (8.60), we find that

$$E_{z_i}(\ell_i^*(\theta)|x_i, N_i^{(\infty)}) = \ell_i(\theta)\,.$$

As a result, estimating functions based on (8.62) are unbiased with respect to $N_i^{(\infty)}$ and Z_i, and can be used for estimation of θ. A common approach is to assume the errors in (8.60) are $N(0, \sigma^2)$, so that $m(t) = \exp(\sigma^2 t^2/2)$ is used in (8.62).

Yi and Lawless (2007) develop this approach for the Cox survival model, using a weakly parametric specification for the baseline hazard function $\lambda_0(t)$. Sandwich-type robust variance estimates for $\widehat{\theta}$ are readily obtained from estimating function theory (Appendix A.2). In (8.60) the parameter σ is assumed known, but the methods can be extended to deal with the estimation of σ through an additional estimating function.

Two other points should be noted. The first is that in many instances where we know that a covariate is measured with some degree of error but there are no observations on the true covariate values, we may choose simply to use only models that include the observed covariates (e.g. x_{1i}, z_i above). The second point is that we can investigate by simulation, and sometimes mathematically, the effects of assuming that $x_{2i} = z_i$, in analyses with a model for a response Y_i, given x_{1i} and x_{2i}. A well-known result is that under (8.60) and a linear model relating a response variable Y_i and covariates x_i, the least squares estimate of β_2 is biased towards zero when measurement error in x_{2i} is ignored (e.g. Carroll et al., 2006, Ch. 2). This type of attenuation also holds for many of the models in this book.

8.8 Bayesian Methods

There is a rather large literature on Bayesian methods for survival data; see Kalbfleisch and Prentice (2002, Ch. 11) and Ibrahim et al. (2001) for detailed discussion and references. Some of this methodology can be applied to recurrent events analysis, but there has been rather little direct consideration of this area, with some notable exceptions. It is not possible to do justice here to Bayesian methods and the many computational advances of recent years. We simply mention some specific examples of Bayesian methodology for event history data, from which many further references can be obtained.

There is a fairly large literature on Bayesian analysis of parametrically specified Poisson processes; Singpurwalla and Wilson (1999) provide references. Clayton (1994) considers general event history analysis and Hjort (1990b) and Sinha (1993) consider Bayesian estimation for nonparametric and semiparametric event history models. Sinha and Maiti (2004) address interval count data along with dependent terminating events. More complex settings involving time-varying covariates and marked processes are considered by Andreev and Arjas (1998), Arjas and Andreev (2000), Eerola et al. (2003), and Ishwaran and James (2004).

8.9 Bibliographic Notes

Marked point processes arise in many fields, including insurance (Grandell, 1997, Ch. 9) where claims have an associated size. Cox and Isham (1980, Ch. 5) and Snyder and Miller (1991, Ch. 4) discuss marked point process models and Andersen et al. (1993, Section 2.4) consider likelihood construction.

Methods for the analysis of cumulative cost processes have been developed in various areas. In actuarial science and insurance these are typically claims processes and there has been considerable emphasis on compound Poisson models (e.g. Grandell, 1997, Ch. 9). Nelson (2003) provides examples in reliability, concerning the costs of equipment repairs. In cancer treatment, various authors consider cumulative quality of life processes. Glasziou et al. (1990), Pepe et al. (1991), and Cook et al. (2003) consider methods based on multistate models as in Section 8.2.1. Gelber et al. (1995) and Zhao and Tsiatis (1997, 1999) deal with estimation of lifetime "costs" via inverse probability weighting. Lin et al. (1997) discuss similar methodology for the analysis of cumulative cost data in health economics, and refinements which lead to more efficient estimation are given in Bang and Tsiatis (2000). Strawderman (2000) highlights the connections between approaches for estimating lifetime cost distributions and investigates relative efficiency.

Prediction methodology is widely discussed in the statistical literature. Lawless and Fredette (2005) give general references on prediction intervals and predictive distributions. Prediction for recurrent events is considered in specific fields such as reliability (e.g. Meeker and Escobar, 1998), warranty claims (e.g. Fredette and Lawless, 2007), and software debugging (e.g. Singpurwalla and Wilson, 1999).

Cook (1995) discusses the design of clinical trials and derives formulae for the required duration of accrual and followup to meet power objectives under Poisson and mixed Poisson models. Bernardo and Harrington (2001) consider designs based on partial likelihood inference under multiplicative intensity models and Hughes (1997) considers design based on a marginal approach. McMahon et al. (1994) discuss selection criteria involving baseline counts, motivated by the ACIP trial (ACIP, 1992). Cook and Wei (2002) discuss the impact of such criteria on the bias and efficiency of estimators. Cook and Wei (2003) consider sample size requirements and show that trials may be more efficient when designed with selection criteria. Robust semiparametric methods based on (8.34) are developed in Cook et al. (2005) for settings with variable followup or multiple treatment periods (e.g. crossover trials). Cook and Lawless (1996) and Jiang (1999) provide the basis for Section 8.4.4. Cook et al. (2007) consider the design of noninferiority trials based on recurrent event responses.

Clustered data for recurrent events is considered by Cai and Schaubel (2005) and Schaubel and Cai (2005a, b), who give methods described in Section 8.6. Schaubel (2005) notes that the robust variance estimator (8.51) can be poor when the number of clusters is small and proposes alternatives.

8.10 Problems and Supplements

8.1. Consider a marked point process in which $\{N(t), 0 \le t\}$ is a renewal process with gap times W_k which are gamma distributed with density $f(w;\gamma) = w^{\gamma_1 - 1}\exp(-w/\gamma_2)/(\Gamma(\gamma_1)\gamma_2^{\gamma_1})$. Derive an expression for $E\{C(t)\}$ if the marks are independently exponentially distributed with mean α^{-1}.

[Section 8.1]

8.2. Consider a marked process where the event process for an individual is Poisson with rate function $u_i\rho(t)$ conditional on an unobserved random variable u_i. The marks C_{ij} for events at times t_{ij} are independent and identically distributed with moment generating function $M_C(s)$. Using the results of Problem 2.5, obtain an expression for the moment generating function for $C_i(t)$ given by (8.1), and evaluate it when u_i has a gamma distribution with density (2.28).

[Section 8.1]

8.3. Consider a compound time homogeneous Poisson process with independent discrete marks with $\Pr(C_j = k) = \pi_k$, $k = 1, \ldots, K$, with $\sum_{k=1}^{K}\pi_k = 1$. Let $\{N(t), 0 \le t\}$ denote the Poisson process with rate ρ, and let $N_k(t) = \sum_{j=1}^{N(t)} I(C_j = k)$ record the number of events over $[0, t]$ where $C_j = k$. Prove that $\{N_k(t), 0 \le t\}$ and $\{N_\ell(t), 0 \le t\}$ are independent Poisson processes with rate functions $\rho\pi_k$ and $\rho\pi_\ell$, for $k \ne \ell$.

[Section 8.1]

8.4. Compare the following modeling strategies for a Poisson process with discrete marks C_j which may be time-dependent:

(i) Events occur according to a Poisson process with rate function $\rho(t)$, and $\Pr(C_j = a_k | j\text{th event at time t}) = \pi_k(t)$, $k = 1, \ldots, K$, where $a_1, \ldots, a_K$ are the possible values of C_j.
(ii) Events with marks a_k occur according to a Poisson process with rate function $\rho_k(t)$, the processes for $k = 1, \ldots, K$ being independent.

Consider the estimation of the distribution of costs (8.1) in each case.

[Section 8.1]

8.5. Consider a marked point process where events occur according to a Poisson process with intensity function $\rho(t)$, $t > 0$, and marks are independent and identically distributed with probability density function $g(c)$, $c > 0$. For any set A in $(0, \infty) \times (0, \infty)$, consider the random variable $N(A) =$ the number of events for which the time t_j and mark c_j satisfy $(t_j, c_j) \in A$.

a. Prove that $N(A)$ has a Poisson distribution with mean

$$\mu(A) = \int_A \lambda(t)g(c)dtdc\,.$$

b. If the marks C have a continuous distribution, thus show that points (t_j, c_j) occur over the region $(0,\infty)\times(0,\infty)$ according to a spatial Poisson process with rate function $\lambda(t)g(c)$.

[Section 8.1; Guttorp, 1995, Section 5.6]

8.6. Lifetime costs with a terminating event.
Consider the case of cumulative costs over the period $[0, T_i]$ for an individual, where T_i denotes the random time at which the process generating the costs terminates. Using the approaches for dealing with event-dependent termination in Section 6.6, devise ways to estimate the distribution of total lifetime cost $C_i = C_i(T_i)$ in the cases where (i) costs are incurred by the occurrence of point events, and (ii) costs accrue continuously at fixed rates according to which state an individual occupies. Assume that data are available on individuals, some of whom may not have been observed to their termination time, but instead only to a censoring time τ_i. Show also that C_i is not independent of the "censoring value" $C_i^* = C(\tau_i)$ in general, so that a naive treatment of data on censored values c_i^*, which assumes independent censoring, is invalid.

[Sections 6.2, 8.2]

8.7. IPCW Estimation for lifetime cost distributions.
Let C_i denote lifetime cost over $[0, T_i]$ as in Problem 8.6, and suppose that τ_i represents a random censoring time for the ith event process, such that τ_i is independent of the event and cost process, and in particular, of (T_i, C_i). Let $G_i(t) = \Pr(\tau_i \geq t)$ denote the survivor function of τ_i.

a. Consider the following estimator based on independent individuals $i = 1, \ldots, m$,

$$\widehat{\Pr}(C_i \geq c) = \frac{1}{m}\sum_{i=1}^{m} \frac{\Delta_i I(C_i \geq c)}{\widehat{G}_i(T_i)}\,,$$

where $\Delta_i = I(T_i \leq \tau_i)$ and $\widehat{G}_i(t)$ is a consistent estimator of $G_i(t)$. Argue that this provides a consistent estimator of $\Pr(C_i \geq c)$.

b. Suppose we wish to estimate $\mu_c = E(C_i)$. Compare an estimator based on (a) with the estimator

$$\widehat{\mu}_c = \int_0^\infty \widehat{S}(t)d\widehat{\mu}_c^*(t)\,,$$

where $S(t) = \Pr(T_i \geq t)$ is estimated by the Kaplan–Meier estimate $\widehat{S}(t)$ based on the data $(\min(T_i, \tau_i),\ \delta_i = I(T_i \leq \tau_i))$, $i = 1, \ldots, m$ and where $d\mu_c^*(t) = E\{dC_i(t)|T_i \geq t\}$ is estimated by

$$d\widehat{\mu}_c^*(t) = \frac{\sum_{i=1}^{m} I\left(T_i \geq t, \tau_i \geq t\right) dC_i(t)}{\sum_{i=1}^{m} I\left(T_i \geq t, \tau_i \geq t\right)} .$$

[Section 8.2; Strawderman, 2000]

8.8. Derive (8.22) by first showing that $\widehat{\lambda} Y$ has an F distribution with $(2, 2n)$ degrees of freedom (e.g. Lawless 2003a, Section 4.6), and then obtaining the distribution of $U = 1 - \exp(-\widehat{\lambda} Y)$ from this.

[Section 8.3]

8.9. Consider a setting where events for individual i occur according to a homogeneous Poisson process with rate $u_i \rho$, given an unobservable random effect u_i. Assume that across individuals $i = 1, \ldots, m$ the u_i are independent gamma random variables with mean 1 and variance ϕ, and density function (2.28). Suppose that for individual i, n_{i1} events have been observed over an interval of length t_{i1}, and that it is desired to predict the number N_{i2} of events over a future interval of length t_{i2}.

a. Show that

$$\Pr(N_{i2} = n_{i2} | N_{i\cdot} = n_{i\cdot}) = \binom{n_{i1} + n_{i2}}{n_{i2}} \left(\frac{t_{i2}}{t_{i1} + t_{i2}}\right)^{n_{i2}} \left(\frac{t_{i1}}{t_{i1} + t_{i2}}\right)^{n_{i1}}$$

where $N_{i\cdot} = N_{i1} + N_{i2}$, and use this to obtain prediction intervals for N_{i2}.

b. Consider, alternatively, the distribution of N_{i2} given N_{i1}, which involves the parameters ρ and ϕ; they can be estimated from the data n_{i1}, $i = 1, \ldots, m$. Show how to base prediction intervals on this.

c. What are the pros and cons of the methods in (a) and (b) ?

[Section 8.3; Faulkenberry, 1973; Vit, 1973]

8.10. Suppose that $\{N_i(t), 0 \leq t\}$ is generated by a mixed time homogeneous Poisson process with rate $u_i \exp(\beta_0 + \beta_1 x_i)$ where u_i is gamma distributed with mean 1 and variance ϕ, and x_i is an observable covariate with $x_i = 1$ with probability 0.5 and $x_i = 0$ with probability 0.5. Let $\exp(\beta_0) = 1.0$, $\exp(\beta_1) = 0.80$, and $\phi = 0.50$. Suppose all subjects are expected to be observed over $[0, 1]$.

a. Derive the number of subjects required to ensure 80% power to reject $H_0 : \beta_1 = 0$ under a two-sided test with size 0.05 at $\beta_1 = 0.80$.

b. Simulate data under the design in (a) to validate the sample size you obtained.

[Section 8.4]

8.11. Consider the setting in Problem 8.10, but suppose a baseline count $N_i(-1,0) = n_{i0}$ is observed over $(-1,0)$ for subject i, which is Poisson distributed with conditional mean $u_i \exp(\beta_0)$. What would be the required sample size to ensure 80% power to reject $H_0 : \beta_1 = 0$ under a two-sided test with size 0.05 at $\beta_1 = 0.80$, if followup were scheduled over $[0,1]$ and if analyses were to be based on the joint or conditional models of Section 8.4?

[Section 8.4]

8.12. Repeat 8.10 and 8.11 where $\exp(\beta_0) = 0.25$. Is the same design as efficient (i.e. associated with a smaller sample size) in this scenario?

[Section 8.4]

8.13. Suppose that given a random effect u with mean 1 and variance ϕ, $\{N(t), 0 \leq t\}$ is a time homogeneous Poisson process with rate $u\rho$.

a. If subjects are to be observed over the interval $[0,\tau]$, derive the relative efficiency of estimators for ρ based on an analysis of the time to the first event and based on the analysis of all events over this interval, if $\phi = 0$.
b. If subjects are to be randomized to an experimental or control treatment with probability 0.5, where $x_i = 1$ or $x_i = 0$ if individual i is assigned to the experimental or control group, respectively, consider the model $\lambda_i(t|H_i(t), u_i) = u_i \rho \exp(\beta x_i)$. Suppose the aim is to achieve 90% power to reject $H_0 : \beta = 0$ when $\beta_A = \log(0.60)$ is the true value, based on a two-sided test of size 0.05. Plot the required sample sizes for analyses based on the time to the first event and based on all events as a function of ρ when $\phi = 0$ and $\tau = 1$.
c. Show that the proportional hazards assumption for the time to the first event does not hold in general when $\phi > 0$. What implications does this have for analysis strategies based on the time to the first event?

[Section 8.4]

8.14. Consider a randomized clinical trial with $x_i = 1$ if individual i is randomized to an experimental treatment, occurring with probability 0.5, and $x_i = 0$ otherwise. Suppose a treatment delays the occurrence of the first event but once this event occurs, subsequent events arise according to the same time homogeneous Poisson process in both groups. Therefore given x_i, W_{i1} is exponential with hazard function $\lambda \exp(\gamma x_i)$ and for $k \geq 2$, W_{ik} is exponential with rate λ independent of x_i.

a. Derive an expression for the mean function in the treatment group and confirm your calculations through a simulation study.
b. Derive the probability limit of $\widehat{\beta}$ obtained under a multiplicative Poisson process model with rate function $\rho \exp(\beta x_i)$. Assume there are common followup times $\tau_i = \tau$.

