Nick T. Thomopoulos

Probability Distributions

With Truncated, Log and Bivariate Extensions

Springer

Probability Distributions

Nick T. Thomopoulos

Probability Distributions

With Truncated, Log and Bivariate Extensions

Springer

Nick T. Thomopoulos
Stuart School of Business
Illinois Institute of Technology
Chicago, IL, USA

ISBN 978-3-319-76041-4 ISBN 978-3-319-76042-1 (eBook)
https://doi.org/10.1007/978-3-319-76042-1

Library of Congress Control Number: 2018934426

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

For my wife,
my children,
and my grandchildren

In Memory of Nick T. Thomopoulos

It is with great sadness that we announce the passing of Springer author Nick T. Thomopoulos. Prior to a teaching career at Illinois Tech which spanned over four decades, Nick served in the US Army and rose to the rank of sergeant. Nick's considerable contributions to the field of management science include developing new mathematical methods in the design of cotton-picker cams and agriculture plows, formulating mixed-model assembly methods, conceiving a way to measure airplane noise by the neighborhoods surrounding airports, creating queuing methods to design radio communication systems for the US Navy, and designing forecast and inventory software which is now used by more than 100 companies worldwide. In addition, he authored over 70 peer-reviewed journal articles and 12 books, 7 of which he published with Springer.

Preface

This book includes several statistical methods and tables that are not readily available in other publications. The content begins with a review of continuous and discrete probability distributions that are in common use, and examples are provided to guide the reader on applications. Some other useful distributions that are less common and less understood are described in full detail. These are the discrete normal, left-partial, right-partial, left-truncated normal, right-truncated normal, lognormal, bivariate normal, and bivariate lognormal. Tables with examples are provided to enable researchers to easily apply the distributions to real applications and sample data. Tables are listed for the partial distribution, and examples show how they are applied in industry. Three of the distributions, left-truncated normal, right-truncated normal, and normal, offer a wide variety of shapes, and a new statistic, the spread ratio, enables the analyst to determine which best fits sample data. A set of tables for the standard lognormal distribution is listed, and examples show how they are used to measure probabilities when the sample data is lognormal. A complete set of 21 tables is listed for the bivariate normal distribution, and examples show how they apply to sample data. Also shown is how to calculate probabilities when an application is from a bivariate lognormal distribution. This book will be highly useful to anyone who does statistical and probability analysis. This includes scientists, economists, management scientists, market researchers, engineers, mathematicians, and students in many disciplines.

Burr Ridge, IL, USA — Nick T. Thomopoulos

Acknowledgments

Thanks especially to my wife, Elaine Thomopoulos, who encouraged me to write this book, and who gave consultation whenever needed. Lauren Sussman assisted in editing the text, while Daniel Sussman provided guidance and advice on creating the graphs and edited the book. Thanks also to the many people who have helped and inspired me over the years, including some former Illinois Institute of Technology (Illinois Tech) Ph.D. students. I can name only a few here: Emanuel Betinis, National University of Health Science; Dick Chiapetta, Chiapetta and Welch; John Garofalakis, Patras University; Marsha Jance, University of Indiana – Richmond; Montira Jantaravareerat, Illinois Tech; Arvid Johnson, St. Francis University; Carol Lindee, Panduit; Anatol Longinow, Illinois Tech; Fotis Mouzakis, Frynon Research; George Resnikoff, California State University; Paul Spirakis, Patras University; Reino Warren, University of Michigan – Flint; and Tolis Xanthopoulos, Mercer Investment Consulting.

Contents

About the Author

Nick T. Thomopoulos has degrees in business (B.S.) and in mathematics (M.A.) from the University of Illinois, and in industrial engineering (Ph.D.) from Illinois Institute of Technology (Illinois Tech). He was supervisor of operations research at International Harvester, senior scientist at the IIT Research Institute, and professor in Industrial Engineering and in the Stuart School of Business at Illinois Tech. He is the author of 11 books including *Fundamentals of Queuing Systems* (Springer), *Essentials of Monte Carlo Simulation* (Springer), *Applied Forecasting Methods* (Prentice Hall), and *Fundamentals of Production, Inventory and the Supply Chain* (Atlantic). He has published many papers and has consulted in a wide variety of industries in the United States, Europe, and Asia. Dr. Thomopoulos has received honors over the years, such as the Rist Prize from the Military Operations Research Society for new developments in queuing theory; the Distinguished Professor Award in Bangkok, Thailand, from the Illinois Tech Asian Alumni Association; and the Professional Achievement Award from the Illinois Tech Alumni Association.

Chapter 1
Continuous Distributions

1.1 Introduction

A variable, x, is continuous when x can be any number between two limits. For example, a scale measures a boy at 150 pounds; and assuming the scale is correct within one-half pound, the boy's actual weight is a continuous variable that could fall anywhere from 149.5 to 150.5 pounds. The variable, x, is a continuous random variable when a mathematical function, called the probability density defines the shape along the admissible range. The density is always zero or larger and the positive area below the density equals one. Each unique continuous random variable is defined by a probability density that flows over the admissible range. Eight of the common continuous distributions are described in the chapter. For each of these, the range of the variable is stated, along with the probability density, and the associated parameters. Also described is the cumulative probability distribution that is needed by an analyst to measure the probability of the x falling in a sub-range of the admissible region. Some of the distributions do not have closed-form solutions, and thereby, quantitative methods are needed to measure the cumulative probability. Sample data is used to estimate the parameter values. Examples are included to demonstrate the features and use of each distribution. The distributions described in this chapter are the following: continuous uniform, exponential, Erlang, gamma, beta, Weibull, normal and lognormal. The continuous uniform occurs when all values between limits a to b are equally likely. The normal density is symmetrical and bell shaped. The exponential happens when the most likely value is at $x = 0$, and the density tails down in a relative way as x increases. The density of the Erlang has many shapes that range between the exponential and the normal. The shape of the gamma density varies from exponential-like to one where the mode (most likely) and the density skews to the right. The beta has many shapes: uniform, ramp down, ramp up, bathtub-like, normal-like, and all shapes that skew to the right and in the same manner they skew to the left. The Weibull density varies from exponential-like to

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_1

shapes that skew to the right. The lognormal density peaks near zero and skews far to the right.

Law and Kelton [1]; Hasting and Peacock [2]; and Hines et al. [3] present thorough descriptions on the properties of the common continuous probability distributions.

1.2 Sample Data Statistics

When n sample data, $(x_1, \ldots, x_n)$, are collected, various statistical measures can be computed as described below:

$$\begin{aligned}
&x(1) = \text{minimum of } (x_1, \ldots, x_n)\\
&x(n) = \text{maximum of } (x_1, \ldots, x_n)\\
&\bar{x} = \text{average}\\
&s = \text{standard deviation}\\
&cov = s/\bar{x} = \text{coefficient of variation}\\
&\tau = s^2/\bar{x} = \text{lexis ratio}
\end{aligned}$$

1.3 Notation

The statistical notation used in this book is the following:

$$\begin{aligned}
&E(x) = \text{expected value of } x\\
&V(x) = \text{variance of } x\\
&\mu = \text{mean}\\
&\sigma^2 = \text{variance}\\
&\sigma = \text{standard deviation}
\end{aligned}$$

Example 1.1 Suppose an experiment yields n = 10 sample data values as follows: [24, 27, 19, 14, 32, 28, 35, 29, 25, 33]. The statistical measures from this data are listed below.

$$\begin{aligned}
&x(1) = \min = 14\\
&x(10) = \max = 35\\
&\bar{x} = 26.6\\
&s = 6.44\\
&cov = 0.24\\
&\tau = 1.56
\end{aligned}$$

1.4 Parameter Estimating Methods

Two popular methods have been developed to estimate the parameters of a distribution from sample data. One is called the maximum-likelihood-estimate method (MLE) that is mathematically formulated to find the parameter estimate that gives the most likely fit with the sample data. The other method is called the method-of-moments (MoM) that substitutes the statistical measures like ($\bar{x}$, s) into their mathematical counterparts [μ, σ] and applies algebra to find the estimates of the parameters.

1.5 Transforming Variables

While analyzing sample data, it is sometimes useful to convert a variable x to another variable, x′, where x′ ranges from zero to one; or where x′ is zero or larger. More discussion is below.

Transform Data to (0,1)

A way to convert a variable from x to x` so that x′ lies between 0 and 1 is described here. Recall the summary statistics of the variable x as listed earlier. For convenience in notation, let a′ = x(1) for the minimum, and b′ = x(n) for the maximum. When x, with average $\bar{x}$ and standard deviation s, is converted to x′ by the relation:

$$x' = (x - a')/(b' - a')$$

the range on x` becomes (0,1). The converted sample average and standard deviation are listed below:

$$\bar{x}' = \left(\bar{x} - a'\right)/(b' - a')$$
$$s' = s/(b' - a')$$

respectively, and the coefficient of variation of x` is:

$$\text{cov}' = s`/\bar{x}'$$

When x lies in the (0,1) range, the cov is sometimes useful to identify the distribution that best fits sample data.