[Section 8.4; Boher and Cook, 2006]

A
Estimation and Statistical Inference

A.1 Maximum Likelihood

A.1.1 Introduction

Consider a parametric model indexed by an $r \times 1$ parameter $\theta = (\theta_1, \ldots, \theta_r)'$ where $\theta \in \Theta$. Suppose D denotes the data available for estimation and model checking, and $\Pr(D;\theta)$ is the probability density or mass function, which is assumed to be known up to the parameter θ. In the case of no covariates, D could, for example, represent the number and timing of events in m recurrent event processes. With fixed (time-independent) covariates, the probability model is taken to be implicitly conditional on them. The *likelihood function* $L(\theta)$ based on D is defined as any function

$$L(\theta) \;\propto\; \Pr(D;\theta) \tag{A.1}$$

or equivalently, $L(\theta) = c\Pr(D;\theta)$ where the proportionality constant c is any positive term not dependent on θ. Many of the models for recurrent events are expressed in terms of intensity functions, and one can obtain expressions for (A.1) as discussed in Chapter 2; for example, see (2.7) in Theorem 2.1.

Let D_i denote the data for process i, so $D = (D_1, \ldots, D_m)'$, and let $x_1, \ldots, x_m$ denote fixed covariate vectors. Assuming the data from the different processes are independent given the x_i, the likelihood function is

$$L(\theta) = c\,\Pr(D|x;\theta) = c\prod_{i=1}^{m}\Pr(D_i|x_i;\theta)\,, \tag{A.2}$$

where x consists of $x_1, \ldots, x_m$.

More generally, a likelihood function may be based on the conditional probability of observed data D, given the value of observed data D', so that $L(\theta) \propto Pr(D|D';\theta)$. This is sometimes used to remove nuisance parameters or to specify a more relevant reference population for inferences (e.g. Cox and

Hinkley, 1974; Barndorff-Neilsen and Cox, 1994; Pawitan, 2001). It should also be noted that D in (A.1) may be only a portion of the available data, in which case $L(\theta)$ is sometimes called a marginal likelihood. Finally, so-called partial likelihoods (Cox, 1975) are often used for inference. Strictly speaking, they are not likelihoods, but are functions $L(\theta)$ that give estimating equations $\partial \log L(\theta)/\partial\theta' = 0$ which yield consistent estimators $\widehat{\theta}$. For simplicity, we present results below for (A.1), but they also apply to conditional or partial likelihoods. See Lawless (2003a, p. 552) for a concise description of partial likelihood.

The maximum likelihood estimate (M.L.E.) $\widehat{\theta}$ is the value of θ that maximizes $L(\theta)$. The M.L.E. therefore has the appealing property of giving the highest probability of realizing the data observed within the class of models considered. The *log-likelihood function* is

$$\ell(\theta) = \log L(\theta)\,, \tag{A.3}$$

and because the log function is monotonic, maximizing $\ell(\theta)$ is equivalent to maximizing $L(\theta)$. There are also computational advantages to maximizing $\ell(\theta)$ because the product in (A.2) becomes a sum.

The *score vector* $U(\theta) = (U_1(\theta), \ldots, U_r(\theta))'$ is the gradient of the log-likelihood so that

$$U_j(\theta) = \partial\ell(\theta)/\partial\theta_j\,, \tag{A.4}$$

$j = 1, 2, \ldots, r$. In the following, $\Pr(D;\theta)$ is assumed to be the true distribution of D, and expectations are with respect to this distribution. Mild regularity conditions on the model are also assumed. Standard results from maximum likelihood theory (e.g. Cox and Hinkley, 1974; van der Vaart, 1998; Severini, 2000) then give $E\{U(\theta)\} = 0$. The *observed information matrix* $I(\theta)$ is the negative of the $r \times r$ Hessian matrix of the log-likelihood so that

$$I_{jk}(\theta) = -\partial^2\ell(\theta)/\partial\theta_j\partial\theta_k = -\partial U_j(\theta)/\partial\theta_k\,, \tag{A.5}$$

$j, k = 1, 2, \ldots, r$. The *expected information matrix* $\mathcal{I}(\theta)$ is both the expectation of the observed information matrix, and the covariance matrix of the score vector, so

$$\mathcal{I}(\theta) = E\left\{-\partial U(\theta)/\partial\theta'\right\} = E\left\{U(\theta)U'(\theta)\right\}\,. \tag{A.6}$$

A.1.2 Asymptotic Pivotals

Pivotal quantities (or *pivotals*, for simplicty) are functions of random variables and parameters with known distributions that do not depend on θ, and provide a very useful basis for inference. They in fact exist only in certain special settings, but it is possible to specify functions that are asymptotically pivotal. Here and elsewhere in the book, "asymptotic" means that the

information about θ is becoming arbitrarily large in some way; this usually occurs through the number of observed processes (m) approaching infinity. The relative likelihood $R(\theta) = L(\theta)/L(\widehat{\theta})$, which gives the relative plausibility of a particular value of θ compared to the M.L.E., is one asymptotic pivotal. Because $0 \leq R(\theta) \leq 1$, this function is often helpful to examine, however, the log-relative likelihood $r(\theta) = \log R(\theta)$ is more convenient to work with and related to this we have the *likelihood ratio pivotal*

$$W(\theta) = -2r(\theta) = 2\ell(\widehat{\theta}) - 2\ell(\theta) \sim \chi^2_r \tag{A.7}$$

asymptotically. A $100p\%$ relative likelihood region, defined as $\{\theta : R(\theta) \geq p\}$, is comprised of all values of θ which are at least $100p\%$ as likely as the M.L.E. They are also asymptotic confidence regions for θ: using (A.7) and letting $\chi^2_{r,\alpha}$ be the upper α-quantile for χ^2_r, for $0 < \alpha < 1$ we have $\{\theta : W(\theta) \leq \chi^2_{r,\alpha}\} = \{\theta : R(\theta) \geq \exp(-\chi^2_{r,\alpha}/2)\}$ as an approximate α confidence region.

Let $\theta = (\theta_1', \theta_2')'$ where $\theta_1 = (\theta_{11}, \ldots, \theta_{1p})'$ and $\theta_2 = (\theta_{21}, \ldots, \theta_{2q})'$ and $p + q = r$. Now suppose interest lies in θ_1 and that θ_2 is viewed as a nuisance parameter. Let $\tilde{\theta}_2(\theta_1)$ denote the value of θ_2 that maximizes $L(\theta)$ when θ_1 is fixed. The function $L(\theta_1, \tilde{\theta}_2(\theta_1))$ is called a *profile likelihood function* for θ_1; θ_2 has been "profiled" out by replacing it with its maximum value (at each θ_1). The *profile relative likelihood function* is

$$R_1(\theta_1) = \frac{L(\theta_1, \tilde{\theta}_2(\theta_1))}{L(\widehat{\theta}_1, \widehat{\theta}_2)}$$

which has properties like the ordinary relative likelihood function. In particular $r_1(\theta_1) = \log R_1(\theta_1)$ is the profile relative log-likelihood function and

$$W_1(\theta_1) = -2r_1(\theta_1) \sim \chi^2_p \tag{A.8}$$

asymptotically. We call (A.8) a *profile likelihood ratio pivotal* which becomes a likelihood ratio statistic when we insert a particular value for θ_1 for testing hypotheses.

Standard asymptotic likelihood arguments show $U(\theta) \sim MVN(0, \mathcal{I}(\theta))$ asymptotically. This gives the asymptotic pivotal

$$U'(\theta)\mathcal{I}^{-1}(\theta)U(\theta) \sim \chi^2_r\,. \tag{A.9}$$

An asymptotic pivotal is also obtained here and in other pivotals below if $\mathcal{I}(\theta)$ is replaced by $I(\theta)$, $\mathcal{I}(\widehat{\theta})$, or $I(\widehat{\theta})$. In most event history settings $\mathcal{I}(\theta)$ is hard to obtain, so this is of major practical importance. We call (A.9) a *score pivotal*, which can in principle be used for the construction of confidence regions for θ; it reduces to a *score statistic* upon inserting a particular value for θ into (A.9).

Suppose also that $U(\theta) = (U_1'(\theta), U_2'(\theta))'$, where $U_k(\theta) = \partial\ell(\theta)/\partial\theta_k$, $k = 1, 2$ for parameter subvectors θ_1 and θ_2 as above. Then the expected information matrix and its inverse can be analogously partitioned as

$$\mathcal{I}(\theta) = \begin{bmatrix} \mathcal{I}_{11}(\theta)\, \mathcal{I}_{12}(\theta) \\ \mathcal{I}_{21}(\theta)\, \mathcal{I}_{22}(\theta) \end{bmatrix} \quad \text{and} \quad \mathcal{I}^{-1}(\theta) = \begin{bmatrix} \mathcal{I}^{11}(\theta)\, \mathcal{I}^{12}(\theta) \\ \mathcal{I}^{21}(\theta)\, \mathcal{I}^{22}(\theta) \end{bmatrix}.$$

It can be shown that

$$U_1'(\theta_1, \tilde{\theta}_2(\theta_1))\mathcal{I}^{11}(\theta_1, \tilde{\theta}_2(\theta_1))U_1(\theta, \tilde{\theta}_2(\theta_1)) \sim \chi_p^2 \tag{A.10}$$

asymptotically. The same result holds if $\mathcal{I}^{11}(\theta)$ is replaced by a consistent estimator. We often call (A.10) a *partial score pivotal*.

There are also several so-called Wald-based approximate pivotals. One is of the form

$$(\widehat{\theta} - \theta)'\mathcal{I}(\widehat{\theta})(\widehat{\theta} - \theta) \sim \chi_r^2 \tag{A.11}$$

but we more often use $I(\theta)$ instead of $\mathcal{I}(\theta)$, giving the *Wald pivotal*

$$(\widehat{\theta} - \theta)'I(\widehat{\theta})(\widehat{\theta} - \theta) \sim \chi_r^2\,. \tag{A.12}$$

The analogous results for the nuisance parameter case are

$$(\widehat{\theta}_1 - \theta_1)'[\mathcal{I}^{11}(\widehat{\theta})]^{-1}(\widehat{\theta}_1 - \theta_1) \sim \chi_p^2 \tag{A.13}$$

and

$$(\widehat{\theta}_1 - \theta_1)'[I^{11}(\widehat{\theta})]^{-1}(\widehat{\theta}_1 - \theta_1) \sim \chi_p^2\,, \tag{A.14}$$

respectively. In the common situation where $p = 1$, (A.14) equals Z^2, where

$$Z = \frac{\widehat{\theta}_1 - \theta_1}{\sqrt{I^{11}(\widehat{\theta})}} \sim N(0, 1) \tag{A.15}$$

asymptotically. We often call $\sqrt{I^{11}(\widehat{\theta})}$ the standard error of $\widehat{\theta}_1$ and write it as s.e.$(\widehat{\theta})$. The results (A.11) to (A.15) are a consequence of the fact that $\widehat{\theta}$ is asymptotically normal; informally, we treat $\widehat{\theta}$ as approximately normal with mean θ and covariance matrix $I^{-1}(\widehat{\theta})$.

A.1.3 Confidence Regions or Intervals

The likelihood ratio, score, and Wald pivotals all provide a basis for constructing confidence regions. As described above, the likelihood ratio pivotal (A.7) gives an approximate α confidence region for θ as

$$\{\theta : -2r(\theta) \leq \chi_{r,\alpha}^2\}$$

which has the added appeal of being a relative likelihood region. If interest lies in θ_1 only, then one obtains

$$\left\{\theta_1 : -2r_1(\theta_1) \leq \chi^2_{p,\alpha}\right\}$$

based on (A.8).

The score pivotal (A.9) gives $\left\{\theta : U'(\theta)\mathcal{I}^{-1}(\theta)U(\theta) \leq \chi^2_{r,\alpha}\right\}$ or

$$\{\theta_1 : U_1'(\theta_1, \tilde{\theta}_2(\theta_1))\mathcal{I}^{11}(\theta_1, \tilde{\theta}_2(\theta_1))U_1(\theta_1, \tilde{\theta}_2(\theta_1)) \leq \chi^2_{p,\alpha}\}$$

which can be computationally less appealing than those based on the likelihood ratio pivotals. Most often, approximate α confidence regions are based on Wald-based approximate pivotals (A.14) or (A.15) to give

$$\{\theta : (\widehat{\theta} - \theta)' I(\widehat{\theta})(\widehat{\theta} - \theta) \leq \chi^2_{r,\alpha}\}$$

or

$$\{\theta_1 : (\widehat{\theta}_1 - \theta_1)' I^{11}(\widehat{\theta})]^{-1}(\widehat{\theta}_1 - \theta_1) \leq \chi^2_{p,\alpha}\}\,.$$

When $p = 1$, (A.15) gives the two-sided $1 - \alpha$ confidence interval $\widehat{\theta} \pm z_{\alpha/2}\,\text{s.e.}(\widehat{\theta})$, where z_α is the standard normal upper α-quantile.

Confidence regions based on the likelihood ratio and some score-based pivotals are invariant to the parameterization of the model, but Wald-based methods generally are not. Furthermore, the closeness of the actual distribution of the approximate Wald pivotals to their limiting normal or χ^2 forms can vary substantially, depending on the parameterization and the amount of information about θ. A good strategy for analyses based on Wald pivotals is to obtain the information matrix using a parameterization in which no component parameters have range restrictions, and for which likelihood ratio statistics are approximately quadratic in θ. This typically improves the accuracy of the Wald-based confidence intervals or regions. The resulting confidence intervals may then be transformed to the scale of interest.

A.1.4 Tests of Hypotheses

The approximate pivotal results above hold when θ is the "true" parameter value and they may therefore be used to test hypotheses regarding particular values. Tests of $H_0 : \theta = \theta_0$, for example, can be based on a likelihood ratio statistic (A.7). Large values of $W(\theta_0)$ indicate evidence against H_0, and the $p-$value is $\Pr(\chi^2_r \geq -2r(\theta_0))$. A score statistic (A.9) with $\theta = \theta_0$ can similarly be used, and gives a $p-$value

$$\Pr(\chi^2_r \geq U'(\theta_0)\mathcal{I}^{-1}(\theta_0)U(\theta_0))\,.$$

Finally, a Wald statistic (A.12) with $\theta = \theta_0$ gives a $p-$value

$$\Pr(\chi^2_r \geq (\widehat{\theta} - \theta_0)' I(\widehat{\theta})(\widehat{\theta} - \theta_0))\,.$$

An alternative is to replace $\widehat{\theta}$ in $I(\widehat{\theta})$ with θ_0; which may be preferable in terms of power depends on the setting. Analogous tests for the nuisance parameter case can be based on (A.8), (A.10), (A.13), and (A.14).

A.2 Estimating Functions

In many contexts, interest lies in the relation between a vector of explanatory variables and an outcome variable, but there is a desire to base inferences on as few parametric assumptions as possible.

Consider a sample of m independent responses $y_1, \ldots, y_m$, with associated covariate vectors $x_1, \ldots, x_m$. Suppose θ is a $p \times 1$ parameter vector and that $U_i(y_i, x_i; \theta)$ is a $p \times 1$ vector $U_i(y_i, x_i; \theta) = (U_{i1}(y_i, x_i; \theta), \ldots, U_{ip}(y_i, x_i; \theta))'$. A set of *estimating equations* for θ is given by

$$U(\theta) = \sum_{i=1}^{m} U_i\left(y_i, x_i; \theta\right) = 0\,.$$

Suitably defined, such estimating equations yield consistent, asymptotically normal estimators $\widehat{\theta}$. We assume here that solving $U(\theta) = 0$ yields a unique estimate $\widehat{\theta}$. The vector $U(\theta)$ is called an *estimating function* and estimating functions which satisfy $E\{U(\theta)\} = 0$ are called *unbiased estimating functions*, where the expectation is taken with respect to the Y_i given the x_i. Unbiased estimating functions yield consistent estimators $\widehat{\theta}$ under appropriate conditions and have considerable appeal. Estimating functions which are only asymptotically unbiased are also useful. Estimating functions of many types are widely used, and Crowder (1987), Heyde (1997), White (1982), and others provide general theory. We quote here the key results needed for the construction of tests and interval estimates.

Let

$$A_m(\theta) = -\frac{1}{m}\partial U(\theta)/\partial\theta', \qquad B_m(\theta) = \frac{1}{m}\sum_{i=1}^{m} U_i(\theta)U_i'(\theta)$$

$$A(\theta) = \lim_{m\to\infty} E\{A_m(\theta)\}\,, \qquad B(\theta) = \lim_{m\to\infty} E\{B_m(\theta)\}\,,$$

where the stated expectations are assumed to exist. Note that $B(\theta) = \lim_{m\to\infty} m^{-1}\mathrm{var}\{U(\theta)\}$ when $U(\theta)$ is unbiased.

Under suitable regularity conditions $m^{-1/2}U(\theta)$ is asymptotically normal with covariance matrix $B(\theta)$ and $\sqrt{m}(\widehat{\theta} - \theta)$ is asymptotically normal with mean zero and covariance matrix $C(\theta) = A^{-1}(\theta)B(\theta)[A^{-1}(\theta)^-]'$ as $m \to \infty$. A consistent estimate of $C(\theta)$ is

$$C_m(\widehat{\theta}) = A_m^{-1}(\widehat{\theta})B_m(\widehat{\theta})\left[A_m^{-1}(\widehat{\theta})\right]'\,.$$

Under maximum likelihood estimation, $U(\theta) = \partial\ell(\theta)/\partial\theta$ and $A(\theta) = B(\theta) = m^{-1}\mathcal{I}(\theta)$, the scaled Fisher information matrix. This leads to $C(\theta) = m\mathcal{I}^{-1}(\theta)$ as given in Section A.1.2.

White (1982) considers estimators $\widehat{\theta}$ obtained from estimating equations in the case where the model on which they are based may not be correct. His

results are very useful in considering the effects of model misspecification on estimators. Briefly, if (Y_i, X_i) are i.i.d. with true distribution G and if there is a unique vector θ^* such that $E_G\{U_i(Y_i, X_i; \theta^*)\} = 0$, then under suitable regularity conditions the solution $\widehat{\theta}$ to $U(\theta) = 0$ is asymptotically normal. In particular $\sqrt{m}(\widehat{\theta} - \theta^*)$ is asymptotically normal with mean vector 0 and covariance matrix $C(\theta^*) = A^{-1}(\theta^*)B(\theta^*)[A^{-1}(\theta^*)]'$, where $A(\theta)$ and $B(\theta)$ are defined as above, but with expectations taken with respect to G. Moreover, $C_m(\widehat{\theta})$ defined above is a consistent estimator of $C(\theta^*)$. It should be noted that in this framework, the interpretation of θ^* under G may not be clear, but we can think of θ^* as the θ-vector that provides the closest approximation to G in the class of models with which $U(\theta)$ is compatible (White, 1982). Boos (1992) discusses pseudo-score tests based on $U(\theta)$, and gives a very useful summary of asymptotic results.

B
Computational Methods

B.1 Software for Recurrent Events

A number of software packages or functions for the analysis of survival data also have the capability of dealing with recurrent events, among them the functions `survreg` and `coxph` in S-PLUS (Insightful Corp.; see http://www.insightful.com) and the procedures LIFEREG and PHREG in SAS (SAS Institute, see http://www.sas.com). We have used the S-PLUS functions and the corresponding R versions (see http://www.r-project.org/˜) for many of the examples in this book; illustrations are provided in Appendix C. Therneau and Grambsch (2000) describe how both S-PLUS and SAS survival analysis procedures for the Cox model can be used more generally for event history analysis. Other major statistical packages such as Stata (Stata Corp.; see http://www.stata.com) and Limdep (Econometric Software Inc.; see http://www.limdep.com) also provide some procedures that will deal with recurrent events. The RELIABILITY procedure in SAS/QC and the SAS package JMP (see http://www.jmp.com) also carry out certain specialized tasks.

New software is constantly being written and made available. Many procedures have been implemented as S functions which are freely available through the shared resources of R (see http://www.r-project.org/ or http://lib.stat.cmu.edu/) or directly from a Web site. For example, W. Q. Meeker and L. A. Escobar have a package called SPLIDA (see http://www.public.iastate.edu/∼splida) that provides many parametric and nonparametric procedures for survival and recurrent event analysis. Software for additive semiparametric models is available from different sources; a recent package is `timereg`, written for R and available at http://staff.pubhealth.ku.dk/∼ts/timereg.html. These are given as useful examples; other resources can be found via usual routes such as papers or Web searches.

B.2 Optimization Methods

There are many problems associated with event history analysis for which specialized software is not available. However, it is usually feasible to use widely available optimization software to maximize likelihood functions, or other objective functions, in order to obtain parameter estimates and fit models. The primary need is for procedures to maximize a function $\ell(\theta)$ for θ lying in a parameter space Ω; in this book $\ell(\theta)$ is usually a log-likelihood function based on some model and observed data.

Extensive accounts of optimization methods and software are available in books such as Nocedal and Wright (1999), and the *Numerical Recipes* books (see http://www.nr.com). Chapter 16 of Venables and Ripley (2002) gives a good overview and is an excellent source of information about optimization software in R and S-PLUS. Broadly speaking, methods can be classified according to whether they use no derivatives, first derivatives only, or first and second derivatives of $\ell(\theta)$. Good optimization software includes numerical methods for closely approximating derivatives, and for most packages it is not required to provide derivatives as input. For maximum likelihood problems, we generally want the observed information matrix $I(\theta) = -\partial^2 \ell(\theta)/\partial\theta\partial\theta'$ evaluated at the maximum likelihood estimate $\widehat{\theta}$; many optimization packages will compute $I(\widehat{\theta})$ or $H(\widehat{\theta}) = -I(\widehat{\theta})$ numerically and return it as part of the output. For various examples in this book, we used either the R function `nlm` or one of the SAS NLP procedures. Other statistical packages, such as Stata, also have general maximum likelihood procedures.

Sometimes a function $\ell(\theta)$ may be maximized on or near the boundary of Ω and software that deals well with constraints on parameters is needed. Log-likelihoods or other objective functions $\ell(\theta)$ may also possess multiple stationary points in some settings, and a numerical optimizer may not necessarily obtain the global maximum. From a statistical perspective, it is important to understand the shape of $\ell(\theta)$. Features such as multiple maxima, nonconvexity, or maxima on the boundary have implications for estimation, and the log-likelihood indicates the amount of information on various parameters, or functions of them. Lawless (2003a, Appendix D1) and Venables and Ripley (2002, Ch. 16) provide a number of additional comments on obtaining parameter estimates.

B.3 Simulation and Resampling Methods

We typically rely on asymptotic approximations, described in Appendix A, to provide confidence intervals for parameters or $p-$values for tests of hypotheses. These approximations can be inaccurate in certain settings, especially when the amount of information about a parameter is small. In that case, and more generally when suitable asymptotic approximations are not available for a test statistic or random quantity used for estimation, we can utilize simulation.

Specifically, suppose that $W = g(D, \theta)$ is a function of the data D and a parameter vector θ that specifies the distribution of D. For example, W may be an approximate pivotal quantity used for estimation, such as the Wald statistic $W = (\widehat{\theta}_1 - \theta_1)/\text{s.e.}(\widehat{\theta}_1)$, or the likelihood ratio statistic $W = 2\ell(\widehat{\theta}) - 2\ell(\theta)$. Alternatively, it could be a test statistic such as (3.43) or (3.52) in Chapter 3. We often assume as well that the distribution of W is either independent of θ, or asymptotically so; this is the case, for example, when considering approximate pivotal quantities for interval estimation, or the distributions of test statistics under the null hypothesis.

If we have a probability model $\Pr(D; \theta)$ for the data, the distribution of W can be estimated by simulation as follows.

1. Generate pseudo-data D^* from the distribution $\Pr(D; \theta_0)$, where θ_0 is specified. If the distribution of W is independent of θ, we can choose any convenient θ_0, but we generally choose $\theta_0 = \widehat{\theta}(d)$, the estimate of θ obtained from the observed data $D = d$.
2. Compute $W^* = g(D^*, \theta_0)$.
3. Repeat this B times, yielding values $W_1^*, \ldots, W_B^*$. The distribution of W can now be estimated from $W_1^*, \ldots, W_B^*$. For example, the empirical distribution function based on these values provides an estimate of the distribution function of W. Ways of generating data from models used in this book have been considered in Chapters 2 and 5; see the end of this section.

In some settings we do not have a full probability model for the data. This is the case, for example, when we are employing estimating functions without a probability distribution or when the data involve censoring or other forms of incompleteness that are random, but not specified in terms of a model. In that case we may hope to use some form of *nonparametric bootstrap*. The fundamental nonparametric bootstrap when there are i.i.d. data from m individuals is to replace step 1 above with

1′. Generate pseudo-data D^* by selecting a random sample of m units, drawn with replacement from the observed data units $d_1, \ldots, d_m$ for the m individuals.

This procedure is frequently applicable; see Efron and Tibshirani (1993), Davison and Hinkley (1997), and other books on the bootstrap for guidance. However, in many event history analysis settings, it is either inapplicable, or its behaviour and validity have not been established. In the latter case, we may wish to conduct simulation studies for specific processes, in order to assess the properties of the nonparametric bootstrap. In other cases, for example, where the data consist of rather long series of events for a small number of processes (perhaps even one), quite different resampling procedures would have to be developed. Little has been done in this area. Finally we note that when interest lies in testing for differences between two groups of individuals in randomized studies, based on (3.52) for example, permutation tests may be used.

Simulation of event processes is also important in applications where we wish to predict the future course of one or more processes. Methods of simulating data from models used in this book have been discussed in Chapters 2 and 5. In particular, (2.9) and (2.46) provide a basis for generating successive events. Variations on this (e.g. See Problem 2.2; Ogata, 1981; Daley and Vere-Jones, 2003) can also be developed, and there are various special algorithms for models such as Poisson and renewal processes (e.g. Ross, 1983, Ch. 11).

C

Code and Remarks for Selected Examples

C.1 Tumorgenicity Data Analysis of Chapter 3

Here we present the data frames and S-PLUS or R code for the analysis of the tumor data of Gail et al. (1980) given in Table 1.1. The results were discussed in Section 3.8.1.

C.1.1 Poisson Analysis with Weibull Baseline Rate

The data for the first five rats in the treated group are displayed below in the data frame "rats", in the "counting process" format.

```
>  rats[1:12, ]
    id start stop status rtrunc tstatus enum trt
 1  1     0  122      1     NA       1    1   1
 2  2     0  122      0     NA       1    1   1
 3  3     0    3      1     NA       1    1   1
 4  3     3   88      1     NA       2    2   1
 5  3    88  122      0     NA       2    3   1
 6  4     0   92      1     NA       1    1   1
 7  4    92  122      0     NA       2    2   1
 8  5     0   70      1     NA       1    1   1
 9  5    70   74      1     NA       2    2   1
10  5    74   85      1     NA       2    3   1
11  5    85   92      1     NA       2    4   1
12  5    92  122      0     NA       2    5   1
```

The `id` variable indicates the rat to which the times correspond. The `start` variable contains the time $t_{i0} = 0$ and the event times (t_{ij}, $j = 1, \ldots, n_i$), and the `stop` variable contains the times of the events (t_{ij}, $j = 1, \ldots, n_i$) and the end-of-followup time (τ_i). By inspection of (3.10) one can see that the `start` and `stop` variables contain the lower and upper limits of integration, respectively, for each "at risk" interval $(t_{i,j-1}, t_{ij}]$ in the likelihood construction. The `status` variable is 1 if the time t_{ij} is an event time, and `status`$= 0$

when it is a right censoring time (i.e. the time of last contact). The variable `tstatus` indicates whether the time in the `start` variable is a left truncation time. If `start`= 0, there is no left truncation (this occurs for the first lines in the data frame for each subject) and these are designated by `tstatus`=1. For subsequent lines, `tstatus`=2 because the `start` variable contains a left truncation time. There is no right truncation, but a corresponding variable must be specified for the "truncation" option in `censorReg` and so the `rtrunc` variable is given with `NA` in each row of the data frame. The variable `enum` simply records the cumulative number of lines in the data frame for each subject and `trt` is a treatment indicator such that `trt`=1 for treated subjects and `trt`=0 for control subjects.

The code for fitting the "Weibull" model (3.56) to the control rats is given below along with part of the resulting output. Note that the "`truncation`" option enables one to indicate that the left truncation times are contained in the `start` variable. More general truncation patterns can be handled using other codes for `tstatus` but we do not need this here.

```
> wfitC <- censorReg(censor(stop, status) ~ 1,
                     data=rats, subset=(trt==0),
                     truncation=censor(start,rtrunc,tstatus),
                     distribution="weibull")

  Coefficients:
               Est. Std.Err. 95% LCL 95% UCL z-value     p-value
  (Intercept) 3.161   0.153    2.861   3.461  20.662 7.55142e-95

  Extreme value distribution: Dispersion (scale) = 0.914
  Observations: 173 Total; 22 Censored
  -2*Log-Likelihood: 1208.5
```

See Section 3.8.1 for guidance on re-expressing these estimates to construct a mean function estimate.

The regression model $\rho_i(t) = \rho_0(t)\exp(x_i\beta)$ is fit by omitting the `subset` option and specifying the `trt` covariate in the model. The resulting S-PLUS (or R) code and output are as follows. Note that because of the way the model is parameterized in `censorReg`, the parameters α_1, α_2, β in Section 3.8.1 are related to those here by $\alpha_1 = \exp(-\text{Intercept})$, $\alpha_2 = 1/\text{Dispersion}$, $\beta = -\text{trt}/\text{Dispersion}$.

```
> censorReg(censor(stop, status) ~ trt, data=rats,
            truncation=censor(start,rtrunc,tstatus),
            distribution="weibull")

  Coefficients:
                Est. Std.Err.  95% LCL 95% UCL  z-value      p-value
  (Intercept) 3.110    0.139    2.837   3.383   22.314 2.67312e-110
          trt 0.775    0.153    0.476   1.074    5.084  3.70418e-07
```

```
Extreme value distribution: Dispersion (scale) = 0.942
Observations: 254 Total; 42 Censored
-2*Log-Likelihood: 1798.05
```

C.1.2 Poisson Analysis with Piecewise-Constant Rates

In Section 3.8.1 a piecewise-constant rate function (Section 3.3) with cut-points at 30, 60, and 90 days is considered, resulting in four parameters in the rate function. The entries in the data frame "rats.pw" are given below for the first five rats in the treated group. The columns contain the variable `id` indexing rats, a variable `interval` indexing the intervals (k) across which the rate functions are constant, the variable `count` containing the counts n_{ik} for the corresponding interval, the variable `len` containing the time "at risk" in the corresponding interval for each rat (S_{ik}), and the treatment indicator `trt`.

```
rats.pw[1:20, ]
   id interval count len trt
 1  1        1     0  30   1
 2  1        2     0  30   1
 3  1        3     0  30   1
 4  1        4     1  32   1
 5  2        1     0  30   1
 6  2        2     0  30   1
 7  2        3     0  30   1
 8  2        4     0  32   1
 9  3        1     1  30   1
10  3        2     0  30   1
11  3        3     1  30   1
12  3        4     0  32   1
13  4        1     0  30   1
14  4        2     0  30   1
15  4        3     0  30   1
16  4        4     1  32   1
17  5        1     0  30   1
18  5        2     0  30   1
19  5        3     3  30   1
20  5        4     1  32   1
```

Here the `interval` variable is declared a factor variable to create the indicators $w_k(t)$, $k = 2, 3, 4$ mentioned in Section 3.3. Using standard notation for generalized linear models (McCullaugh and Nelder, 1989), we let β_0 denote the intercept and β_k denote the coefficient for $w_k(t)$, $k = 2, 3, 4$. Then $\alpha_1 = \beta_0$ and $\alpha_k = \beta_k + \beta_0$, $k = 2, 3, 4$. The log of the length of time a subject was observed to be at risk in interval k is the offset. The following is the code to fit the piecewise-constant model for control rats, and the corresponding output.

```
> pfitC <- glm(count ~ offset(log(len)) + factor(interval),
              family=poisson(link=log),
              data=rats.pw, subset=(trt==0))

     Min      1Q  Median     3Q    Max
-1.9183 -1.5748 -0.2736 0.6262 2.8959

Coefficients:
                    Value Std. Error  t value
      (Intercept) -3.0937     0.1714 -18.0503
factor(interval)2  0.1625     0.2330   0.6977
factor(interval)3  0.3023     0.2260   1.3377
factor(interval)4 -0.1569     0.2481  -0.6325

(Dispersion Parameter for Poisson family taken to be 1 )

    Null Deviance: 167.7858 on 99 degrees of freedom
Residual Deviance: 163.3078 on 96 degrees of freedom
Number of Fisher Scoring Iterations: 4
```