Transform Data to ($x \geq 0$)

A way to convert a variable, x to, x`, where $x' \geq 0$ is given here. The summary statistics described earlier of the variable x are used again with $a' = x(1)$ for the minimum. When x is converted to x' by the relation:

$$x' = (x - a')$$

the range of x' becomes zero and larger. The corresponding sample average and standard deviation become:

$$\bar{x}' = (\bar{x} - a')$$
$$s' = s$$

respectively. Finally, the coefficient of variation is:

$$cov' = s'/\bar{x}'$$

1.6 Continuous Random Variables

A continuous random variable, x, can take on any value in a range that spans between limits a and b. Note where the low limit, a, could be minus infinity; and the high limit, b, could be plus infinity. An example is the amount of rainwater found is a five-gallon bucket after a rainfall. A probability density function, f(x), defines how the probability varies along the range, where the sum of the area within the admissible region sums to one. Below defines the probability density function, f(x), and the cumulative distribution function, F(x):

$$f(x) \geq 0 \qquad a \leq x \leq b$$
$$F(x) = \int_a^x f(w)dw \qquad a \leq x \leq b$$

This chapter describes some of the common continuous probability distributions and their properties. The random variable of each is denoted as x, and below is a list of the distributions with their designations and parameters.

continuous uniform	$x \sim CU(a,b)$
exponential	$x \sim Exp(\theta)$
Erlang	$x \sim Erl(k,\theta)$
gamma	$x \sim Gam(k_1,k_2,a,b)$
beta	$x \sim Bet(k_1,k_2,a,b)$
Weibull	$x \sim We(k_1,k_2)$
normal	$x \sim N(\mu,\sigma^2)$
lognormal	$x \sim LN(\mu_y,\sigma_y^2)$

Of particular interest with each distribution is the coefficient of variation (cov) and its range of values that apply. When sample data is available, the sample cov can be measured and compared to each distribution's cov range to help narrow the choice of the distribution that applies.

1.7 Continuous Uniform

A variable, x, follows a continuous uniform probability distribution, Cu(a,b), when it has two parameters a and b, where x can fall equally likely anywhere from a to b. See Fig. 1.1. The probability density, and the cumulative distribution function of x are below:

$$f(x) = 1/(b - a) \qquad a \le x \le b$$
$$F(x) = (x - a)/(b - a) \qquad a \le x \le b$$

The expected value, variance, and standard deviation of x are listed below:

$$E(x) = \mu = (b + a)/2$$
$$V(x) = \sigma^2 = (b - a)^2/12$$
$$\sigma = (b - a)/\sqrt{12}$$

Coefficient of Variation

Note when the low limit is set to zero, (a = 0):

$$\mu = b/2$$
$$\sigma = b/\sqrt{12}$$
$$cov = 2/\sqrt{12} = 0.577$$

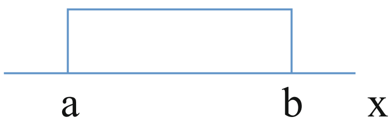

Fig. 1.1 The continuous uniform distribution

Parameter Estimates

When sample data, (x_1, …, x_n), is available, the parameters (a,b) are estimated as shown below by either the maximum-likelihood estimate (MLE) method, or by the method-of-moment estimate.

From MLE, the estimates of the two parameters are the following:

$$\hat{a} = x(1) = \min(x_1, \ldots, x_n)$$
$$\hat{b} = x(n) = \max(x_1, \ldots, x_n)$$

The method-of-moment way uses the two equations: $\mu = (b + a)/2$, and $\sigma = (b - a)/\sqrt{12}$, to estimate the parameters (a,b). as below:

$$\hat{a} = \bar{x} - \sqrt{12}s/2$$
$$\hat{b} = \bar{x} + \sqrt{12}s/2$$

Recall, $\bar{x}$ is the sample average and s is the sample standard deviation.

Example 1.2 Suppose a continuous uniform variable x has min = 0 and max = 1, yielding: x ~ CU(0,1). Some statistics are below:

$$f(x) = 1 \quad 0 \le x \le 1$$
$$F(x) = x \quad 0 \le x \le 1$$
$$\mu = 0.5$$
$$\sigma^2 = 1/12 = 0.083$$
$$\sigma = \sqrt{1/12} = 0.289$$
$$\text{cov} = 0.289/0.500 = 0.578$$

The probability of x less or equal to 0.45, say, is:

$$P(x \le 0.45) = F(0.45) = 0.45.$$

Example 1.3 The yield strength on a copper tube was measured at 70.23 from a device with accuracy of ± 0.40, evenly distributed. Hence, the true yield strength, denoted as x, follows a continuous uniform distribution with parameters:

$$a = 70.23 - 0.40 = 69.83$$
$$b = 70.23 + 0.40 = 70.63$$

The probability density becomes:

$$f(x) = 1/0.80 \quad 69.83 \le x \le 70.63$$

and the cumulative distribution is:

$$F(x) = (x - 69.83)/0.80 \quad 69.83 \le x \le 70.63$$

The mean, variance and standard deviation are the following:

$$\mu = 70.23$$
$$\sigma^2 = (0.80)^2/12 = 0.053$$
$$\sigma = \sqrt{0.53} = 0.231$$

The probability that the true yield strength is below 70, say, becomes:

$$F(70.00) = (70.00 - 69.83)/0.80 = 0.212$$

Note, the cov is 0.231/70.23 = 0.003

But when x is converted to x` = x – a, the mean, standard deviation, and coefficient of variation become:

$$\mu` = (70.23 - 69.83) = 0.40$$
$$\sigma` = 0.231$$
$$cov` = 0.231/0.40 = 0.577$$

Example 1.4 An experiment yields the following ten sample data entries: (12.7, 11.4, 15.3, 20.5, 13.6, 17.4, 15.6, 14.9, 19.7, 18.3). The analyst assumes the data comes from a continuous uniform distribution and seeks to estimate the parameters, (a, b). To accomplish, the following statistics are measured:

$$x(1) = \min = 11.4$$
$$x(n) = \max = 20.5$$
$$\bar{x} = 15.93$$
$$s = 3.00$$

The two methods of estimating the parameters (a,b) are applied. The MLE estimates are the following:

$$\hat{a} = \min = 11.4$$
$$\hat{b} = \max = 20.5$$

The method-of-moment estimates become:

$$\hat{a} = 15.93 - \sqrt{12} \times 3.00/2 = 10.73$$
$$\hat{b} = 15.93 + \sqrt{12} \times 3.00/2 = 21.13$$

Note, when x` = (x – a):

$$\bar{x}` = (15.93 - 11.4) = 4.53$$
$$s` = s = 3.00$$

and

$$cov = s/\bar{x}` = 3.00/4.53 = 0.662$$

which is reasonably close to the continuous uniform value of 0.577.

1.8 Exponential

The exponential distribution, Ex(θ), is used in many areas of science and is the primary distribution that applies in queuing theory to represent the time between arrivals and the time to service a unit. The variable, x, has its peak at x = 0 and a density that continually decreases as x increases. See Fig. 1.2 where $\theta = 1$. The density has one parameter, θ, and is defined as below:

$$f(x) = \theta e^{-\theta x} \quad \text{for}\, x \geq 0$$

The cumulative probability distribution becomes,

$$F(x) = 1 - e^{-\theta x} \quad \text{for} \;\; x \geq 0$$

The mean, variance, and standard deviation of x are listed below:

$$\mu = 1/\theta$$
$$\sigma^2 = 1/\theta^2$$
$$\sigma = 1/\theta$$

Since $\mu = \sigma$, the coefficient-of-variation becomes:

$$cov = 1.00$$

The median, $x_{0.50}$, occurs when F(x) = 0.50; and thereby,

$$F(x_{0.50}) = 0.50 = 1 - e^{-\theta x_{0.50}}$$

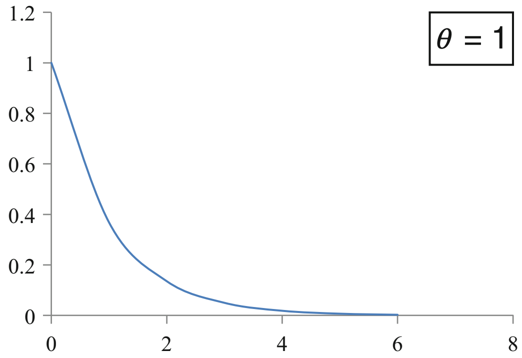

Fig. 1.2 The Exponential Distribution when $\mu = 1.0$

Solving for $x_{0.50}$, yields:

$$x_{0.50} = -\ln(1 - 0.50)/\theta = 0.693/\theta = 0.693\mu$$

where ln = the natural logarithm.

Parameter Estimate

When a sample of size n yields sample data (x_1, . . ., x_n) and an average, $\bar{x}$, the estimate of θ becomes:

$$\widehat{\theta} = 1/\widehat{x}$$

This estimate is derived from the MLE and also from the method-of-moments.

Example 1.5 Assume a variable, x, is exponentially distributed with mean equal to ten. Some measures of x are below:

$$\mu = 1/\theta = 10$$
$$\theta = 0.10$$
$$f(x) = 0.10e^{-0.1x} \quad x > 0$$
$$F(x) = 1 - e^{-0.1x} \quad x > 0$$
$$\sigma^2 = 100$$
$$\sigma = 10$$
$$cov = \sigma/\mu = 10/10 = 1.00$$

Example 1.6 The α-percent-point of x, denoted as xα, is computed as below.

$$\text{Since, } \alpha = F(x) = 1 - e^{-\theta x}$$
$$x\alpha = -\left(\frac{1}{\theta}\right)\ln(1 - \alpha)$$

where ln = natural logarithm, and,

$$P(x \le x\alpha) = \alpha$$

When the exponential variable, x`, has $\theta = 1.0$, giving μ = 1.0, the distribution is sometimes called the standard exponential distribution; and the α-percent-point of x` becomes:

$$x\grave{}\alpha = -\ln(1 - \alpha)$$

When $\theta \neq 1.0$, the α-percent-point of x, xα, is related to its counterpart of x`, x`α, as follows:

$$x\alpha = -\left(\frac{1}{\theta}\right)\ln(1-\alpha) = (1/\theta)x`\alpha = \mu x`\alpha$$

With $\theta = 1.0$ and $\alpha = 0.90$, say,

$$x`_{0.90} = -\ln(1-0.90) = 2.303.$$

$$\text{If } \theta = 2.0 : \mu = 1/\theta = 0.5 \text{ and } x_{0.90} = 0.5 \times 2.303 = 1.152$$

$$\text{If } \theta = 0.4 : \mu = 2.5 \text{ and } x_{0.90} = 2.5 \times 2.303 = 5.757$$

so forth,

Example 1.7 Suppose an analyst has n = 10 sample observations: 6.0, 11.4, 21.6, 0.6, 3.0, 5.0, 9.0, 14.6, 2.6, 4.2, and suspects the data is from an exponential distribution. The sample average and standard deviation are computed and as follows:

$$\bar{x} = 7.80$$

$$s = 6.48$$

The estimate of the exponential parameter is $\widehat{\theta} = 1/\bar{x} = 1/7.80 = 0.128$. Note also, the coefficient-of variation becomes:

$$cov = 6.48/7.80 = 0.83$$

which is similar to cov = 1.00 of the exponential distribution.

1.9 Erlang

The Erlang distribution, Erl(k,θ), is named after its founder, Agner Erlang, a Danish mathematician from the early 1900s, who was studying how many circuits were necessary to provide acceptable telephone service while working for the Copenhagen Telephone Company. The distribution has many shapes ranging from the exponential to normal, and is often used in studying queuing systems. The Erlang has two parameters, θ and k. The parameter θ is the scale parameter and k, is an integer that identifies the number of independent exponential variables that are summed to form the Erlang variable.

The Erlang variable x is related to the exponential variable y as below:

$$x = (y_1 + \ldots + y_k)$$

The probability density of x is,

$$f(x) = x^{k-1}\theta^k e^{-\theta x}/(k-1)! \quad x > 0$$

and the cumulative distribution function is,

$$F(x) = 1 - e^{-\theta x}\sum_{j=0}^{k-1}(\theta x)^j/j! \quad x > 0$$

The expected value of x is related to the expected value of y as below:

$$E(x) = \mu = kE(y) = k/\theta$$

In the same way, the variance of x is formed from adding k variances of y, V(y), as below:

$$V(x) = \sigma^2 = kV(y) = k/\theta^2$$

The standard deviation becomes:

$$\sigma = \sqrt{k}/\theta$$

When k = 1, the Erlang distribution is the same as an exponential distribution where the peak is at x = 0, from where the density skews downward to the right. Note via the central-limit theorem, as k increases, the shape of the Erlang density resembles a normal density.

Coefficient-of-Variation

The coefficient of variation of the Erlang is computed as below:

$$\begin{aligned} cov &= \sigma/\mu \\ &= \sqrt{k}/\theta/k/\theta \\ &= 1/\sqrt{k} \end{aligned}$$

The following is a list of the cov as k ranges from 1 to 10. Note how the distribution begins as an exponential (cov = 1.00), and trends as a normal variable (cov $\leq$ 0.33).

k	cov
1	1.000
2	0.707
3	0.577
4	0.500
5	0.447
6	0.408

7	0.378
8	0.354
9	0.333
10	0.316

Parameter Estimates

When sample data ($x_1, \ldots, x_n$) is available, with the sample average, $\bar{x}$, and variance, s^2, parameter estimates, ($\widehat{k}, \widehat{\theta}$) can be obtained in a two step manner. Using the relations for μ and σ^2 listed earlier, the relations below now apply:

$$\bar{x} = \widehat{k}/\widehat{\theta}$$

$$s^2 = \widehat{k}/\widehat{\theta}^2$$

1. Using the above,

$$\widehat{\theta} = \bar{x}/s^2$$

$$\widehat{k} = \bar{x}\widehat{\theta}$$

2. Set $\widehat{k}$ equal to its closest integer, and re-compute $\widehat{\theta}$ as:

$$\widehat{\theta} = \widehat{k}/\bar{x}$$

Example 1.8 Suppose n samples ($x_1, \ldots, x_n$) yields: $\bar{x} = 20$ and $s^2 = 64$. Using the two steps given above,

$$(1) \quad \widehat{\theta} = \frac{20}{64} = 0.3125$$

$$\widehat{k} = 20 \times 0.3125 = 6.25$$

$$(2) \quad \widehat{k} = 6$$

$$\widehat{\theta} = 6/20 = 0.30$$

Example 1.9 Suppose the same data of Example 1.8 where n samples ($x_1, \ldots, x_n$) yield $\bar{x} = 20$ and s = 8, and cov = 8/20 = 0.40. Using the cov listings given earlier, the closest k to cov = 0.40 is $\widehat{k} = 6$; and thereby, the estimate of θ becomes $\widehat{\theta} = \widehat{k}/\bar{x} =$ 6/20 = 0.30.

Example 1.10 Consider an Erlang variable with parameters, k = 3 and θ = 0.5. The mean, variance, standard deviation and coefficient-of-variation are below:

$$\mu = 3/0.5 = 6.0$$
$$\sigma^2 = 3/0.5^2 = 12.00$$
$$\sigma = \sqrt{12} = 3.46$$
$$\text{cov} = 3.46/6.0 = 0.577$$

The cumulative distribution function is obtained as follows:

$$F(x) = 1 - e^{-0.5x}\left[(0.5x)^0/0! + (0.5x)^1/1! + (0.5x)^2/2!\right]$$

The probability of x = 8.0 or less, say, is computed as below:

$$F(8.0) = 1 - e^{-4}[1 + 4 + 8] = 0.761$$

Cumulative Probability

A standard Erlang distribution with parameter k is introduced by setting the parameter θ to one, $\theta = 1.0$. By doing so, the cumulative distribution of the variable x is computed as follows:

$$F(x) = 1 - e^{-x}\sum_{j=0}^{k-1} x^j/j! \quad x > 0$$

where F(x) = probability of x or less. Calculated values of F(x) are listed in Table 1.1 below for selective entries of x when the Erlang parameter varies from k = 1 to k = 6. Note at k = 1, the Erlang is the same as the exponential distribution.

For clarity, let $F(x`)_\theta$ represent the cumulative distribution for variable x` and for any θ; and let, $F(x)_1$ be the same for variable x when $\theta = 1$. Hence, x`$\theta$ = x, and x` = x/θ. For example, when k = 1 and $\theta = 1$, Table 1.1 shows $F(2)_1 = 0.86$. So when k = 1, $\theta = 2$, x` = x/ $\theta$ = 2/2 = 1.0 and $F(1)_2 = F(2)_1 = 0.86$. Also note when k = 1 and $\theta = 0.5$, x` = x/θ = 2/0.5 = 4 and thereby $F(4)_{0.5} = F(2)_1 = 0.86$, So forth.