An estimate of the treatment effect based on a multiplicative model is obtained by introducing the `trt` covariate into the regression model.

```
> options(contrasts = c("contr.treatment", "contr.poly"))

> glm(count ~ offset(log(len)) + factor(interval) + trt,
      family=poisson(link=log), data=rats.pw)

Deviance Residuals:
     Min      1Q  Median     3Q    Max
-1.8336 -1.1994 -0.3302 0.4701 3.0551

Coefficients:
                   Value Std. Error  t value
      (Intercept) -3.0882     0.1506 -20.5007
factor(interval)2  0.1719     0.1956   0.8785
factor(interval)3  0.2063     0.1941   1.0631
factor(interval)4 -0.0645     0.2039  -0.3166
              trt -0.8230     0.1514  -5.4354

(Dispersion Parameter for Poisson family taken to be 1 )

    Null Deviance: 301.3706 on 191 degrees of freedom
Residual Deviance: 266.3228 on 187 degrees of freedom
Number of Fisher Scoring Iterations: 4
```

C.1.3 Nonparametric and Semiparametric Poisson Analysis

The nonparametric Nelson-Aalen estimate of the mean function for control individuals given in Figures 3.2 and 3.3, is obtained by fitting a null Cox model (i.e. no covariates) using the "rats" data frame. The code is as follows.

```
> NPfit <- coxph(Surv(start,stop,status) ~ 1,
                 data=rats, subset=(trt==0), method="breslow")

> KM <- survfit(NPfit, type="aalen")
> NA.MF <- data.frame(time=c(0,KM$time),
                      na=-log(c(1,KM$surv)))
```

The "`method="breslow"`" specification indicates a way of dealing with ties in the event times and is a standard approach for survival analysis (although not the default in S-PLUS); see Lawless (2003a). Specification of "`type="aalen"`" in the `survfit` function ensures that the resulting nonparametric estimate of the mean function is of the form (3.17).

The semiparametric multiplicative Poisson model with a treatment covariate can be fit with the `coxph` function by including all subjects and adding the covariate `trt`. In the data frame below, `start` and `stop` contain the left truncation time and event/censoring times, respectively, where `status` indicates which times are event times. The following code is used for fitting the semiparametric model.

```
> coxph(Surv(start,stop,status) ~ trt, data=rats, method="breslow")
n= 254
      coef exp(coef) se(coef)     z       p
trt -0.816     0.442    0.152 -5.37 7.8e-08

    exp(coef) exp(-coef) lower .95 upper .95
trt     0.442       2.26     0.328     0.596

Likelihood ratio test= 31.7  on 1 df,   p=1.81e-08
Wald test            = 28.9  on 1 df,   p=7.76e-08
Score (logrank) test = 30.5  on 1 df,   p=3.27e-08
```

C.1.4 Semiparametric Mixed Poisson Analysis

The `coxph` function in S-PLUS or R can also be used to fit semiparametric mixed Poisson models using the `frailty(id)` option. The particular forms one can use are gamma, log-normal, and t distributions, but only the gamma random effect distribution (the default) gives exact maximum likelihood estimates and we emphasis the use of this distribution here. As discussed in

Section 3.5, the gamma frailty distribution is parameterized to have mean 1 and variance ϕ.

Therneau and Grambsch (2000, Ch. 9) provide extensive discussion on the use of the frailty option in the `coxph` function. Therneau et al. (2003) show that for a fixed random effect variance parameter, the EM equations for β coincide with those arising from an analysis based on penalized likelihood.

```
>  coxph(Surv(start,stop,status) ~ trt + frailty(id),
        data=rats, method="breslow")
   n= 254
                 coef se(coef)   se2 Chisq   DF       p
          trt -0.816 0.211    0.153 15.0   1.0 0.00011
   frailty(id)                       49.5  24.3 0.00190

       exp(coef) exp(-coef) lower .95 upper .95
   trt     0.442       2.26     0.292     0.668

   Iterations: 6 outer, 18 Newton-Raphson
      Variance of random effect= 0.27   I-likelihood = -791.9
```

The standard error `se(coef)` is preferred to `se2` as it is apparently close to correct, and gives more conservative statements about covariate effects. The `I-likelihood` is the marginal log-likelihood for the model (obtained after integrating out the random effect) and should therefore be used for likelihood ratio tests.

C.1.5 Robust Semiparametric Analysis

Robust variance estimates and standard errors as given in (3.38) are obtained by using the `cluster(id)` option in `coxph`.

```
coxph(Surv(start,stop,status) ~ trt + cluster(id),
            data=rats, method="breslow")
  n= 254
           coef exp(coef) se(coef) robust se        z          p
  trt -0.815774  0.442297 0.151836   0.19809 -4.11819 3.8186e-05

      exp(coef) exp(-coef) lower .95 upper .95
  trt  0.442297    2.26092  0.299985  0.652122

  Likelihood ratio test= 31.69 on 1 df, p=1.81146e-08
  Wald test            = 16.96 on 1 df, p=3.8186e-05
  Score (logrank) test = 30.54 on 1 df, p=3.26554e-08, Robust = 11.2
p=0.000816617

(Note: the likelihood ratio and score tests assume
```

```
independence of observations within a cluster, the
Wald and robust score tests do not).
```

The note at the end of the output is a reminder that likelihood-based inferences are not robust. The Wald test and the score tests are robust however, where the latter is based on a score statistic for $\beta = 0$ from a Poisson process, but uses a robust variance estimate.

C.2 Code for rhDNase Data Analyses of Chapter 4

Following some manipulation, the data given in Section D.3 of Appendix D for the first few subjects can be reformatted and read into S-PLUS or R as a data frame in the counting process formulation as follows.

```
> rhDNase <- read.table("rhDNase.dat", header=T)
> rhDNase[1:18, c("id","trt","fev","time1","time2","status","etype",
                  "enum","enum1")]

       id trt  fev time1 time2 status etype enum enum1
1  493301   1 28.8     0   168      0     1    1     1
2  493303   1 64.0     0   169      0     1    1     1
3  493305   0 67.2     0    65      1     1    1     1
4  493305   0 67.2    65    75      1     2    2     1
5  493305   0 67.2    75   168      0     1    3     2
6  493309   1 57.6     0   168      0     1    1     1
7  493310   0 57.6     0   171      0     1    1     1
8  493311   1 25.6     0   166      0     1    1     1
9  493312   0 86.4     0   168      0     1    1     1
10 493313   0 32.0     0    90      1     1    1     1
11 493313   0 32.0    90   104      1     2    2     1
12 493313   0 32.0   104   166      0     1    3     2
13 589301   1 86.4     0   169      0     1    1     1
14 589302   0 28.8     0     8      1     1    1     1
15 589302   0 28.8     8    22      1     2    2     1
16 589302   0 28.8    22    63      1     1    3     2
17 589302   0 28.8    63    88      1     2    4     2
18 589302   0 28.8    88   169      0     1    5     3
```

Here `id` is the patient ID number, `trt` equals 1 for patients receiving rhDNase and 0 if they receive placebo, and `fev` is the forced respiratory volume measured at randomization. The variable `time1` is the start of a period indicating when subjects become "at risk" for a transition and `time2` is the time of the transition or a censoring time. The `status` variable equals 1 if `time2` is a transition time and equals 0 if it is a censoring time (i.e. the end of followup). The `etype` variable indicates the nature of the event time recorded

in time2; specifically, if etype=1 then time2 corresponds to the onset of an exacerbation (or censoring) and if etype=2, time2 corresponds to the time of a resolution of an exacerbation (or censoring). The enum variable simply records the cumulative number of lines in the data frame for each individual and enum1 the cumulative number of exacerbation-free periods.

Here we consider the data for the first two events, create a gap time variable, and center FEV.

```
> rhDNase.etype1 <- rhDNase[rhDNase$etype == 1,]
> rhDNase.etype1$gtime<-rhDNase.etype1$time2-rhDNase.etype1$time1
> rhDNase.etype1$fevc<-rhDNase.etype1$fev-
                 mean(rhDNase.etype1$fev[rhDNase.etype1$enum1==1])

Data for First and Second Gaps :
> rhDNase1 <- rhDNase.etype1[rhDNase.etype1$enum1 == 1,]
> rhDNase2 <- rhDNase.etype1[rhDNase.etype1$enum1 == 2,]
```

Here we create a data frame in which we can use the first gap time as a covariate in the second gap time analysis.

```
> temp <- rhDNase1[rhDNase1$status==1, c("id","gtime")]
> dimnames(temp)[[2]] <- c("id","gtime1")
> rhDNase2 <- merge(rhDNase2, temp, by="id", all.x=T)
> rhDNase2$gtime1c <- rhDNase2$gtime1 - mean(rhDNase2$gtime1)
```

Below is the code for fitting a Cox model to first and second gap times.

```
> fit1 <- coxph(Surv(gtime,status) ~ trt + fevc,
                data = rhDNase1, method = "breslow")
> summary(fit1)
  n= 645
          coef exp(coef) se(coef)     z       p
   trt -0.3828     0.682  0.12971 -2.95 3.2e-03
  fevc -0.0206     0.980  0.00277 -7.44 1.0e-13

       exp(coef) exp(-coef) lower .95 upper .95
   trt     0.682       1.47     0.529     0.879
  fevc     0.980       1.02     0.974     0.985

  Rsquare= 0.103   (max possible= 0.991 )
  Likelihood ratio test= 69.8  on 2 df,   p=6.66e-16
  Wald test            = 63.5  on 2 df,   p=1.65e-14
  Score (logrank) test = 65.9  on 2 df,   p=4.77e-15


> fit2 <- coxph(Surv(gtime,status) ~ trt + fevc + gtime1c,
                data = rhDNase2, method = "breslow")
> summary(fit2)
  n= 227
               coef exp(coef) se(coef)      z       p
```

```
    trt  0.358071     1.431  0.22456  1.595 0.11000
   fevc  0.000932     1.001  0.00538  0.173 0.86000
gtime1c -0.014310     0.986  0.00394 -3.634 0.00028

        exp(coef) exp(-coef) lower .95 upper .95
    trt     1.431      0.699     0.921     2.222
   fevc     1.001      0.999     0.990     1.012
gtime1c     0.986      1.014     0.978     0.993

Rsquare= 0.068   (max possible= 0.968 )
Likelihood ratio test= 15.9  on 3 df,   p=0.00121
Wald test            = 14.8  on 3 df,   p=0.00196
Score (logrank) test = 15.4  on 3 df,   p=0.00153
```

Model checking can be carried out using the S-PLUS function `cox.zph`, which does not provide evidence of problems with the assumption of multiplicative covariate effects.

```
> check1 <- cox.zph(fit1, transform="identity")
> check1
           rho chisq     p
   trt 0.0239 0.139 0.709
  fevc 0.0363 0.288 0.591
GLOBAL     NA 0.422 0.810
```

Gap time analysis may also be carried out using parametric log-normal models as in Section 4.3.2. The results summarized in Table 4.2 are based on models with treatment and FEV or treatment, FEV, and the previous gap time; the associated code is given below.

```
> fit1 <- survReg(Surv(gtime, status) ~ trt + fevc,
                  data=rhDNase1, dist="lognormal")
> summary(fit1)

              Value Std. Error     z        p
(Intercept) 5.4030     0.1048 51.53 0.00e+00
        trt 0.4302     0.1371  3.14 1.71e-03
       fevc 0.0217     0.0029  7.47 8.03e-14
 Log(scale) 0.3688     0.0512  7.21 5.68e-13

Scale= 1.45

Log Normal distribution
Loglik(model)= -1625   Loglik(intercept only)= -1660.8
        Chisq= 71.51 on 2 degrees of freedom, p= 3.3e-16
Number of Newton-Raphson Iterations: 3
n= 645
```

```
> rhDNase2$lgtime1 <- log(rhDNase2$gtime1)
> fit2 <- survReg(Surv(gtime, status) ~ trt + fevc + lgtime1,
                  data=rhDNase2, dist="lognormal")
> summary(fit2)

                Value Std. Error      z        p
(Intercept)  3.20899     0.4867  6.593 4.30e-11
        trt -0.22657     0.2105 -1.077 2.82e-01
       fevc -0.00454     0.0047 -0.965 3.34e-01
    lgtime1  0.41730     0.1322  3.157 1.59e-03
Log(scale)   0.20559     0.0859  2.392 1.67e-02

Scale= 1.23

Log Normal distribution
Loglik(model)= -484.9   Loglik(intercept only)= -490.9
        Chisq= 11.99 on 3 degrees of freedom, p= 0.0074
Number of Newton-Raphson Iterations: 3
n= 227
```

C.3 Code for Chronic Bronchitis Trial of Chapter 6

Below are the first few lines of the data frame for the semiparametric analysis of the exacerbation and interexacerbation times for the chronic bronchitis study given in Section 6.7.2. It is somewhat more involved than some of the previous data frames in that information is given for the time since the previous transition, the time since disease onset, the season, treatment, and other covariates.

```
 id enum etype estart estop gstart gstop gtime estatus xmark
1101    0     1      4    36      4    36    32       1     3
1101    1     2     36    97      0    61    61       1     3
1101    2     1     97   137      0    40    40       1     3
1101    3     2    137   369      0   232   232       0     3
1202    0     1      1    16      1    16    15       1     5
1202    1     2     16   178      0   162   162       1     5
1202    2     1    178   193      0    15    15       1     5
1202    3     2    193   263      0    70    70       1     5
1202    4     1    263   266      0     3     3       1     5
1202    5     2    266   327      0    61    61       1     5
1202    6     1    327   330      0     3     3       1     5
1202    7     2    330   363      0    33    33       0     5

xseason trt symptoms symptomsc gender severity
      1   1        4 -1.824324      0        0
      2   1        4 -1.824324      0        0
```

```
2   1        4 -1.824324      0        0
1   1        4 -1.824324      0        0
2   0        1 -4.824324      1        0
3   0        1 -4.824324      1        0
4   0        1 -4.824324      1        0
4   0        1 -4.824324      1        0
4   0        1 -4.824324      1        0
1   0        1 -4.824324      1        0
1   0        1 -4.824324      1        0
1   0        1 -4.824324      1        0
```

In this example, state 1 corresponded to the AECB-free condition and state 2 to the AECB condition. Here `enum` records the cumulative number of transitions of any sort over $[0, t]$, denoted $N_{i1}(t^-) + N_{i2}(t^-)$, and `enum` records the total number of transitions over $[0, t^-]$. The `etype` variable indicates whether the times associated with each row relate to transitions from the AECB to the AECB-free state (`etype=1`) or transitions from the AECB-free to the AECB state (`etype=2`). The variables `estart` and `estop` indicate the beginning and end of an "at risk" interval with time measured from the start of the exacerbation experienced at study entry, and `gstart` and `gstop` give the corresponding times on a semi-Markov scale (i.e. measured from the occurrence of the previous transition). The `estatus` variable is the censoring indicator which indicates whether the `estop` and `gstop` times correspond to transition times (`estatus=1`) or censoring times (`estatus=0`). To examine the possible effect of the time since diagnosis with chronic bronchitis, `xmark` reflects which of the following five year intervals the patient history falls into: `xmark` = 1, 2, 3, 4, 5, 6, 7, 8 for 0–5, 5–10, 10–15, 15–20, 20–25, 25–30, 30–35, 35–40 years duration, respectively. Seasonal effects are addressed through inclusion of `xseason` which takes the values 1 (Jan–March); 2 (April–Jun), 3 (July–Sept), or 4 (Oct–Dec). Additional covariates include `gender` (1 = female, 0 = male), `severity` (1 = severe disease, 0 = not severe), `trt` (1 = Ciprofloxacin, 0 = standard care), `trt.dt` (1 = Ciprofloxacin received and in the first exacerbation, 0 = standard care), and `symptomsc` represents the preceding days of AECB symptoms at randomization (centered by subtracting the mean number of symptom days, which is given by the variable `symptoms`).