1.10 Gamma

The gamma distribution, Gam(k,θ), is the same as the Erlang, except the parameter k is a positive integer for the Erlang, and k is any value larger than zero for the gamma. The gamma variable, x, is zero or larger. The density of the gamma is,

Table 1.1 Values of the cumulative probability, F(x), for selective values of x from the Erlang Distribution for k = 1 to 6 when θ = 1.0

	k					
x	1	2	3	4	5	6
0.0	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.39	0.09	0.01	0.00	0.00	0.00
1.0	0.63	0.26	0.08	0.02	0.00	0.00
1.5	0.78	0.44	0.19	0.07	0.02	0.00
2.0	0.86	0.59	0.32	0.14	0.05	0.02
2.5	0.92	0.71	0.46	0.24	0.11	0.04
3.0	0.95	0.80	0.58	0.35	0.18	0.08
3.5	0.97	0.86	0.68	0.46	0.27	0.14
4.0	0.98	0.91	0.76	0.57	0.37	0.21
4.5	0.99	0.94	0.83	0.66	0.47	0.30
5.0	0.99	0.96	0.88	0.73	0.56	0.38
5.5	1.00	0.97	0.91	0.80	0.64	0.47
6.0	1.00	0.98	0.94	0.85	0.71	0.55
6.5	1.00	0.99	0.96	0.89	0.78	0.63
7.0	1.00	0.99	0.97	0.92	0.83	0.70
7.5	1.00	1.00	0.98	0.94	0.87	0.76
8.0	1.00	1.00	0.99	0.96	0.90	0.81
8.5	1.00	1.00	0.99	0.97	0.93	0.85
9.0	1.00	1.00	0.99	0.98	0.95	0.88
9.5	1.00	1.00	1.00	0.99	0.96	0.91
10.0	1.00	1.00	1.00	0.99	0.97	0.93
10.5	1.00	1.00	1.00	0.99	0.98	0.95
11.0	1.00	1.00	1.00	1.00	0.98	0.96
11.5	1.00	1.00	1.00	1.00	0.99	0.97
12.0	1.00	1.00	1.00	1.00	0.99	0.98
12.5	1.00	1.00	1.00	1.00	0.99	0.99
13.0	1.00	1.00	1.00	1.00	1.00	0.99
13.5	1.00	1.00	1.00	1.00	1.00	0.99
14.0	1.00	1.00	1.00	1.00	1.00	0.99
14.5	1.00	1.00	1.00	1.00	1.00	1.00
15.0	1.00	1.00	1.00	1.00	1.00	1.00

$$f(x) = x^{k-1}\theta^k \exp(-\theta x)/\Gamma(k) \quad x > 0$$

where Γ(k) is called the gamma function, (not a density), defined as

$$\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt \quad \text{for}\, k > 0$$

When k is a positive integer, Γ(k) = (k-1)! Note: Γ(1) = 1, Γ(2) = 1, Γ(3) = 2, Γ(4) = 6, Γ(5) = 24, and so forth. Also note, $\Gamma(0.5) = \sqrt{\pi} = 1.77$. The non-integer values, in-between, are estimated using quantitative methods.

F(x) is computed as an Erlang when k is a positive integer, but when k is not an integer, quantitative methods are needed.

The mean, variance, and standard deviation of x are the following:

$$\mu = k/\theta$$

$$\sigma^2 = k/\theta^2$$

$$\sigma = \sqrt{k}/\theta$$

Hence,

$$\text{cov} = \sigma/\mu = 1/\sqrt{k}$$

When $k \le 1$: cov ≥ 1, and the mode is zero.

When $k > 1$: cov < 1, and the mode is larger than zero.

Since the shape of the gamma looks anywhere from an exponential to a normal, the cov ranges from above one to below 0.33.

Parameter Estimates

The parameters, k, θ, can be estimated from sample data via the method-of-moments as shown below:

Since, $\mu = k/\theta$ and $\sigma^2 = k/\theta^2$; substituting the sample counterparts, $\bar{x}$, s^2, yields: $\bar{x} = \widehat{k}/\widehat{\theta}$ and $s^2 = \widehat{k}/\widehat{\theta}^2$. Thereby, the estimates of θ and k are the following:

$$\widehat{\theta} = \bar{x}/s^2$$

$$\widehat{k} = \bar{x}\widehat{\theta}$$

Example 1.11 An experiment yields the following sample data: $(x_1, \ldots, x_n)$ with mean $\bar{x} = 20.4$ and variance $s^2 = 95.6$. Assuming the gamma distribution applies, the estimates of the gamma parameters become:

$$\widehat{\theta} = \bar{x}/s^2 = 20.4/95.5 = 0.214$$
$$\widehat{k} = \bar{x}\widehat{\theta} = 20.4 \times 0.214 = 4.37$$

Cumulative Probability Estimates

When the gamma parameter θ is set to one, θ = 1, the cumulative probability of the gamma variable, x, can be estimated from use of Table 1.1 results that are based on the Erlang distribution. Recall the parameter k for the Erlang includes only the positive integers, k = 1, 2, ...; while the same for the gamma is any value of k larger than zero. The estimates are based on interpolation and include only the values of k > 1. The method is described below in Example 1.12.

Example 1.12 Suppose x is a gamma variable with θ = 1, k = 1.7 and the analyst wants an estimate of $F(2)_{k=1.7} = P(x < 2.0)$. Note the particular values from Table 1.1 listed below:

k	F(2)
1	0.86
2	0.59

Hence,

$$0.59 < F(2)_{k=1.7} < 0.86$$

and with interpolation:

$$F(2)_{k=1.7} \approx 0.67.$$

Assume further, the analyst wants an estimate of $F(1.8)_{k=1.7} = P(x < 1.8)$ when θ = 1 and k = 1.7. The pertinent entries from Table 1.1 are listed below:

k	F(1.5)	F(2.0)
1	0.78	0.86
2	0.44	0.59

Hence,

$$0.44 < F(1.8)_{k=1.7} < 0.86$$

and with interpolation,

$$F(1.8)_{k=1.7} \approx 0.62.$$

1.11 Beta

The beta distribution, Bet(k_1,k_2,a,b), with variable x, has four parameters (k_1,k_2,a,b) where $k_1 > 0$, $k_2 > 0$, and ($a < x < b$). The distribution has many shapes depending on the values of k_1 and k_2 as described below:

Parameters	Shape
$k_1 < 1$ and $k_2 \geq 1$	mode at x = a (right skewed)
$k_1 \geq 1$ and $k_2 < 1$	mode at x = b (left skewed)
$k_1 = 1$ and $k_2 > 1$	ramp down from x = a to x = b
$k_1 > 1$ and $k_2 = 1$	ramp down from x = b to x = a

$k_1 < 1$ and $k_2 < 1$	bathtub shape
$k_1 > 1$ and $k_2 > 1$ and $k_1 > k_2$	mode closer to x = b and left skewed
$k_1 > 1$ and $k_2 > 1$ and $k_2 > k_1$	mode closer to x = a and right skewed
$k_1 > 1$ and $k_2 > 1$ & $k_1 = k_2$	mode in middle, symmetrical, normal like
$k_1 = k_2 = 1$	uniform

Standard Beta

A related distribution is the standard beta, Bet(k_1,k_2,0,1),with variable x`, having the same parameters ($k_1$,$k_2$) as the beta, and is constrained to the range (0,1). The relation between x and x` is below:

$$x` = (x - a)/(b - a)$$

and,

$$x = a + x`(b - a)$$

The probability density for x` is,

$$f(x`) = (x`)^{k1-1}(1 - x`)^{k2-1}/B(k_1, k_2) \quad (0 < x` < 1)$$

where

$$B(c, d) = \text{beta function} = \int_0^1 t^{c-1}(1 - t)^{d-1}dt$$

There is no closed-form solution for the cumulative distribution function, F(x`).
The expected value and variance of x` are below:

$$E(x`) = \mu = \frac{k_1}{k_1 + k_2}$$

$$V(x`) = \sigma^2 = \frac{(k_1k_2)}{(k_1 + k_2)^2(k_1 + k_2 + 1)}$$

When $k_1 > 0$ and $k_2 > 0$, the mode is the following:

$$\widetilde{\mu} = \frac{k_1 - 1}{(k_1 + k_2 - 2)}$$

Mean and Variance

The expected value and variance of the beta x becomes:

$$E(x) = a + E(x`)(b - a)$$
$$V(x) = (b - a)^2 V(x`)$$

Parameter Estimates

When an analyst has sample data, $(x_1, \ldots, x_n)$, and obtains measures of the sample mean, $\bar{x}$`, and mode, $\tilde{x}$`, estimates of the parameters k_1 and k_2 can be derived. This is by recalling the equations given above on the mean, μ, and the mode, $\widetilde{\mu}$, of the standard beta, and applying some algebra to arrive at the following estimates:

$$\widehat{k}_1 = \bar{x}` \left[2\tilde{x}` - 1\right] / \left[\tilde{x}` - \bar{x}`\right]$$
$$\widehat{k}_2 = \left[1 - \bar{x}`\right] \widehat{k}_1 / \bar{x}`$$

Example 1.13 Assume an analyst has sample data from a beta distribution with measures of the sample mean, $\bar{x} = 22$, mode, $\tilde{x} = 18$, a` = 10 and b` = 50. The data are converted to standard beta measures as below:

$$\bar{x}` = (22 - 10)/(50 - 10) = 0.30$$
$$\tilde{x}` = (18 - 10)/(50 - 10) = 0.20$$

So now, the estimates of the parameters, k_1, k_2, are measured:

$$\widehat{k}_1 = 0.30(2 \times 0.2 - 1.0)/(0.20 - 0.30) = 1.80$$
$$\widehat{k}_2 = (1 - 0.30) \times 1.8/0.30 = 4.20$$

Note, since $\widehat{k}_1 > 1$, $\widehat{k}_2 > 1$ and $\widehat{k}_2 > \widehat{k}_1$, the beta variable, x, lies within a = 10 and b = 50 and is skewed to the right.

1.12 Weibull

The Weibull distribution, $We(k_1,k_2)$, is named after Wallodi Weibull, a Swedish mathematician who described its use during 1951 with applications in reliability and life testing. The distribution has two parameters, $k_1 > 0$ and $k_2 > 0$, and the random variable, denoted as x, ranges from zero and above. The probability density is:

$$f(x) = k_1 k_2^{-k1} x^{k1-1} \exp\left[-(x/k_2)^{k1}\right] \quad x > 0$$

and the cumulative distribution function is:

$$F(x) = 1 - \exp\left[-(x/k_2)^{k1}\right] \quad x > 0$$

The mean and variance of x are listed below,

$$\mu = \frac{k_2}{k_1}\Gamma\left(\frac{1}{k_1}\right)$$

$$\sigma^2 = \frac{\left(k_2{}^2\right)}{k_1}\left[2\Gamma\left(\frac{2}{k_1}\right) - 1/k_1\Gamma\left(\frac{1}{k_1}\right)^2\right]$$

Recall Γ denotes the gamma function described earlier. When the parameter $k_1 \leq 1$, the shape of the density is exponential-like with a peak at $x = 0$. When $k_1 > 1$, the shape has a mode greater than zero and skews to the right, and at $k_1 \geq 3$, the density shape starts looking like a normal distribution.

Since the shape of the Weibull goes from an exponential-like to a normal-like, the cov ranges from above one to below 0.33.

The mode is computed as below:

$$\widetilde{\mu} = k_2[(k_1 - 1)/k_1]^{1/k1}$$

Note when $k_1 = 1$, the mode is zero.

Weibull Plot

The analyst can plot sample data on special designed graph paper to determine if the data fits a Weibull distribution, and if so, the plot gives estimates on the parameters: k_1 and k_2.

Parameter Estimates

Law and Kelton [1] provide the following way to estimate the Weibull parameters (k_1, k_2) when the following data is available:

$\tilde{x}$ = an estimate of the most likely value (mode) of x

x_α = an estimate of the α – percent – point value of x

The analysis here is when $k_1 \geq 1$ and the mode of x is greater than zero. When x represents time-to-fail, $k_1 = 1$ indicates the fail rate is constant, and $k_1 > 1$ is when the fail rate increases over time. For this situation, the mode is measured as below:

$$\tilde{x} = k_2[(k_1 - 1)/k_1]^{1/k1}$$

Using algebra, k_2 becomes

$$k_2 = \tilde{x}/[(k_1 - 1)/k_1]^{1/k_1}$$

The cumulative distribution for the α-percent-point of x is x_α and is obtained by the following,

$$F(x_\alpha) = 1 - \exp\left[-(x_\alpha/k_2)^{k1}\right] = \alpha$$

Hence,

$$\ln(1 - \alpha) = \left[-(x_\alpha/k_2)^{k1}\right]$$

Applying algebra and solving for k_2,

$$k_2 = (x_\alpha)/\ln[1/(1 - \alpha)]^{1/k1}$$

So now,

$$\tilde{x}/[(k_1 - 1)/k_1]^{1/k_1} = (x_\alpha)/\ln[1/(1 - \alpha)]^{1/k_1}$$

and thereby,

$$\tilde{x}/x_\alpha = [(k_1 - 1)/k_1]^{1/k1}/\ln(1/1 - \alpha)^{1/k1}$$

Solving for k_1 The only unknown in the above equation is k_1. At this point, an iterative search is made to find the value of k_1 where the right side of the above equation is equal to the left side. The result is $\widehat{k}_1$.

Solving for k_2 Having found $\widehat{k}_1$, the other parameter, k_2, is now obtained from

$$\widehat{k}_2 = \tilde{x}/\left[\left(\widehat{k}_1 - 1\right)/\widehat{k}_1\right]^{1/\widehat{k}1}$$

Example 1.14 Suppose sample data provides estimates of $\tilde{x} = 100$ and $x_{0.90} = 300$. Note $\alpha = 0.90$. To find the estimate of k_1, the following computations are needed to begin the iterative search.

$$\tilde{x}/x_\alpha = 100/300 = 0.33$$

$$[(k_1 - 1)/k_1]^{1/k1}/\ln(1/1 - a)^{1/k1} = [(k_1 - 1)/k_1]^{1/k1}/\ln(1/1 - 0.90)^{1/k1}$$
$$= [(k_1 - 1)/k_1]^{1/k1}/\ln(10)^{1/k1}$$

Note the left-hand-side (LHS) of the equation below. An iterative search of k_1 is now followed until the LHS is near to 0.33, i.e.,

$$\text{LHS} = [(k_1 - 1)/k_1]^{1/k1}/\ln(10)^{1/k1}$$
$$= 0.33$$

The search for k_1 begins with $k_1 = 2.00$, and continues until $k_1 = 1.62$.

$$\text{At } k_1 = 2.00, \text{LHS} = 0.465$$
$$\text{At } k_1 = 1.70, \text{LHS} = 0.363$$
$$\text{At } k_1 = 1.60, \text{LHS} = 0.321$$
$$\text{At } k_1 = 1.65, \text{LHS} = 0.343$$
$$\text{At } k_1 = 1.62, \text{LHS} = 0.330$$

Hence, $\widehat{k}_1 = 1.62$.

So now, the estimate of k_2 is the following:

$$\widehat{k}_2 = (\tilde{x})/\left[\left(\widehat{k}_1 - 1\right)/\widehat{k}_1\right]^{1/\widehat{k}_1}$$
$$= (100)/[(1.62 - 1.00)/1.62]^{1/1.62}$$
$$= 180.9$$

Example 1.15 Suppose an analyst has the following data for an item: $k_1 = 1.62$, $k_2 = 180.9$, and the location parameter is $\gamma = 1000$ h. He/she wants to find the probability of an item failing prior to $w = 1200$ h, where $w = \gamma + x$. The probability is obtained as below:

$$F(w = 1200) = F(x = 200) = 1 - \exp\left[-(200/180.9)^{1.62}\right] = 0.692$$

Note, the probability of a failure, F(w), listed below for selective values of $w = \gamma + x$ ranging from 1010 to 1600 h when $k_1 = 1.62$ and $k_2 = 180.9$.

w	1010	1025	1050	1100	1200	1300	1400	1500	1600
F(w)	0.01	0.04	0.12	0.31	0.69	0.90	0.97	0.99	1.00

1.13 Normal

The normal distribution, $N(\mu,\sigma^2)$, is symmetrical with a bell shaped density. This distribution originates from Johann Gauss, a German mathematician of the early 1800s. Its parameters are: the mean μ, and the standard deviation σ. This is perhaps the most widely used probability distribution in business and scientific applications. A full description on the standard normal is given in Chap. 3 (Standard Normal).

Standard Normal Distribution

A companion distribution, the standard normal distribution, N(0,1), has a mean of zero, a standard deviation of one, and has the same shape as the normal distribution. The notations for the normal and standard normal variables are the following:

$$\mathrm{x} = \text{normal variable}$$
$$\mathrm{z} = \text{standard normal variable}$$

and the conversion from one to the other are shown below:

$$\mathrm{z} = (\mathrm{x} - \mu)/\sigma$$

$$\mathrm{x} = \mu + z\sigma$$

When k represents a particular value of z, the probability density is:

$$\mathrm{f(k)} = \left(1/\sqrt{2\pi}\right)\exp\left(-\mathrm{k}^2/2\right)$$

The cumulative distribution function yielding the probability (z less than k) becomes:

$$\mathrm{F(k)} = \int_{-\infty}^{k} f(z)dz$$

Although there is no closed-form solution for the cumulative distribution, over the years, a series of approximate methods have been developed. In Chap. 3 (Standard Normal), one such approximation method is described, and is used throughout this book. Fig. 1.3 depicts the shape of the standard normal distribution.