The result of fitting a full model for the resolution of exacerbations is as follows.

```
> coxph(Surv(gstart,gstop,estatus) ~
            (trt+gender+severity+symptomsc)*strata(Ienum.gt.0) +
             factor(xmark)+factor(xseason)+
             strata(Ienum.gt.0)+frailty(id,distribution="gamma"),
             data=chest, subset=(etype == 1), method="breslow",
             control=coxph.control(eps=1e-06, iter.max=100))

  n= 820
```

```
                              coef se(coef)   se2 Chisq  DF       p
                       trt  0.5132 0.1585  0.1437 10.48 1.0 1.2e-03
                    gender -0.1993 0.1598  0.1441  1.56 1.0 2.1e-01
                  severity -0.2821 0.2485  0.2243  1.29 1.0 2.6e-01
                 symptomsc -0.1091 0.0159  0.0149 47.06 1.0 6.9e-12
             factor(xmark)2 -0.1590 0.1651  0.1343  0.93 1.0 3.4e-01
             factor(xmark)3 -0.2837 0.1648  0.1306  2.97 1.0 8.5e-02
             factor(xmark)4 -0.4618 0.1993  0.1576  5.37 1.0 2.1e-02
             factor(xmark)5 -0.4225 0.2247  0.1835  3.54 1.0 6.0e-02
             factor(xmark)6 -0.3069 0.2755  0.2167  1.24 1.0 2.7e-01
             factor(xmark)7  0.2637 0.3345  0.2672  0.62 1.0 4.3e-01
             factor(xmark)8 -0.7072 0.4193  0.3463  2.84 1.0 9.2e-02
             factor(xmark)9 -0.4688 0.3028  0.2307  2.40 1.0 1.2e-01
          factor(xseason)2  0.2911 0.1204  0.1135  5.84 1.0 1.6e-02
          factor(xseason)3  0.1803 0.1474  0.1406  1.50 1.0 2.2e-01
          factor(xseason)4  0.2037 0.1242  0.1182  2.69 1.0 1.0e-01
frailty(id, distribution                      100.29 58.4 5.4e-04
   trt:strata(Ienum.gt.0) -0.4528 0.1879  0.1794  5.81 1.0 1.6e-02
gender:strata(Ienum.gt.0)  0.1103 0.1907  0.1816  0.33 1.0 5.6e-01
severity:strata(Ienum.gt.  0.2879 0.2745  0.2633  1.10 1.0 2.9e-01
symptomsc:strata(Ienum.gt  0.0979 0.0177  0.0172 30.41 1.0 3.5e-08

Iterations: 8 outer, 36 Newton-Raphson
   Variance of random effect= 0.164   I-likelihood = -2845.4
Degrees of freedom for terms= 0.8 0.8 0.8 0.9 5.1 2.7 58.4 0.9
                              0.9 0.9 0.9
Rsquare= 0.277   (max possible= 0.999 )
Likelihood ratio test= 266  on 73.21 df,   p=0
```

Because the first exacerbation is incomplete and treatment was initiated in mid-exacerbation, the model is stratified according to whether it is the first or a subsequent exacerbation. We also fit covariate by strata interactions to give estimates of different covariate effects in the two strata.

The complementary aspect of this process is the onset of exacerbations. The result of fitting a full model for the associated events is given below; see Table 6.5.

```
> coxph(Surv(gstart,gstop,estatus) ~
            trt + gender + severity + factor(xmark) +
            factor(xseason) + frailty(id, distribution="gamma"),
            data=chest, subset=(etype == 2), method="breslow",
            control=coxph.control(eps=1e-06, iter.max=100))

  n= 595
                              coef se(coef) se2 Chisq   DF      p
                       trt -0.0369 0.124  0.105  0.09  1.0 0.7700
                    gender  0.2361 0.127  0.109  3.47  1.0 0.0630
                  severity  0.5621 0.172  0.141 10.69  1.0 0.0011
             factor(xmark)2 -0.3328 0.189  0.166  3.10  1.0 0.0780
```

```
          factor(xmark)3  0.2964 0.187  0.159  2.52  1.0 0.1100
          factor(xmark)4  0.2839 0.223  0.188  1.61  1.0 0.2000
          factor(xmark)5 -0.1330 0.263  0.227  0.26  1.0 0.6100
          factor(xmark)6  0.1054 0.294  0.251  0.13  1.0 0.7200
          factor(xmark)7  0.4789 0.375  0.307  1.63  1.0 0.2000
          factor(xmark)8  0.2534 0.504  0.428  0.25  1.0 0.6200
          factor(xmark)9  0.4307 0.324  0.265  1.77  1.0 0.1800
        factor(xseason)2 -0.0996 0.161  0.153  0.38  1.0 0.5400
        factor(xseason)3  0.3822 0.148  0.143  6.63  1.0 0.0100
        factor(xseason)4  0.0216 0.150  0.142  0.02  1.0 0.8900
frailty(id, distribution                      74.00 51.5 0.0220

  Iterations: 7 outer, 21 Newton-Raphson
     Variance of random effect= 0.198   I-likelihood = -2151.2
  Degrees of freedom for terms=  0.7  0.7  0.7  5.7  2.7 51.5
  Rsquare= 0.261   (max possible= 0.999 )
  Likelihood ratio test= 180  on 62.03 df,   p=2.15e-13
```

D

Datasets

The datasets below are available from the Web site for this book (www.stats.uwaterloo.ca/cook-lawless/book.shtml).

D.1 Bladder Cancer Data

Byar (1980) reported on patients with superficial bladder cancer who participated in a randomized clinical trial to assess the effect of experimental treatments on the recurrence of tumors. We consider data on 38 patients assigned to receive Thiotepa and 47 subjects assigned to placebo. The covariates include the treatment ($z_{i1} = 1$ for Thiotepa and $z_{i1} = 0$ for placebo), the diameter (in centimeters) of the largest tumor present at randomization (z_{i2}), and the number of tumors present at randomization (z_{i3}). The followup data include the times of inspection and the numbers of tumors detected at the inspections. In Table D.1 the times of inspections are listed for the first 20 subjects; if one or more tumors were detected at an inspection time the number is given in parentheses.

In the corresponding data frame, the counting process formulation is used with `visit` indicating the visit number, `estart` and `estop` the beginning and end of an interval, and `m` the number of tumors detected upon inspection at `estop`, followed by the covariates. Data for the first five subjects are as follows.

id	visit	estart	estop	m	z1	z2	z3
1	1	0	1	0	0	3	1
2	1	0	1	0	0	1	2
2	2	1	4	0	0	1	2
3	1	0	7	0	0	1	1
4	1	0	3	0	0	1	5
4	2	3	9	0	0	1	5
4	3	9	10	0	0	1	5
5	1	0	1	0	0	1	4

Table D.1. Data from 20 individuals in the bladder cancer study of Byar (1980).

ID	z_1	z_2	z_3		Months from Randomization (# Tumors Detected)							
1	0	3	1	1								
2	0	1	2	1	4							
3	0	1	1	7								
4	0	1	5	3	9	10						
5	0	1	4	1	4	6(1)	7	10				
6	0	1	1	3	10	14						
7	0	1	1	2	10	12(2)	16(3)	18				
8	0	1	1	3	14	18						
9	0	3	1	5(2)	10	12	18					
10	0	3	1	3	7	10(6)	15(3)	19	23			
11	0	1	1	1	3(8)	8	10	13	16(8)	19	23(8)	
12	0	1	3	3(1)	6	9(1)	10	13	15	17	20	21(8) 23
13	0	3	3	3	6	9	11	14	18	23		
14	0	3	2	3	7(8)	10(7)	13	16(5)	24(7)			
15	0	1	1	3(1)	6	9	12	15(1)	17	20	22	25(3)
16	0	1	8	1(8)	4	7	11	14	17	20	25	26
17	0	4	1	2(4)	6	26(8)						
18	0	2	1	3	6	13	16	22	23	26		
19	0	2	1	6	25(3)	28						
20	0	4	1	5	8	17	29					

```
5     2     1     4 0  0  1  4
5     3     4     6 1  0  1  4
5     4     6     7 0  0  1  4
5     5     7    10 0  0  1  4
```

D.2 Bowel Motility Data

Aalen and Husebye (1991) report on a study of the motility of the small bowel discussed in Section 1.3.2, where the data are given in Table 1.4. They provide data on 19 healthy subjects who were provided with a standard meal at 6:00 p.m. to induce a "fed state", and had their intraluminal pressure monitored overnight for 13 hours and 40 minutes. Interest lies in the migrating

motor complex (MMC), an activity front of a fasting cycle during the digestion process. Several MMC were observed for each individual and the times between these events are of interest. The first MMC defines the start of the fasting cycle and the times between the consecutive cycles define the gaps. The last MMC period is censored by the end of monitoring. A data frame for an analysis in S-PLUS or R is shown below, where `time` is the duration of a cycle, `status` indicates whether it was observed completely, and `enum` counts the number of cycles for an individual. Data for the first five individuals are shown.

id	time	status	enum
1	112	1	1
1	145	1	2
1	39	1	3
1	52	1	4
1	21	1	5
1	34	1	6
1	33	1	7
1	51	1	8
1	54	0	9
2	206	1	1
2	147	1	2
2	30	0	3
3	284	1	1
3	59	1	2
3	186	1	3
3	4	0	4
4	94	1	1
4	98	1	2
4	84	1	3
4	87	0	4
5	67	1	1
5	131	0	2

D.3 Pulmonary Exacerbations and rhDNase

Fuchs et al. (1994) report on a double-blind randomized multicenter clinical trial designed to assess the effect of rhDNase, a recombinant deoxyribonuclease I enzyme, versus placebo on the occurrence of respiratory exacerbations among patients with cystic fibrosis. The rhDNase operates by digesting the extracellular DNA released by leukocytes that accumulate in the lung as a result of bacterial infection (Therneau and Hamilton, 1997), and so it was expected that aerosol administration of rhDNase would reduce the incidence of exacerbations.

Data on the occurrence and resolution of all exacerbations were recorded over approximately 169 days of followup for 645 patients in this trial; the

data are discussed in some detail in Therneau and Grambsch (2000). Part of the data is given in Table D.2 for the first 20 patients. We include a patient identifier, the treatment assignment (T) (1 = rhDNase, 0 = placebo), two baseline measurements of forced expiratory volume (FEV_1 and FEV_2) reflecting lung capacity, and the date of randomization. In addition, the number of days from randomization to the beginning (B) of the exacerbations is recorded, as well as the day on which treatment for each exacerbation ended (E) and patients became at risk of a new exacerbation. Therefore, for patient number 589302, the first exacerbation began 8 days after randomization and antibiotic therapy for this exacerbation ended 22 days after randomization. The patient then remained at risk until the second exacerbation, which began 63 days after randomization, and became at risk again after therapy ended on day 88; the patient did not have another exacerbation over the remainder of followup which ended on day 169.

Table D.2. Data from rhDNase study for first 20 subjects.

					Onset and Resolution Times						
					1st		2nd		3rd		Cens.
ID	T	FEV_1	FEV_2	Rand. Date	B	E	B	E	B	E	Time
493301	1	28.8	28.1	20/03/1992							168
493303	1	64.0	63.0	24/03/1992							169
493305	0	67.2	68.7	24/03/1992	65	75					168
493309	1	57.6	56.5	26/03/1992							168
493310	0	57.6	56.3	24/03/1992							171
493311	1	25.6	25.3	27/03/1992							166
493312	0	86.4	85.4	27/03/1992							168
493313	0	32.0	32.4	28/03/1992	90	104					166
589301	1	86.4	86.0	27/02/1992							169
589302	0	28.8	29.2	06/03/1992	8	22	63	88			169
589303	0	112.0	110.7	28/02/1992	60	74	83	124			169
589305	0	70.4	71.7	04/03/1992	50	68					169
589307	1	96.0	94.5	05/03/1992							169
589309	0	44.8	44.6	05/03/1992	99	114					169
589310	1	70.4	70.1	06/03/1992	35	64	71	108			169
589311	1	54.4	53.8	11/03/1992							169
589312	0	73.6	73.2	12/03/1992	8	13					196
589313	1	96.0	97.2	12/03/1992							169
589314	0	105.6	107.0	12/03/1992							169
589316	1	80.0	79.4	19/03/1992							167

D.4 Software Debugging Data

Section 1.2.2 introduced the software debugging example in which the number of faults found in a software system was recorded over a 160-day testing period. Table D.3 contains the full dataset. The data frame, whose first several lines are shown below, is constructed like Table D.3, with `t` representing the total testing time in terms of total staff days to date, `Nt` the number of faults detected up to time t, and `Ct` the number of lines of code changed (or added) up to time t.

t	Nt	Ct
0	0	0
4.8	0	16012
6	0	16012
14.3	7	32027
22.8	7	48042
32.1	7	58854
41.4	7	69669
51.2	11	80483
60.6	12	91295
70	13	102110
79.9	15	112925
91.3	20	120367
97	21	127812
107.7	22	135257
119.1	28	142702
127.6	40	150147
135.1	44	152806
142.8	46	155464
148.9	48	158123
156.6	52	167081
163.9	52	167704
169.7	59	174626
170.1	59	174626
170.6	59	174626
174.7	63	181548
179.6	68	188473
185.5	71	194626
194.0	88	200782
200.3	93	206937

D.5 Artificial Field Repair Data

In Example 8.1 some results were presented on the analysis of an artificial dataset on field repairs. The data were generated as follows. The time-homogeneous event rate for subject i was gamma distributed with mean 2 and variance 0.5 and this

Table D.3. Software debugging data from Dalal and McIntosh (1994).

t	N(t)	C(t)	t	N(t)	C(t)	t	N(t)	C(t)
0.0	0	0	285.5	186	247946	578.3	457	306902
4.8	0	16012	294.2	190	251016	587.2	467	307849
6.0	0	16012	295.7	190	251016	595.5	473	308795
14.3	7	32027	298.0	190	254086	605.6	480	309742
22.8	7	48042	305.2	195	257155	613.9	491	310688
32.1	7	58854	312.3	201	260225	621.6	496	311635
41.4	7	69669	318.2	209	260705	623.4	496	311635
51.2	11	80483	328.9	224	261188	636.3	502	311750
60.6	12	91295	334.8	231	261669	649.7	517	311866
70.0	13	102110	342.7	243	262889	663.9	527	312467
79.9	15	112925	350.5	252	263629	675.1	540	313069
91.3	20	120367	356.3	259	264367	677.4	543	313069
97.0	21	127812	360.6	271	265107	677.9	544	313069
107.7	22	135257	365.7	277	265845	688.4	553	313671
119.1	28	142702	374.9	282	266585	698.1	561	314273
127.6	40	150147	386.5	290	267325	710.5	573	314783
135.1	44	152806	396.5	300	268607	720.9	581	315294
142.8	46	155464	408.0	310	269891	731.6	584	315805
148.9	48	158123	417.3	312	271175	732.7	585	315805
156.6	52	167081	424.9	321	272457	733.6	585	315805
163.9	52	167704	434.2	326	273741	746.7	586	316316
169.7	59	174626	442.7	339	275025	761.0	598	316827
170.1	59	174626	451.4	346	276556	776.5	612	318476
170.6	59	174626	456.1	347	278087	793.5	621	320125
174.7	63	181548	460.8	351	279618	807.2	636	321774
179.6	68	188473	466.0	356	281149	811.8	639	321774
185.5	71	194626	472.3	359	283592	812.5	639	321774
194.0	88	200782	476.4	362	286036	829.0	648	323423
200.3	93	206937	480.9	367	288480	844.4	658	325072
207.2	97	213093	486.8	374	290923	860.5	666	326179
211.9	98	219248	495.8	376	293367	876.7	674	327286
217.0	105	221355	505.7	380	295811	892.0	679	328393
223.5	113	223462	516.0	392	298254	895.5	686	328393
227.0	113	225568	526.2	399	300698	910.8	690	329500
234.1	122	227675	527.3	401	300698	925.1	701	330608
241.6	129	229784	535.8	405	303142	938.3	710	330435
250.7	141	233557	546.3	415	304063	952.0	720	330263
259.8	155	237330	556.1	425	305009	965.0	729	330091
268.3	166	241103	568.1	440	305956	967.7	729	330091
277.2	178	244879	577.2	457	306902	968.6	731	330091

Table D.3. Software debugging data from Dalal and McIntosh (1994) - continued.

t	N(t)	C(t)	t	N(t)	C(t)	t	N(t)	C(t)
981.3	740	329919	1139.1	805	331852	1279.8	854	339943
1013.9	759	330036	1163.2	823	332167	1287.4	855	341955
1030.1	776	330326	1174.3	827	332391	1295.1	859	341967
1044.0	781	330616	1184.6	832	332615	1304.8	860	341979
1047.0	782	330616	1198.3	834	332839	1305.8	865	342073
1059.7	783	330906	1210.3	836	333053	1313.3	867	342168
1072.6	787	331196	1221.1	839	333267	1314.4	867	342168
1085.7	793	331486	1230.5	842	333481	1320.0	867	342262
1098.4	796	331577	1231.6	842	333481	1325.3	867	342357
1112.4	797	331669	1240.9	844	333695	1330.6	870	342357
1113.5	798	331669	1249.5	845	333909	1334.2	870	342358
1114.1	798	331669	1262.2	849	335920	1336.7	870	342358
1128.0	802	331760	1271.3	851	337932			

was used to generate events over $(0, \tau_i]$, where $\tau_i \sim$ Unif(1,3). At the jth event time for subject i, the cost was generated independently as $C_{ij} \sim N(10, 2.5^2)$. Table D.4 contains the simulated data for the first ten individuals.