Coefficient of Variation

The range of the normal variable x is almost entirely from L to H where:

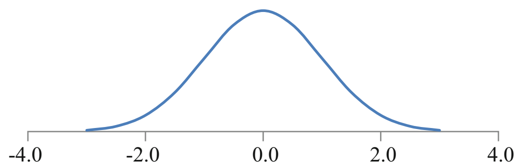

Fig. 1.3 The Standard Normal Distribution

$$L = \mu - 3\sigma$$
$$H = \mu + 3\sigma$$

When L = 0,

$$\mu = 3\sigma$$

and

$$\text{cov} = \sigma/\mu = 0.33$$

Parameter Estimates

When n sample data, (x_1, . . ., x_n), is available and the sample average, $\bar{x}$, and variance, s^2, are computed, the corresponding estimates of the normal parameters are the following:

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma}^2 = s^2$$

Example 1.16 Suppose an experiment yields ten sample data entries as: [2.3, 7.4, 8.1, 9.7, 10.1, 11.2, 12.5, 15.3, 16.1, 19.0]. The average is $\bar{x}$ = 11.17 and the standard deviation is s = 4.83. Hence,

$$\hat{\mu} = 11.17$$
$$\hat{\sigma}^2 = 4.83^2$$

1.14 Lognormal

The lognormal distribution, LN(μ_y,σ_y^2), with variable x > 0, has a density that peaks early and skews far to the right. A full description of the distribution is given in Chap. 9 (Lognormal). The lognormal variable is related to a counterpart normal variable y, in the following way:

$$y = \ln(x)$$
$$x = e^y$$

The variable y is normally distributed with mean and variance, μ_y and σ_y^2, respectively, and x is lognormal with mean and variance, μ_x and σ_x^2. The notation for x and y are as below:

$$x \sim LN\left(\mu_y, \sigma_y^2\right)$$

$$y \sim N\left(\mu_y, \sigma_y^2\right)$$

The parameters that define the distribution of x, are the mean and variance of y. The parameters between x and y are related in the following way.

$$\mu_x = \exp\left[\mu_y + \sigma_y^2/2\right]$$

$$\sigma_x^2 = \exp\left[2\mu_y + \sigma_y^2\right]\left[\exp\left(\sigma_y^2\right) - 1\right]$$

$$\mu_y = \ln\left[\mu_x^2/\sqrt{\mu_x^2 + \sigma_x^2}\right]$$

$$\sigma_y^2 = \ln\left[1 + \sigma_x^2/\mu_x^2\right]$$

The cov of lognormal variable x is much larger than its counterpart normal variable y.

Parameter Estimates

Consider a lognormal variable x with sample data, $(x_1, \ldots, x_n)$. The data are transformed to their normal counterparts by $y = \ln(x)$, and thereby, the data entries become: $(y_1, \ldots, y_n)$. So now, the average and variance of the y entries are computed and labeled as: $\bar{y}$, and s_y^2. The estimates of the parameters for the lognormal distribution become:

$$\hat{\mu}_y = \bar{y}$$

$$\hat{\sigma}_y^2 = s_y^2$$

Example 1.17 Assume the analyst has collected ten sample entries as: [4.3, 5.8, 3.2, 2.4, 5.8, 6.1, 7.6, 9.2, 10.3, 33.9]. The sample mean and standard deviation of x are $\bar{x} = 8.86$, $s = 9.13$, respectively, and thereby the cov = 1.03. Upon taking the natural logarithm of each, $y = \ln(x)$, the sample now has ten variables on y. The corresponding values of y are: [1.459, 1.758, 1.163, 0.875, 1.758, 1.808, 2.028, 2.219, 2.332, 3.523]. The mean and variance of the $n = 10$ observations of y are $\bar{y} = 1.892$ and $s_y^2 = 0.144^2$. Hence, the lognormal variable is denoted as:

$$x \sim LN\left(1.892, 0.144^2\right)$$

1.15 Summary

Eight of the common continuous distributions are described and these are: continuous uniform, exponential, Erlang, gamma, beta, Weibull, normal and lognormal. For each, the probability density, parameters, and the admissible range are defined. The cumulative probability function is also listed when available; but on some distributions, there is no closed-form solution and quantitative methods are needed to compute the cumulative probability. When sample data is obtainable, the computations to estimate the parameter values from the data are shown. Examples are presented to demonstrate the use of the distributions and the way to estimate their parameters.

References

1. Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis*. Boston: McGraw Hill.
2. Hasting, N. A. J., & Peacock, J. B. (1974). *Statistical distributions*. New York: Wiley & Sons.
3. Hines, W. W., Montgomery, L. D. C., Goldsman, D. M., & Burror, C. M. (2003). *Probability and statistics for engineers*. New York: Wiley & Sons.

Chapter 2
Discrete Distributions

2.1 Introduction

When the outcome of an independent experiment trial includes a specified set of values, usually integers, the outcome is a discrete variable. For example, the number of dots in a roll of two dice can take on only the integer numbers 2 to 12. When probabilities are assigned to each outcome and the sum over all possible outcomes is one, the variable, x, becomes a discrete random variable. The chapter describes the following six common discrete probability distributions: discrete uniform, binomial, geometric, Pascal, Poisson, and hyper geometric. For each of these, the probability function is listed, along with parameters of the function. Included, also are the mean, variance and applications on each distribution. When sample data is available, the analyst applies measures from the data to estimate the parameter values. The discrete uniform distribution applies when all the discrete integers between and including two limits (a,b) are possible outcomes and each can occur with an equal probability. The binomial occurs when n trials with a constant probability per trial yields zero to n successes. The geometric happens when n trials are needed to obtain the first success and the probability of a success is constant per trial. The Pascal is when n trials are needed to obtain k successes and the probability of a success is constant per trial. The Poisson occurs when the average rate of events occurring in specified duration is known. The hyper geometric happens with n trials are taken without replacement from a population of size N that has D defectives.

Law and Kelton [1]; Hasting and Peacock [2]; and Hines et al. [3] present thorough descriptions on the properties of the common discrete probability distributions.

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_2

2.2 Discrete Random Variables

A discrete random variable includes a set of exact values where each has a specific probability of occurrence by chance, and the sum of all the probabilities is equal to one. The event could be the number of dots on a roll of die (1, 2, 3, 4, 5, 6), or the number of heads in four tosses of a coin (0, 1, 2, 3, 4,), so forth. The more common discrete random variables are listed below with their typical designations along with parameters.

Discrete uniform	x ~ DU(a,b)
Binomial	x ~ Bin(n,p)
Geometric	x ~ Geo(p)
Pascal	x ~ Pa(k,p)
Poisson	x ~ Po(θ)
Hyper Geometric	x ~ HG(n, N, D)

Lexis Ratio

A statistic that helps to determine the discrete distribution that applies to a set of sample data is the lexis ratio, $\tau = \frac{\sigma^2}{\mu}$. The lexis ratio is computed from sample data by $\widehat{\tau} = \frac{\widehat{\sigma}^2}{\widehat{\mu}}$, where $\widehat{\mu} = \bar{x}$ = sample average, and $\widehat{\sigma}^2 = s^2$ = sample variance.

2.3 Discrete Uniform

When a variable x takes on all integers from a to b with equal probabilities, the variable follows the discrete uniform distribution, DU(a,b). The probability of x is:

$$P(x) = 1/(b - a + 1) \quad x = a, a + 1, \ldots, b$$

and the cumulative distribution function is:

$$F(x) = (x - a + 1)/(b - a + 1) \quad x = a, a + 1, \ldots, b$$

The expected value and variance of x are listed below:

$$E(x) = \mu = (a + b)/2$$

$$V(x) = \sigma^2 = \left[(b - a + 1)^2 - 1\right]/12$$

Parameter Estimates

When sample data is collected, $(x_1, \ldots, x_n)$, the common statistics measured are as follows:

$\bar{x}$ = sample average
s^2 = sample variance
x(1) = minimum
x(n) = maximum

The above stats are used to estimate the parameters as shown below. The MLE method yields:

$$\widehat{a} = x(1)$$

$$\widehat{b} = x(n)$$

Another way to estimate the parameters is by the method-of-moments as shown below:

$$\widehat{a} = \text{floor integer of } \left(\bar{x} + 0.5 - 0.5\sqrt{12s^2 + 1}\right)$$

$$\widehat{b} = \text{ceiling integer of } \left(2\bar{x} - \widehat{a}\right)$$

Example 2.1 Suppose an experiment yields ten discrete sample data entries [7, 5, 4, 8, 5, 4, 12, 9, 2, 8], and estimates of the min and max parameters from a discrete uniform distribution are needed. The following statistics are computed from the data:

$$\bar{x} = 6.4$$

$$s^2 = 8.711$$

$$x(1) = 2$$

$$x(n) = 12$$

Using the MLE method, the parameter estimates are:

$$\widehat{a} = 2$$

$$\widehat{b} = 12$$

The method-of-moment estimate of the parameters are as follows:

$$\widehat{a} = \text{floor}\left(6.4 + 0.5 - 0.5\sqrt{12 \times 8.711 + 1}\right) = \text{floor}\,(1.764) = 1$$

$$\widehat{b} = \text{ceiling}\,(2 \times 6.4 - 1.764) = \text{ceiling}\,(11.036) = 12$$

2.4 Binomial

When a variable, x, is the number of successes from n independent trials, with p the probability of success per trial, x follows the binomial distribution, Bin(n,p). The variable x can take on the integer values of zero to n.

The probability of x in n trials is listed below:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, \ldots, n$$

The parameters are:

n = number of trials
p = probability of a success per trial

The cumulative distribution function of x, F(x), is the probability of the variable achieving the value of x or smaller. When $x = x_o$,

$$F(x_o) = P(x \le x_o).$$

The mean, variance, and standard deviation of x are below:

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

$$\sigma = \sqrt{np(1-p)}$$

Lexis Ratio

The lexis ratio, τ, is always less than one as shown below:

$$\tau = \sigma^2/\mu = np(1-p)/np = (1-p) < 1.$$

Parameter Estimates

When a sample of n trials yields x successes, the MLE of p is,

$$\hat{p} = x/n$$

The associated standard deviation becomes:

$$s_p = \sqrt{\hat{p}(1 - \hat{p})/n}$$

In the event the n trial experiment is run m times, and the results are $(x_1, \ldots, x_m)$, the estimate of p becomes,

$$\hat{p} = \sum x_i/(nm)$$

When the i-th experiment has n_i trials and x_i successes, the estimate of p becomes:

$$\hat{p} = \sum x_i / \sum n_i$$

Normal Approximation

When n is large, and,

$$p \leq 0.5 \text{ with } np > 5,$$

or

$$p > 0.5 \text{ with } n(1 - p) > 5,$$

x can be approximated with the normal distribution, as follows:

$$x \sim N[np, np(1 - p)]$$

Using the normal approximation, where $\mu = np$ and $\sigma = \sqrt{np(1-p)}$, the probability of the number of successes equal or below x_o becomes:

$$\begin{aligned} P(x \leq x_o) &= P(x \leq x_o + .5) \\ &= F\left[(x_o + .5 - np)/\sqrt{np(1-p)}\right] \end{aligned}$$

where F(z) is the cumulative probability from the standard normal distribution. The probability that the number of successes is equal to x_o is obtained as below:

$$\begin{aligned} P(x = x_o) &= P(x_o - .5 \leq x \leq x_o + .5) \\ &= F\left[(x_o + .5 - np)/\sqrt{np(1-p)}\right] - F\left[(x_o - .5 - np)/\sqrt{np(1-p)}\right] \end{aligned}$$

Poisson Approximation

When n is large and p is small, and the above normal approximation does not apply, x can be approximated by the Poisson distribution with parameter θ. The probability of any x becomes:

$$P(x) = e^{-\theta}\theta^{x}/x! \qquad x = 0, 1, \ldots..$$

The estimate of θ is below:

$$\theta = np$$

Example 2.2 Suppose x is binomial with n = 10 and p = 0.30. The mean and standard deviation of the number of successes are below:

$$\begin{aligned}\mu &= np = 10 \times 0.3\\ &= 3.0\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{np(1-p)}\\ &= \sqrt{10(.3)(.7)}\\ &= 1.45\end{aligned}$$

The probability distribution on x is:

$$P(x) = \binom{10}{x} 0.3^{x} 0.7^{10-x} \qquad x = 0, 1, \ldots, 10$$

Note,

$$P(0) = \binom{10}{0} 0.3^{0} 0.7^{10} = 0.028$$

$$P(1) = \binom{10}{1} 0.3^{1} 0.7^{9} = 0.121$$

so on.

Example 2.3 An experiment is run n = 5 times and the number of successes is x = 2. The estimate on the probability of a success becomes:

$$\hat{p} = x/n = 2/5 = 0.40$$

The standard deviation on the estimate of p, denoted as s_p, is below:

$$\begin{aligned}s_p &= \left(\hat{p}\left(1-\hat{p}\right)/n\right)^{0.5}\\ &= \sqrt{.4(.6)/5}\\ &= 0.22\end{aligned}$$

Example 2.4 An experiment of n = 10 trials is run m = 4 times with success results as: (3, 4, 2, 3). The estimate on the probability of success per trial is:

$$\begin{aligned}\widehat{p} &= \sum x_i/(\mathrm{nm}) \\ &= 12/(4 \times 10) = 0.30\end{aligned}$$

The associate standard deviation becomes:

$$\begin{aligned}s_p &= \sqrt{.3(.7)/40} \\ &= 0.072\end{aligned}$$

Example 2.5 An experiment is run n = 80 times with a probability of success of p = 0.20. The analyst wants to find the probability of x less or equal to 20. Since, p =0.2≤ 0.5 and np = 16 > 5, the normal distribution can be used to approximate the probability. The mean and standard deviation of x are the following:

$$\begin{aligned}\mu &= 80 \times .2 = 16 \\ \sigma &= \sqrt{0.2(.8)80} = 3.58\end{aligned}$$

Applying Table 3.1 of Chap. 3 (Standard Normal), the cumulative probability becomes:

$$\begin{aligned}P(x \leq 20) &= F[(20.5 - 16)/3.58] \\ &= F(1.256) \\ &\approx 0.89\end{aligned}$$

Example 2.6 Suppose n = 100 trials with p = 0.01 and the analyst wants to estimate the probability of x at one or less. Because n is large and p is small, the Poisson distribution is used to estimate the probability. The mean of x becomes: θ = np = 100 × 0.01 = 1.0; and the probability is computed below:

$$\begin{aligned}P(x \leq 1) &= F(1) \\ &= P(0) + P(1) \\ &= e^{-1}1^0/0! + e^{-1}1^1/1! = 0.736\end{aligned}$$

2.5 Geometric

The variable of the geometric distribution, Geo(p), could be set as the number of trials, n, till the first success; or the number of fails, x, till the first success. Both are described below.

Number of Trials

The probability of the number of trials, n, to achieve the first success, when p is the probability of a success per trial, is obtained by the geometric distribution where n ≥ 1. The probability of n is below:

$$P(n) = p(1-p)^{n-1} \quad n = 1, 2, \ldots$$

The cumulative probability of n or less is obtained as shown here:

$$F(n) = 1 - (1-p)^{n} \quad n = 1, 2, \ldots$$

The expected value and variance of n are listed below.

$$E(n) = \mu_n = 1/p$$
$$V(n) = {\sigma_n}^2 = (1-p)/p^2$$

Number of Failures

The number of fails till the first success, x, is related to the number of trials, n, in the following way;

$$x = n - 1$$

The probability becomes,

$$P(x) = p(1-p)^{x} \quad x = 0, 1, 2, \ldots$$

The mean and variance of x are listed below:

$$\mu_x = (1-p)/p$$
$${\sigma_x}^2 = (1-p)/p^2$$

Lexis Ratio

The lexis ratio of x:

$$\begin{aligned} \tau &= {\sigma_x}^2/\mu_x \\ &= [(1-p)/p^2]/[(1-p)/p] \\ &= 1/p \end{aligned}$$

is always greater than one.

The lexis ratio of n:

$$\tau = \sigma_n{}^2/\mu_n = (1 - p)/p$$

could be larger or smaller than one.

Parameter Estimate

When a series of experiments yield m samples on the number of trials till a success as: $(n_1, \ldots, n_m)$, the MLE of p becomes:

$$\hat{p} = 1/\bar{n}$$

where:

$$\bar{n} = \sum n_i/m$$

When the variable is the number of fails till a success and the average number of fails is $\bar{x}$, the estimates of p becomes:

$$\hat{p} = 1/(\bar{x} + 1)$$

Example 2.7 Suppose the probability of a success is p = 0.2 and the variable of interest is the number of trials, n, till the first success. The probability of n is:

$$P(n) = 0.2\left(0.8^{n-1}\right) \quad n = 1, 2, \ldots.$$

Note, P(1) = 0.200; P(2) = 0.160; P(3) = 0.128; so on, and,

$$F(3) = P(n \leq 3) = 0.200 + 0.160 + 0.128 = 0.488$$

Also observe:

$$F(3) = 1 - (1 - 0.2)^3 = 1 - 0.512 = 0.488$$

Example 2.8 Assume m = 5 samples on n (number of trials till a success) are the following: (3, 8, 4, 2, 5). The average of n becomes $\bar{n}$ = 22/5 = 4.4, and hence, the estimate on the probability of a success is:

$$\hat{p} = 1/4.4 = 0.227$$

Example 2.9 Consider a situation where p = 0.10 is the probability of a success per trial and x is the number of fails till the first success. The probability of x becomes:

$$P(x) = 0.10(0.90^x) \quad x = 0, 1, \ldots$$

The mean and variance of x are:

$$\mu_x = 0.90/0.10 = 9.0$$

$$\sigma_x{}^2 = 0.90/0.10^2 = 90.0$$

and the lexis ratio is:

$$\tau = 90.0/9.0 = 10.0$$

Example 2.10 Suppose m = 8 samples from geometric data are obtained and yield the following values of x: [13, 16, 12, 15, 14, 14, 11, 15] where x is the number of failures till the first success. The analyst wants to estimate the probability of a success, p, and since the average of x is $\bar{x} = 13.75$, the estimate becomes:

$$\widehat{p} = 1/(13.75 + 1) = 0.068.$$

2.6 Pascal

A French mathematician, Blaise Pascal, first formulated the Pascal distribution during the early seventeenth century. He is also credited with other accomplishments in mathematics, and with the invention of a mechanical calculator. The random variable of the Pascal distribution, Pa(k,p), is either of two type: the number of trials, n, to obtain k successes; or the number of failures, x, to obtain k successes. Both are described below.