Table D.4. Artifical field repair data.

ID	τ	Variable	Time and Cost							
1	1.66	Time	0.11	0.36	0.86	0.97	1.22			
		Cost	8.36	11.73	12.77	6.36	10.86			
2	2.17	Time	1.19							
		Cost	11.15							
3	2.65	Time								
		Cost								
4	2.12	Time	0.04	0.23	0.67	1.13	1.24	1.34	1.66	2.12
		Cost	12.09	11.33	8.38	11.51	5.54	10.84	11.40	13.05
5	1.83	Time								
		Cost								
6	2.81	Time	0.41	1.00	1.18	1.61	2.00	2.40		
		Cost	8.15	6.63	8.71	13.53	10.46	9.89		
7	1.83	Time	0.02	0.99	1.15					
		Cost	6.55	7.58	10.63					
8	1.27	Time	0.20	0.38	0.69	0.89	1.06			
		Cost	11.44	10.61	6.62	8.20	4.96			
9	1.60	Time	0.15	0.45	0.50	0.62	0.86	1.16	1.42	
		Cost	5.83	8.76	12.47	7.41	10.05	10.64	9.30	
10	2.27	Time	0.16	1.29	1.41	2.09				
		Cost	11.82	10.84	12.92	10.62				

References

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726.

Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Springer Lect. Notes Statist.* **2**, 1–25. Springer–Verlag, New York.

Aalen, O.O. (1992). Modelling heterogeneity in survival analysis by the compound Poisson distribution. *Ann. Appl. Prob.* **2**, 951–972.

Aalen, O.O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statist. Med.* **10**, 1227–1240.

Aalen, O.O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on on censored observations. *Scand. J. Statist.*. **5**, 141–150.

Aalen, O., Borgan, O., and Fekjaer, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* **57**, 993–1001.

Aalen, O., Fosen, J., Weedon-Fekjaer, H., Borgan, O., and Husebye, E. (2004). Dynamic analysis of multivariate failure time data. *Biometrics* **60**, 764–773.

Abu-Libdeh, H., Turnbull, B.W., and Clark, L.C. (1990). Analysis of multi-type recurrent events in longitudinal studies: application to a skin cancer prevention trial. *Biometrics* **46**, 1017–1034.

ACIP Investigators (1992). Asymptomatic cardiac ischaemia pilot study (ACIP). *Am. J. Cardiology* **70**, 744–747.

Allison, P.D. (1984). *Event History Analysis: Regression for Longitudinal Data.* Sage, Beverly Hills, CA.

Allison, P.D. (1985). Survival analysis of backward recurrence times. *J. Am. Statist. Assoc.* **80**, 315–322.

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1120.

Andersen, P.K. and Green, A. (1985). Robustness to differential mortality of incidence estimation in an illness-death-emigration model. *Scand. J. Statist.* **12**, 63–68.

Andersen, P.K. and Liestol, K. (2003). Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics* **4**, 633–649.

Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer–Verlag, New York.

Andreev, A. and Arjas, E. (1998). Acute middle ear infections in small children: A Bayesian analysis using multiple time scales. *Lifetime Data Anal.* **4**, 121–137.

Arjas, E. and Andreev, A. (2000). Predictive inference, causal reasoning, and model assessment in nonparametric Bayesian analysis: A case study. *Lifetime Data Anal.* **6**, 187–205.

Ascher, H. and Feingold, H. (1984). *Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes*. Marcel Dekker, New York.

Asgharian, M., M'Lan, C.E., and Wolfson, D.B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Am. Statist. Assoc.* **97**, 201–209.

Baker, S.G., Wax, Y., and Patterson, B.H. (1993). Regression analysis of grouped survival data: Informative censoring and double sampling. *Biometrics* **49**, 379–389.

Balshaw, R.F. and Dean, C.B. (2002). A semiparametric model for the analysis of recurrent event panel data. *Biometrics* **58**, 324–331.

Bang, H. and Tsiatis, A.A. (2000). Estimating medical costs with censored data. *Biometrika* **87**, 329–343.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Baxter, L.D. (1994). Estimation from quasi life tables. *Biometrika* **81**, 567–577.

Baydar, N. and White, M. (1988). A method for analyzing backward recurrence time data on residential mobility. *Sociol. Methodol.* **18**, 105–135.

Beran, R. (1990). Calibrating prediction regions. *J. Am. Statist. Assoc.* **85**, 715–723.

Berman, M. (1981). Inhomogeneous and modulated gamma processes. *Biometrika* **68**, 163–152.

Berman, M. and Turner, T.R. (1992). Approximating point process likelihoods. *Applied Statist.* **41**, 31–38.

Bernardo, M.V and Harrington D.P. (2001). Sample size calculations for the two-sample problem using the multiplicative intensity model. *Statist. Med.* **20**, 557–579.

Blossfeld, H.P. and Rohwer, G. (1995). *Techniques of Event History Modeling*. L. Erlbaum, Hillsdale, NJ.

Boher, J. and Cook, R.J.(2006). Implications of model misspefication in robust tests for recurrent events. *Lifetime Data Anal.* **12**, 69–95.

Boos, D.D. (1992). On generalized score tests. *Am. Statistician* **46**, 327–333.

Borgan, O. and Hoem, J. (1988). Demographic reproduction rates and the estimation of an expected total count per person in an open population. *J. Am. Statist. Assoc.* **83**, 886–891.

Borgan, D., Fiaccone, R.L., Henderson, R., and Barreto, M.L. (2005). Dynamic analysis of recurrent event data with missing observations, with application to infant diarrhoea in Brazil. University of Oslo Dept. of Math. Statistical Research Report No 9.

Box-Steffensmeier, J.M. and De Boef. S. (2006). Repeated events survival models: The conditional frailty model. *Statist. Med.* **25**, 3518–3533.

Brookmeyer, R. and Gail, M.H. (1994). *AIDS Epidemiolgy: A Quantitative Approach*. Oxford University Press, Oxford.

Brown, T.C. and Nair, M.G. (1988). A simple proof of the multivariate random time change theorem for point processes. *J. Appl. Prob.* **25**, 210–214.

Byar, D.P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage 1 bladder tumors: Comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, 363–370. Eds. M. Pavone-Macaluso, P.H. Smith, and F. Edsmyr. Plenum, New York.

Byar, D., Kaihara, R., Sylvester, R., Freedman, L., Hannigan, J., Koiso, K., Oohashi, Y., and Tsugawa, R. (1986). Statistical analysis techniques and sample size determination for clinical trials of treatments for bladder cancer. *In Developments in Bladder Cancer*, 49–64, Alan R. Liss, New York.

Cai, J. and Schaubel, D.E. (2004a). Analysis of recurrent event data. In *Handbook of Statistics*, vol. 23. Elsevier, New York.

Cai, J. and Schaubel, D.E. (2004b). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Anal.* **10**, 121–138.

Cai, J. and Schaubel D.E. (2005). Analysis of clustered recurrent event data with application to hospitalization rate among renal failure patients. *Biostatistics* **6**, 404–419.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, UK.

Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models*, 2nd edition. Chapman and Hall/CRC Press, Boca Raton, FL.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In *Longitudinal Analysis of Labor Market Data*. Eds. J.J. Heckman and B. Singer. Cambridge University Press, Cambridge, UK.

Chang, S.H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Anal.* **10**, 175–190.

Chang, S.H. and Tzeng, S.J. (2006). Nonparametric estimation of sojourn time distribution for truncated serial event data - A weight-adjusted approach. *Lifetime Data Anal.* **12**, 53–67.

Chang, S.H. and Wang, M.C. (1999). Conditional regression analysis for recurrence time data. *J. Am. Statist. Assoc.* **94**, 1221–1230.

Chen, H.Y. and Little, R.J.A. (1999). Proportional hazards regression with missing covariates. *J. Am. Statist. Assoc.* **94**, 896–908.

Chen, B.E., Cook, R.J., Lawless, J.F., and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statist. Med.* **24**, 671–691.

Chen, B.E. and Cook, R.J. (2003). Regression modeling with recurrent events and time-dependent interval-censored marker data. *Lifetime Data Anal.* **9**, 275–291.

Chen, E.B. and Cook, R.J. (2004). Tests for multivariate recurrent events in the presence of a terminal event: Application to studies of cancer metastatic to bone. *Biostatistics* **5**, 129–143.

Cheuvart, B. (1988). A nonparametric model for multiple occurrences. *Applied Statist.* **37**, 157–168.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrics* **65**, 141–151.

Clayton, D.G. (1994). Some approaches to the analysis of recurrent event data. *Statist. Meth. Med. Res.* **3**, 244–262.

Cook, J.A. and Lawless, J.F. (1991). Two-sample tests with multinomial or grouped failure time data. *Biometrics* **47**, 445–459.

Cook, R.J. (1995). The design and analysis of randomized trials with recurrent events. *Statist. Med.* **14**, 2081–2098.

Cook, R.J. and Lawless, J.F. (1996). Interim monitoring of longitudinal comparative studies with recurrent event responses. *Biometrics* **52**, 1311–1323.

Cook, R.J. and Lawless, J.F. (1997). Marginal analysis of recurrent events and a terminating event. *Statist. Med.* **16**, 911–924.

Cook, R.J. and Lawless, J.F. (2002). Analysis of repeated events. *Statist. Meth. Med. Res.* **11**, 141–166.

Cook, R.J. and Wei, W. (2002). Selection effects in randomized trials with count data. *Statist. Med.* **21**, 515–531.

Cook, R.J. and Wei, W. (2003). Conditional analysis of mixed Poisson processes with baseline counts: Implications for trial design and analysis. *Biostatistics* **4**, 479–494.

Cook, R.J., Lawless, J.F., and Lee, K.A. (2003). Cumulative processes related to event histories. *SORT* **27**, 13–30.

Cook, R.J., Lawless, J.F., and Nadeau, J.C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52**, 732–739.

Cook, R.J., Lee, K.A., and Li, H. (2007). Non-inferiority trial design for recurrent events. *Statist. Med.*.

Cook, R.J., Ng, E.J.M., Mukherjee, J., and Vaughan, D. (1999). Two-state mixed renewal processes for chronic disease. *Statist. Med.* **18**, 175–188.

Cook, R.J., Wei, W., and Yi, G.Y. (2005). Robust tests for treatment effects based on censored recurrent event data over multiple periods. *Biometrics* **61**, 692–701.

Couper, D. and Pepe, M.S. (1997). Modelling prevalence of a condition: Chronic graft-versus-host disease after bone marrow transplantation. *Statist. Med.* **16**, 1551–1571.

Cox, D.R. (1969). Discussion of paper by P.A.P. Moran, *J. Roy. Statist. Soc. A* **132**, 521–522.

Cox, D.R. (1972a). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B* **34**, 187–220.

Cox, D.R. (1972b). The statistical analysis of dependencies in point processes. In *Symposium on Point Processes*, Ed. P.A.W. Lewis. John Wiley and Sons, New York.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Cox, D.R. and Isham, V. (1980). *Point Processes*. Chapman and Hall, London.

Cox, D.R. and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events*. Chapman and Hall, London.

Craiu, R.V. and Duchesne, T. (2004). Inference based on the EM algorithm for the competing risk model with masked causes of failure. *Biometrika* **91**, 543–558.

Craiu, R.V. and Reiser, B. (2006). Inference for the dependent competing risks model with masked causes of failure. *Lifetime Data Anal.* **12**, 21–33.

Crowder, M.J. (1987). On linear and quadratic estimating functions. *Biometrika* **74**, 591–597.

Dabrowska, D.M., Sun, G., and Horowitz, M.M. (1994). Cox regression in a Markov renewal model: An application to the analysis of bone marrow transplant data. *J. Am. Statist. Assoc.* **89**, 867–877.

Dalal, S.R. and McIntosh, A.A. (1994). When to stop testing for large software systems with changing code. *IEEE Trans. Software Eng.* **20**, 318–323.

Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.

Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Vol 1: Elementary Theory and Methods*. Springer, New York.

Datta, S. and Satten, G.A. (2001). Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. *Statist. Prob. Letters* **55**, 403–411.

Datta, S. and Satten, G.A. (2002). Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under dependent censoring. *Biometrics*, **58**, 792–802.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.

Dean, C. and Lawless, J.F. (1989). Tests for detecting overdispersion in Poisson regression models. *J. Am. Statist. Assoc.* **84**, 467–472.

Dean, C.B. (1991). Estimating equations for mixed Poisson models. In *Estimating Functions*, 35–46. Ed. V.P. Godambe. Clarendon Press, Oxford.

Dean, C.B. (1992). Testing for overdispersion in Poisson and binomial regression models. *J. Am. Statist. Assoc.* **87**, 451–457.

Dean, C.B. and Balshaw, R. (1997). Efficiency lost by analysing counts rather than event times in Poisson and overdispersed Poisson regression models. *J. Am. Statist. Assoc.* **92**, 1387–1398.

DeMets, D.L. (1984). Stopping guidelines vs. stopping rules: A practitioner's point of view. *Commun. Statist. (A) – Theory and Methods* **13**, 2395–2417.

Dempster, A.P., Laird N.M., and Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* **39**, 1–38.

Dewanji, A. and Sengupta, D. (2003). Estimation of competing risks with general missing pattern in failure types. *Biometrics* **59**, 1063–1070.

Diamond, I.R. and McDonald, J.W. (1992). The analysis of current status data. In *Demographic Applications of Event History Analysis*, Eds. J. Trussel, R. Hankinson, and J. Tilton. Clarendon Press, Oxford.

Diggle, P.J. and Kenward, M.G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statist.* **43**, 49–94.

Diggle, P., Farewell, D., and Henderson, R. (2007). Analysis of longitudinal data with drop-out: Objectives, assumptions and a proposal (with discussion). *Applied Statist.* **56**.

Diggle, P., Heagerty, P., Liang K-Y., and Zeger, S. (2002). *Analysis Of Longitudinal Data.* Oxford University Press, London.

Eerola, M., Gasbarra, D., Makela, P.H., Linden, H., and Andreev, A. (2003). Joint modelling of recurrent infections and antibody response by Bayesian data augmentation. *Scand. J. Statist.* **30**, 677–698.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Erto, P. (1989). Reliability assessments by repair shops via maintenance data. *J. Appl. Statist.* **16**, 303–313.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd edition. Springer–Verlag, New York.

Farewell, V.T., Lawless, J.F., Gladman, D.D., and Urowitz, M.B. (2003). Tracing studies and analysis of the effect of loss to follow-up on the estimation of mortality from patient registry data. *Applied Statist.* **52**, 445–456.

Farrington, C.P. and Whitaker, H.J. (2006). Semiparametric analysis of case series data (with discussion). *Applied Statist.* **55**, 553–594.

Faulkenberry, D.G. (1973). A method of obtaining prediction intervals. *J. Am. Statist. Assoc.* **68**, 433–435.

Feigin, P.D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Adv. Appl. Prob.* **8**, 712–736.

Fisher, R.A. (1950). The significance of deviations from expectation in a Poisson series. *Biometrics* **6**, 17–24.

Fitzmaurice, G.M., Laird, N.M., and Ware, J.H. (2004). *Applied Longitudinal Analysis*. John Wiley & Sons, Hoboken, NJ.

Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. John Wiley and Sons, New York.

Follman, D. and Goldberg, M. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* **30**, 389–396.

Fong, D.Y.T., Lam, K.F., Lawless, J.F., and Lee, Y.W. (2001). Dynamic random effects models for times between repeated events. *Lifetime Data Anal.* **7**, 345–362.

Fosen, J., Ferkingstad, E., Borgan, O., and Aalen, O.O. (2006a). Dynamic data analysis – A new approach to analyzing time-dependent covariates. *Lifetime Data Anal.* **12**, 143–167.

Fosen, J., Borgan, O., Weedon-Fekjaer, H., and Aalen, O.O. (2006b). Dynamic analysis of recurrent event data using the additive hazard model. *Biometrical J.* **48**, 381–398.

Frangakis, C.E. and Rubin D.B. (2001). Addressing the idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics* **57**, 333–353.

Fredette, M. and Lawless, J.F. (2007). Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims. *Technometrics* **49**, 66–80.

Freedman, L., Sylvester, R., and Byar, D.P. (1989). Using permutation tests and bootstrap confidence intervals to analyse repeated events data from clinical trials. *Controlled Clin. Trials* **10**, 129–141.

Fuchs, H.J., Borowitz, D.S., Christiansen, D.H., Morris, E.M., Nash, M.L., Ramsey, B.W., Rosenstein, B.J., Smith, A.L., and Wohl, M.E.

(1994). Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *New Eng. J. Med.* **331**, 637–642.

Gail, M.H., Santner, T.J., and Brown, C.C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36**, 255–266.

Gao, S. and Zhou, X-H. (1997). An empirical comparison of two semi-parametric approaches for the estimation of covariate effects from multivariate failure time data. *Statist. Med.* **16**, 2049–2062.

Gaver, D.P. and O'Muircheartaigh, I.G. (1987). Robust empirical Bayes analyses of event rates. *Technometrics* **29**, 1–15.

Gelber, R.D., Cole, B.F., Gelber, S., and Goldhirsch, A. (1995). Comparing treatments using quality-adjusted survival: The Q-TWIST method. *Am. Statistician* **49**, 161–169.

Ghosh, D. (2004). Accelerated rates regression models for recurrent events. *Lifetime Data Anal.* **10**, 247–261.

Ghosh, D. and Lin, D.Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554–562.

Ghosh, D. and Lin, D.Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* **12**, 663–688.

Ghosh, D. and Lin, D.Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59**, 877–885.

Gill, R. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahr. Verw. Geb.* **53**, 97–116.

Gill, R.D., van der Laan, M.J., and Robins, J.M. (1997). Coarsening at random: Characterizations, conjectures and counter-examples. In *Proc. First Seattle Symposium in Biostatistics*. Eds. D.Y. Lin and T.R. Fleming. Springer–Verlag, New York.

Glasziou, P.P., Simes, R.J., and Gelber, R.D. (1990). Quality adjusted survival analysis. *Statist. Med.* **9**, 1259–1276.

Glidden, D.V. (2002). Robust inference for event probabilities with non-Markov event data. *Biometrics* **58**, 361–368.

Gomez, G., Serrat, C., and Ruiz, L. (2006). Weighted conditional survival estimator for ordered failure times subject to a common censoring process. Manuscript.

Grambsch, P.M. and Therneau, T.M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.

Grandell, J. (1997). *Mixed Poisson Processes*. Chapman and Hall, London.

Grossman, R., Mukherjee, J., Vaughan, D., Eastwood, C., Cook, R.J., LaForge, J., and Lampron, N. (1998). The Canadian Ciprofloxacin Health Economic Study Group. A 1-year community-based health economic study of ciprofloxacin vs usual antibiotic treatment in acute exacerbations of chronic bronchitis. *CHEST* **113**, 131–141.

Gruger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics* **47**, 595–605.

Guo, G. (1993). Event-history analysis for left-truncated data. *Sociol. Methodol.* **23**, 217–243.

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. Chapman and Hall, London.

Hamerle, A. (1991). On the treatment of interrupted spells and initial conditions in event history analysis. *Sociol. Methods and Res.* **19**, 388–414.

He, W. and Lawless, J.F. (2003). Flexible maximum likelihood methods for bivariate proportional hazards models. *Biometrics* **59**, 837–848.

He. W. and Lawless, J.F. (2005). Bivariate location-scale models for regression analysis, with applications to lifetime data. *J. Roy. Statist. Soc. B* **67**, 63–78.

Heckman, J.J. and Singer, B. (1985). *Longitudinal Analysis of Labor Market Data* (editors). Cambridge University Press, Cambridge, UK.

Heckman, J.J. and Singer, B. (1986). Econometric analysis of longitudinal data. Ch. 29 in *Handbook of Econometrics*, Vol. 3, Eds. Z. Griliches and M.D. Intriligator. North Holland, Amsterdam.

Heckman, J.J. and Walker, J.R. (1992). Understanding third births in Sweden. Ch. 7 in *Demographic Applications of Event History Analysis*, Eds. J. Trussel, R. Hankinson, and J. Tilton. Clarendon Press, Oxford.

Heyde, C. (1997). *Quasi-Likelihood and Its Application*. Springer, New York.

Hjort, N.L. (1990a). Goodness-of-fit tests for composite hypotheses in hazard based models. *Ann. Statist.* **18**, 1221–1258.

Hjort, N.L. (1990b). Nonparametric Bayes estimators based on beta processes in models of life history data. *Ann. Statist.* **18**, 1259–1294.

Hoem, J.M. (1985). Weighting, misclassification, and other issues in the analysis of survey samples of life histories. Chapter 5 in *Longitudinal Analysis of Labor Market Data*, Eds. J.J. Heckman and B. Singer. Cambridge University Press, Cambridge, UK.

Holford, T.R. 1980. The analysis of rates and survivorship using log-linear models. *Biometrics* **36**, 299–305.

Hortobagyi, G.N., Theriault, R.L., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J.F., Seaman, J., Knight, R.D., Heffernan, M., Reitsma, D.J., Kennedy, I., Allan, S.G., and Mellars, K. for the Protocol 19 Aredia Breast Cancer Study Group (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. Protocol 19 Aredia Breast Cancer Study Group. *New Eng. J. Med.* **335**, 1785–1791.

Hortobagyi, G.N., Theriault, R.L., Lipton, A., Porter, L., Blayney, D., Sinoff, C., Wheeler, H., Simeone, J.F., Seaman, J., Knight, R.D., Heffernan, M., Mellars, K., and Reitsma, D.J. (1998). Long-term prevention

of skeletal complications of metastatic breast cancer with pamidronate. *J. Clin. Oncol.* **16**, 2038–2044.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer–Verlag, New York.

Hu, C. and Lin, D.Y. (2002). Cox regression with covariate measurement error. *Scand. J. Statist.* **29**, 637–655.

Hu, X.J. and Lawless, J.F. (1996a). Estimation of rate and mean functions from truncated recurrent event data. *J. Amer. Statist. Assoc.* **91**, 300–310.

Hu, X.J. and Lawless J.F. (1996b). Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika* **83**, 747–761.

Hu, X.J., Sun, J., and Wei, L.J. (2002). Regression parameter estimation from panel counts. *Scand. J. Statist.* **29**, 1–19.

Huang, Y. (2002). Censored regression with the multistate accelerated sojourns model. *J. Roy. Statist. Soc. B* **64**, 17–29.

Huang, Y. and Chen, Y.Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Anal.* **9**, 293–303.

Huang, Y. and Wang, M.C. (2003). Frequency of recurrent events at failure time: Modeling and inference. *J. Am. Statist. Assoc.* **98**, 663–670.

Huang, Y. and Wang, M.C. (2004). Joint modeling and estimation of recurrent event processes and failure time. *J. Am. Statist. Assoc.* **99**, 1153–1165.

Hughes, M.D. (1997). Power considerations for clinical trials using multivariate time-to-event data. *Statist. Med.* **16**, 865–882.

Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis.* Springer, New York.

Ibrahim, J.G., Chen, M.-H., Lipsitz, S.R., and Herring, A.H. (2005). Missing-data methods for generalized linear models. *J. Am. Statist. Assoc.* **100**, 332–346.

Ishwaran, H. and James, L.F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. *J. Am. Statist. Assoc.* **99**, 175–190.

Jennison, C. and Turnbull, B.W. (2000). *Group Sequential Methods with Applications to Clinical Trials.* Chapman and Hall/CRC Press, London.

Jiang, S-T., Landers, T.L., and Rhoads, T.R. (2006). Assessment of repairable-system reliability using proportional intensity models: A review. *IEEE Trans. Rel.* **55**, 328–336.

Jiang, W. (1999). Group sequential procedures for repeated events data with frailty. *J. Biopharm. Statist.* **9**, 379–399.

Jiang, W., Turnbull, B.W., and Clark, L.E. (1999). Semiparametric regression models for repeated events with random effects and measurement error. *J. Am. Statist. Assoc.* **94**, 853–864.

Joe, H. (1997). *Multivariate Dependence Concepts*. Chapman and Hall, London.

Kalbfleisch J.D. and Lawless, J.F.(1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statist. Med.* **7**, 149–160.

Kalbfleisch J.D. and Lawless, J.F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Am. Statist. Assoc.* **84**, 360–372.

Kalbfleisch, J.D. and Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edition. John Wiley and Sons, New York.

Karlis, D. and Xekalaki (2005). Mixed Poisson distributions. *Int. Statist. Rev.* **73**, 35–58.

Karr, A.F. (1991). *Point Processes and Their Statistical Inference*, 2nd edition. Marcel Dekker, New York.

Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective. *J. Roy. Statist. Soc. A* **154**, 371–412.

Keiding, N. (2006). Event history analysis and the cross-section. *Statist. Med.* **14**, 2343–2364.

Keiding, N., Andersen, C., and Fledelius, P. (1998). Cox regression model for claims data in non-life insurance. *Astin Bull.* **28**, 95–118.

Keiding, N., Kvist, K., Hartvig, H., Tvede, M., and Juul, S. (2002). Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* **3**, 565–578.

Kessing, L.V., Andersen, P.K., Martensen, P.B., and Bolwig, T.G. (1998). Recurrence in affective disorder. *Brit. J. Psych.* **172**, 23–28.

Kessing, L.V., Hansen, M.G., and Andersen, P.K. (2004). Course of illness in depressive bipolar disorders: Naturalistic study, 1994–1999. *Brit. J. Psych.* **185**, 372–377.

Kessing, L.V., Olsen, E.W., and Andersen, P.K. (1999). Recurrence in affective disorder: Analyses with frailty models. *Am. J. Epidem.* **149**, 404–411.

Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–806.

Korn, E.L. and Graubard, B.I. (1999). *Analysis of Health Surveys*. John Wiley and Sons, New York.

Lan, K.K.G. and DeMets, D.L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–663.

Lan, K.K.G. and DeMets, D.L. (1989). Group sequential procedures: Calendar versus information time. *Statist. Med.* **8**, 1191–1198.

Laslett, G.M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika* **69**, 153–160.

Lawless, J.F. (1987a). Regression methods for Poisson process data. *J. Am. Statist. Assoc.* **82**, 808–815.

Lawless, J.F. (1987b). Negative binomial and mixed Poisson regression. *Canad. J. Statist.* **15**, 209–225.

Lawless, J.F. (1995). The analysis of recurrent events for multiple subjects. *Applied Statist.* **44**, 487–498.

Lawless, J.F. (1998). Statistical analysis of product warranty data. *Int. Statist. Rev.* **66**, 41–60.

Lawless, J.F. (2003a). *Statistical Models and Methods for Lifetime Data*, 2nd edition. Wiley, Hoboken, NJ.

Lawless, J.F. (2003b). Event history analysis and longitudinal surveys. In *Analysis of Survey Data*, Eds. R.L. Chambers and C.J. Skinner. Wiley, Chichester.

Lawless, J.F. (2006). Models for software testing and prediction. Manuscript.

Lawless, J.F. and Fong, D.Y.T. (1999). State duration models in clinical and observational studies. *Statist. Med.* **18**, 2365–2376.

Lawless, J.F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92**, 529–542.

Lawless, J.F. and Kalbfleisch, J.D. (1992). Some issues arising in the collection and analysis of field reliability data. In *Survival Analysis: State of the Art*, 141–152. Eds. J.P. Klein and P. Goel, Kluwer, Amsterdam.

Lawless, J.F. and Nadeau, J.C. (1995). Nonparametric estimation of cumulative mean functions for recurrent events. *Technometrics* **37**, 158–168.

Lawless, J.F. and Thiagarajah, K. (1996). A point process model incorporating renewals and time trends. *Technometrics* **38**, 131–138.

Lawless, J.F. and Yan, P. (1992). Some statistical methods for follow-up studies of disease with intermittent monitoring. In *Multiple Comparisons, Selection, and Applications in Medicine*, Ed. F.M. Hoppe. Marcel Dekker, New York.

Lawless, J.F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canad. J. Statist.* **26**, 549–565.

Lawless, J.F., Hu, J., and Cao, J. (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Anal.* **1**, 227–240.

Lawless, J.F., Kalbfleisch, J.D., and Wild, C.J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *J. Roy. Statist. Soc.* B **61**, 413–438.

Lawless, J.F., Wigg, M.B., Tuli, S., Drake, J., and Lamberti-Pasculli, M. (2001). Analysis of repeated failures or durations, with application to shunt failures for patients with paediatric hydrocephalus. *Applied Statist.* **50**, 449–465.

Lee, L. (1980). Testing adequacy of the Weibull and loglinear rate models for a Poisson process. *Technometrics* **22**, 195–199.

Lee, S.Y. and Tsai, W.Y. (2005). An estimator of the survivor function based on the semi-Markov model under dependent censoring. *Lifetime Data Anal.* **11**, 193–211.

Lee, S.Y. and Wolfe, R.A. (1998). A simple test for independent censoring under the proportional hazards model. *Biometrics* **54**, 1176–1182.

Lewis, P.A.W. (1972). Recent results in the statistical analysis of univariate point processes. In *Stochastic Point Processes*, 1–54. Ed. P.A.W. Lewis. Wiley, New York.

Liang, K-Y. (1987). Estimating functions and approximate conditional likelihood. *Biometrika* **74**, 695–702.

Lin, D.Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Med.* **13**, 2233–2247.

Lin, D.Y. Feuer, E.J., Etzioni, R., and Wax, Y. (1997). Estimating medical costs from incomplete follow-up data. *Biometrics* **53**, 419–434.

Lin, D.Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika* **86**, 59–70.

Lin, D.Y., Wei, L.J., and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605–618.

Lin, D.Y, Wei, L.J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. B* **62**, 711–730.

Lin, H., Scharfstein, D.O., and Rosenheck, R.A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *J. Roy. Statist. Soc. B* **66**, 791–813.

Lindqvist, B.H., Elvebakk, G., and Heggland, K. (2003). The trend-renewal process for statistical analysis of repairable systems. *Technometrics* **45**, 31–44.

Little, R.J.A. (1992). Incomplete data in event history analysis. Ch. 8 in *Demographic Applications of Event History Analysis*. Eds. J. Trussel, R. Hankinson, and J. Tilton. Clarendon Press, Oxford.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Statist. Assoc.* **90**, 1112–1121.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis of Missing Data*, 2nd edition. Wiley, New York.

Liu L., Wolfe, R.A., and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, 747–756.

Maller, R.A., Sun, L., and Zhou, X. (2002). Etimating the expected total number of events in a process. *J. Am. Statist. Assoc.* **97**, 577–589.

Martinussen, T. and Scheike, T.H. (2006). *Dynamic Regression Models for Survival Data*. Springer, New York.

McCullagh, P. and Nelder J. (1989). *Generalized Linear Models*. Chapman and Hall, London.

McGilchrist, C.A. and Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics* **47**, 461–466.

McKeague, I.W. and Sasieni, P.D. (1994). A partly additive risk model. *Biometrika* 81, 501–514.

McLeish, D.L. and Struthers, C. (2006). Estimation of regression parameters in missing data problems. *Canad. J. Statist.* **34**, 233–259.

McMahon, R.P., Proschan, M., Geller, N.L., Stone, P.H., and Sopko, G. (1994). Sample size calculations for clinical trials in which entry criteria and outcomes are counts of events. *Statist. Med.* **13**, 859–870.

Meeker, W.Q. and Escobar, L.E. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

Metcalfe, C. and Thompson, S.G. (2006). The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Statist. Med.* **25**, 165–179.

Miloslavsky, M., Keles, S., van der Laan, M.J., and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *J. Roy. Statist. Soc. B* **66**, 239–257.

Moran, P.A.P. (1971). Maximum likelihood estimation in non-standard conditions. *Proc. Cambridge Phil. Soc.* **70**, 441–450.

Murphy, S.A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–198.

Nadeau, C. and Lawless, J.F. (1998). Inference for means and covariances of point processes through estimating functions. *Biometrika* **85**, 893–906.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829–838.

Naylor, J.C. and Smith, A.F.M. (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statist.* **31**, 214–225.

Nelson, W. (1995). Confidence limits for recurrence data: Applied to cost or number of product repairs. *Technometrics* **37**, 147–157.

Nelson, W.B. (1988). Graphical analysis of system repair data. *J. Qual. Tech.* **20**, 24–35.

Nelson, W.B. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. ASA-SIAM Series on Statistics and Applied Probability #10, Philadelphia.

Ng, E.T.M. and Cook, R.J. (1999a). Adjusted score tests of homogeneity for Poisson processes. *J. Am. Statist. Assoc.* **99**, 308–319.

Ng, E.T.M. and Cook, R.J. (1999b). Robust inference for bivariate point processes. *Canad. J. Statist.* **27**, 509–524.

Nielsen, G.G., Gill, R.D., Andersen, P.K., and Sorensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, 25–43.

Nielsen, J.D. and Dean, C.B. (2005). Regression splines in the quasi-likelihood analysis of recurrent event data. *J. Statist. Plan. Inf.* **134**, 521–535.

Nocedal, J. and Wright, S.J. (1999). *Numerical Optimization*. Springer–Verlag, New York.

Oakes, D. and Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika* **81**, 83–90.

O'Brien, P.C. and Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549–556.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theor.* **IT-27**, 23–31.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Statist. Assoc.* **83**, 9–27.

Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Am. Math. Soc.* **165**, 483–506.

Parner, E. (1998). Asymptotic theory for correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.

Parzen, E. (1999). *Stochastic Processes*. Society for Industrial and Applied Mathematics, Philadelphia.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford.

Pena, E., Strawderman, R., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *J. Am. Statist. Assoc.* **96**, 1299–1315.

Pena, E.A. (1998). Smooth goodness-of-fit tests for composite hypotheses in hazard based models. *Ann. Statist.* **26**, 1935–1971.

Pepe, M.S. (1991). Inference for events with dependent risks in multiple endpoint studies. *J. Am. Statist. Assoc.* **86**, 770–778.

Pepe, M.S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times. *J. Am. Statist. Assoc.* **88**, 811–820.

Pepe, M.S. and Fleming, T.R. (1991). A non-parametric method for dealing with mismeasured covariate data. *J. Am. Statist. Assoc.* **86**, 108–113.

Pepe, M.S., Longton, G., and Thornquist, M. (1991) A qualifier Q for the survival function to describe the prevalence of a transient condition. *Statist. Med.* **10**, 413–421.

Pickles, A. and Crouchley, R. (1994). Generalizations and applications of frailty models for survival and event data. *Statist. Meth. Med. Res.* **3**, 263–278.

Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191–199.

Prentice, R.L., Williams, B.J., and Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.

Press, W.H., Fleming, B.P., Teukolsky, S.A., and Vetterling, W.T. (1986). *Numerical Recipes*. Cambridge University Press, Cambridge, UK.

Præstgaard, J. (1991). Nonparametric estimation of actuarial values. *Scand. Act. J.* **2**, 129–143.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics* **5**, 375–383.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453–466.

Rigdon, S.E. and Basu, A.P. (2000). *Statistical Methods for the Reliability of Repairable Systems*. John Wiley and Sons, New York.

Robins, J.M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proc. Amer. Statist. Assoc. Biopharm. Section*, 24–33.

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues*, 279–331. Eds. N. Jewell, K. Dietz, and V. Farewell. Birkäuser, Boston.

Robins, J.M. and Rotnitzky, A. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**, 805–820.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–866.

Romanowski, B., Marina, R.B., and Roberts, J.N. (2003). Patients' preference of valacyclovir once-daily suppressive therapy versus twice-daily episodic therapy for recurrent genital herpes: A randomized study. *Sexually Transmitted Diseases* **30**, 226–231.

Ross, S.M. (1983). *Stochastic Processes*. John Wiley and Sons, New York.

Satten, G.A. and Sternberg, M.R. (1999). Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics* **53**, 507–513.

Satten, G.A., Datta, S. and Robins, J. (2001). Estimating the marginal survival function in the presence of time dependent covariates. *Statist. Prob. Letters* **54**, 397–403.

Schaubel, D.E. (2005). Variance estimation for clustered recurrent event data with a small number of clusters. *Statist. Med.* **24**, 3037–3051.

Schaubel, D.E. and Cai, J. (2004a). Non-parametric estimation of gap time survival functions for ordered multivariate failure time data. *Statist. Med.* **23**, 1885–1900.

Schaubel, D.E. and Cai, J. (2004b). Regression methods for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika* **91**, 291–303.

Schaubel, D.E. and Cai, J. (2005a). Semiparametric methods for clustered recurrent event data. *Lifetime Data Anal.* **11**, 405–425.

Schaubel, D.E. and Cai, J. (2005b). Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. *Biostatistics* **6**, 404–419.

Schaubel, D.E. and Cai, J. (2006a). Rate/mean regression for multiple-sequence recurrent event data with missing event category. *Scand. J. Statist.* **33**, 191–207.

Schaubel, D.E. and Cai, J. (2006b). Multiple imputation methods for recurrent event data with missing event category. *Canad. J. Statist.* **34**, 677–692.

Scheike, T.H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function of a counting process. *Lifetime Data Anal.* **8**, 247–262.

Scheike, T.H. and Zhang, M.J. (2003). Extensions and applications of the Cox-Aalen survival model. *Biometrics* **59**, 1036–1045.

Segall, A. and Kailath, T. (1975). The modeling of randomly modulated jump processes. *IEEE Trans. Inf. Theory* **21**, 135–143.

Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Singpurwalla, N.D. and Wilson, S.P. (1999). *Statistical Methods in Software Engineering: Reliability and Risk*. Springer, New York.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *J. Am. Statist. Assoc.* **88**, 979–983.

Sinha, D. and Maiti, T. (2004). A Bayesian approach for the analysis of panel-count data with dependent termination. *Biometrics* **60**, 34–40.

Snyder, D.L. and Miller, M.I. (1991). *Random Point Processes in Time and Space*. Springer–Verlag, New York.

Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* **61**, 702–714.

Song, X. and Huang, Y. (2006). A corrected pseudo-score approach for additive hazards model with longitudinal covariates measured with error. *Lifetime Data Anal.* **12**, 97–110.

Staniswalis, J.G., Thall, P.F., and Salch, J. (1997). Semiparametric regression analysis for recurrent event interval counts. *Biometrics* **53**, 1334–1353.

Sternberg, M.R. and Satten, G.A. (1999). Discrete time nonparametric estimation for semi-Markov models of chain-of-events data subject to interval censoring and truncation. *Biometrics* **55**, 514–522.

Strawderman, R. (2000). Estimating the mean of an increasing stochastic process at a censored stopping time. *J. Am. Statist. Assoc.* **95**, 1192–1208.

Strawderman, R. (2005). The accelerated gap times model. *Biometrika* **92**, 647–666.

Stukel, T.A. (1993). Comparison of methods for the analysis of longitudinal interval count data. *Statist. Med.* **12**, 1339–1351.

Sun, J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer, New York.

Sun, J. and Kalbfleisch, J.D. (1993). The analysis of current status data on point processes. *J. Am. Statist. Assoc.* **88**, 1449–1454.

Sun, J. and Kalbfleisch, J.D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica* **5**, 279–290.

Sun, J. and Matthews, D.E. (1997). A random-effect regression model for medical follow-up studies. *Canad. J. Statist.* **25**, 101–111.

Sun, J. and Wei, L.J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Roy. Statist. Soc. B* **62**, 293–302.

Sun, J. and Yang, I. (2000). Nonparametric tests for stratum effects in the Cox model. *Lifetime Data Anal.* **6**, 321–330.

Sun, J., Park, D., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *J. Am. Statist. Assoc.* **100**, 882–889.

Thall, P.F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197–209.

Thall, P.F. and Lachin, J.M. (1988). Analysis of recurrent events: nonparametric methods for random interval count data. *J. Am. Statist. Assoc.* **83**, 339–347.

Thall, P.F. and Vail S.C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46**, 657–671.

Therneau, T.A. and Hamilton, S.A. (1997). rhDNase as an example of recurrent event analysis. *Statist. Med.* **16**, 2029–2047.

Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Therneau, T.M., Grambsch, P.M., and Pankratz, V.S. (2003). Penalized survival models and frailty. *J. Comp. Graph. Statist.* **12**, 156–175.

Trussell. J., Hankinson, R., and Tilton, J. (1992). Eds., *Demographic Applications of Event History Analysis.* Clarendon Press, Oxford.

Tuli, S., Drake, J., Lawless, J.F., Wigg, M., and Lamberti-Pasculli, M. (2000). Risk factors for repeat cerebrospinal shunt failures in pediatric hydrocephalus. *J. Neurosurgery* **92**, 31–38.

Turnbull, B.W., Jiang, W., and Clark, L.C. (1997). Regression models for recurrent event data: Parametric random effects models with measurement error. *Statist. Med.* **16**, 853–864.

van der Laan, M., Hubbard, A., and Robins, J. (2002). Locally efficient estimation of a multivariate survivor function in longitudinal studies. *J. Am. Statist. Assoc.* **98**, 494–507.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.

Vardi, Y. (1982a). Nonparametric estimation in renewal processes. *Ann. Statist.* **10**, 772–785.

Vardi, Y. (1982b). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10**, 616–620.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*, 4th edition. Springer–Verlag, New York.

Verona, E., Petrov, D., Cserhati, E., Hofman, J., Geppe, N., Medley, H., and Hughes, S. (2003). Fluticasone propionate in asthma: A long term dose comparison study. *Arch. Dis. Child* **88**, 503–509.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika* **83**, 507–518.

Vit, P. (1973). Interval prediction for a Poisson process. *Biometrika* **60**, 667–668.

Wang, M.C. (1999). Gap time bias in incident and prevalent cohorts. *Statistica Sinica* **9**, 999–1010.

Wang, M.C. and Chang, S.H. (1999). Nonparametric estimation of a recurrent survival function. *J. Am. Statist. Assoc.* **94**, 146–153.

Wang, M.C., Qin, J., and Chiang, C.T. (2001). Analyzing recurrent event data with informative censoring. *J. Am. Statist. Assoc.* **96**, 1057–1065.

Wang, W. and Wells, M.T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* **85**, 561–572.

Wei, L.J. and Glidden, D.V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statist. Med.* **16**, 833–839.

Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–1073.

Wellner, J.A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.

Wellner, J.A., Zhang, Y., and Liu, H. (2004). A semiparametric regression model for panel count data: When do pseudolikelihood estimators become badly inefficient?, 143–174. *Proc. Second Seattle Symp. in Biostatistics*, Eds. D.Y. Lin and T.R. Fleming. Springer, New York.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–26.

Xue, X. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate failure time. *Lifetime Data Anal.* **2**, 277–289.

Ye, Y., Kalbfleisch, J.D., and Schaubel, D.E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63**, 78–87.

Yi, G.Y. and Lawless, J.F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *J. Statist. Plan. Inf.* **137**, 1816–1828.

Yusuf, S., Wittes, J., Probstfield, J., and Tyroler, H.A. (1991). Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *J. Am. Med. Assoc.* **266**, 93–98.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.

Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* **59**, 1099–1106.

Zhang, Z. and Rockette, H.E. (2005). On maximum likelihood estimation in parametric regression with missing covariates. *J. Statist. Plan. Inf.* **134**, 206–223.

Zhang, Z. and Rockette, H.E. (2006). Semiparametric maximum likelihood for missing covariates in parametric regression. *Ann. Inst. Statist. Math.* **58**, 687–706.

Zhang, Z. and Rockette, H.E. (2007). An EM algorithm for regression analysis with incomplete covariate information. *J. Statist. Comp. Sim.* **77**, 163–173.

Zhao, H. and Tsiatis, A.A. (1997). A consistent estimator for the distribution of quality-adjusted survival time. *Biometrika* **84**, 339–348.

Zhao, H. and Tsiatis, A.A. (1999). Efficient estimation of the distribution of quality adjusted survival time. *Biometrics* **55**, 231–236.

Zhao, Y., Lawless, J.F., and McLeish, D.L. (2006). Maximum likelihood methods for regression problems with missing covariates or responses and two-phase sampling. Manuscript.

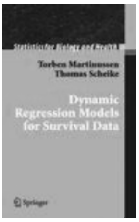
Zucker, D.M. (2005). A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *J. Am. Statist. Assoc.* **100**, 1264–1277.

Author Index

Subject Index

Springer
the language of science

springer.com



Dynamic Regression Models for Survival Data

T. Martinussen and T. Scheike

This book studies and applies modern flexible regression models for survival data with a special focus on extensions of the Cox model and alternative models with the specific aim of describing time-varying effects of explanatory variables. The book covers the use of residuals and resampling techniques to assess the fit of the models and also points out how the suggested models can be utilised for clustered survival data.

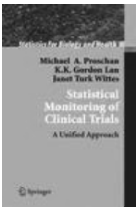
2005. 472 pp. (Statistics for Biology and Health) Hardcover
ISBN 978-0-387-20274-7

The Statistical Analysis of Interval-censored Failure Time Data

Jianguo Sun

This book collects and unifies statistical models and methods that have been proposed for analyzing interval-censored failure time data. It provides the first comprehensive coverage of the topic of interval-censored data and complements the books on right-censored data. The focus of the book is on nonparametric and semiparametric inferences, but it also describes parametric and imputation approaches.

2006. 304 pp. (Statistics for Biology and Health) Hardcover
ISBN 978-0-387-32905-5



Statistical Monitoring of Clinical Trials

Michael A. Proschan, K.K. Gordon Lan, and Janet Turk Wittes

The approach taken in this book is to studies monitored over time, what the Central Limit Theorem is to studies with only one analysis. Just as the Central Limit Theorem shows that test statistics involving very different types of clinical trial outcomes are asymptotically normal, this book shows that the joint distribution of the test statistics at different analysis times is asymptotically multivariate normal with the correlation structure of Brownian motion ("the B-value") irrespective of the test statistic.

2006. 258 pp. (Statistics for Biology and Health) Hardcover
ISBN 978-0-387-30059-7