Number of Trials

The Pascal distribution applies when the variable is the number of trials, n, needed to achieve k successes with p the probability of a success per trial. This distribution is also called the Negative Binomial distribution. The probability of n is listed below:

$$P(n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad n = k, k+1, \ldots$$

The cumulative distribution function becomes:

$$F(n) = \sum_{y=k}^{n} P(y) \quad n = k, k+1, \ldots$$

The mean and variance of n are listed below.

$$E(n) = \mu_n = k/p$$
$$V(n) = \sigma_n{}^2 = k(1 - p)/p^2$$

Lexis Ratio

The lexis ratio on n becomes:

$$\begin{aligned}\tau &= \sigma_n{}^2/\mu_n \\ &= [k(1 - p)/p^2]/[k/p] \\ &= (1 - p)/p\end{aligned}$$

where τ could be larger or smaller than one.

Parameter Estimate

Suppose m samples on the number of trials to achieve k successes are: $(n_1, \ldots, n_m)$; with the average:

$$\bar{n} = \sum n_i/m$$

The MLE of the probability of a success becomes:

$$\widehat{p} = k/\bar{n}$$

Number of Failures

When x is the number of failures to achieve k successes, the variable x is as follows:

$$x = n - k \quad x = 0, 1, 2, \ldots$$

The mean and variance of x become:

$$\mu_x = \mu_n - k = k(1 - p)/p$$
$$\sigma_x{}^2 = \sigma_n{}^2 = k(1 - p)/p^2$$

Lexis Ratio

The lexis ratio on x becomes:

$$\begin{aligned}\tau &= {\sigma_x}^2/\mu_x \\ &= [k(1-p)/p^2]/[k(1-p)/p] \\ &= 1/p\end{aligned}$$

where τ is always larger than one.

Parameter Estimate

If m samples of x are: $(x_1, \ldots, x_m)$ and $\bar{x}$ is the average of the m samples, the estimate of p becomes:

$$\widehat{p} = k/(\bar{x} + k)$$

Example 2.11 Suppose n is the number of trials to reach k = 4 successes when the probability of a success is p = 0.60. The probability of n is listed below:

$$P(n) = \binom{n-1}{4-1}.6^4(.4)^{n-4} \qquad n = 4, 5, \ldots..$$

Note:

$$P(4) = \binom{4-1}{4-1}.6^4(.4)^{4-4} \quad = 0.1296$$

$$P(5) = \binom{5-1}{4-1}.6^4(.4)^{5-4} \quad = 0.2074$$

so on.

The probability of n = 5 or less becomes:

$$P(n \le 5) = 0.1296 + 0.2074 = 0.3370$$

The mean and variance of n are below:

$$\mu_n = 4/0.6 = 6.66$$

$${\sigma_n}^2 = 4(1-.6)/.6^2 = 4.44$$

Note, the counterpart mean and variance of x are the following:

$$\mu_x = 6.66 - 4 = 2.66$$

$$\sigma_x{}^2 = 4.44$$

The lexis ratio of n is:

$$\tau = 4.44/6.66 = 0.66$$

and, the lexis ratio of x is:

$$\tau = 4.44/2.66 = 1.66$$

Example 2.12 Suppose m = 5 samples from the Pascal distribution with parameter k = 4 are observed and yield the following data entries of n (number of trials): [16, 14, 17, 15, 16], and the analyst wants to estimate the probability of a success. Since the average is $\bar{n} = 15.60$, the estimate of p is $\widehat{p} = 4/15.60 = 0.256$.

2.7 Poisson

The Poisson distribution is named after its founder, Simeon Poisson, a famous French mathematician who, during the early 1800s, is credited with advancements in mathematics, geometry and physics. The variable, x, is distributed like a Poisson distribution, Po(θ), when events under study occur randomly with an average rate of θ for a specified time duration, area, or space. It could be the number of demands for a specific product per day at a retail location. The probability of x is the following:

$$P(x) = \theta^x e^{-\theta}/x! \quad x = 0, 1, 2, \ldots..$$

The expected value and variance of x are shown below.

$$E(x) = \mu = \theta$$

$$V(x) = \sigma^2 = \theta$$

Lexis Ratio

Since the mean and the variance are both equal to θ, the lexis ratio becomes:

$$\tau = \sigma^2/\mu = 1$$

Relation to the Exponential Distribution

The Poisson and the exponential distributions are related since the time, t, between events from a Poisson variable are distributed as exponential, Exp(θ), with E(t) = 1/θ.

Parameter Estimate

When n samples (x_1, ..., x_n) are collected, and the sample average is $\bar{x}$, the MLE of θ is:

$$\widehat{\theta} = \bar{x}$$

Example 2.11 Cars arrive to a parking lot on Monday mornings at an average rate of 2 per 15 min. For this time duration, the mean and variance of the number of cars arriving are:

$$\mu = 2$$
$$\sigma^2 = 2$$

The probability of x arrivals in 15 min becomes:

$$P(x) = e^{-2}2^x/x! \quad x = 0, 1, 2, \ldots.$$

Note,

$$P(0) = e^{-2}2^0/0! = 0.135$$
$$P(1) = e^{-2}2^1/1! = 0.271$$
$$P(2) = e^{-2}2^2/2! = 0.271$$

so on.

The probability x is two or less becomes:

$$P(x \leq 2) = 0.677$$

Example 2.12 Suppose n = 10 samples from Poisson data are observed and yield the following values of x: [1, 1, 2, 3, 3, 1, 2, 3, 1, 2], and the analyst wants to estimate the Poisson parameter, θ. Since the average of x is $\bar{x} = 1.90$, the estimate is $\widehat{\theta} = 1.90$.

2.8 Hyper Geometric

The variable x is distributed by the hyper geometric distribution, HG(n,N,D), where x represents the number of defectives in n samples taken without replacement from a population of size N with D defectives. The variable x can take on the integer values of zero to the smaller of D and n. The probability of x is the following:

$$\mathrm{P(x)} = \binom{N-D}{n-x}\binom{D}{x}/\binom{N}{n} \qquad \mathrm{x} = 0, \ldots, \min(\mathrm{n},\mathrm{D})$$

The expected value and variance of x are listed below:

$$\mathrm{E(x)} = \mu = \mathrm{nD/N}$$

$$\mathrm{V(x)} = \sigma^2 = \mathrm{n[D/N][1 - D/N][N - n]/[N - 1]}$$

The lexis ratio is smaller than one as shown below:

$$\tau = \sigma^2/\mu = [1 - \mathrm{D/N}][\mathrm{N - n}]/[\mathrm{N} - 1] < 1$$

Parameter Estimate

The ratio of defectives per unit (D/N) can be estimated from sample data. Assume, N units and n samples without replacement each day; and m days of sample number of defectives are: ($x_1, \ldots, x_m$). The average of the m samples is denoted as $\bar{x}$. Hence, the estimate of the ratio of defectives becomes:

$$\frac{\widehat{D}}{N} = \bar{x}/\mathrm{n}$$

Example 2.11 Each day a lot of N = 5 units come into a store and a sample of n = 2 are taken without replacement to seek out any defectives. If the number of defectives is D = 1, the probability of finding x defectives is below:

$$\mathrm{P(x)} = \binom{4}{2-x}\binom{1}{x}/\binom{5}{2} \qquad \mathrm{x} = 0, 1$$

The probabilities become:

$$\mathrm{P(0)} = \binom{4}{2-0}\binom{1}{0}/\binom{5}{2} \quad = 0.60$$

$$P(1) = \binom{4}{2-1}\binom{1}{1}/\binom{5}{2} \quad = 0.40$$

The mean and variance of x are computed below:

$$\mu = 2 \times 1/5 = 0.40$$

$$\sigma^2 = 2 \times 1/5[1 - 1/5][5 - 2]/[5 - 1] = 0.24$$

The lexis becomes

$$\tau = 0.24/0.40 = 0.60$$

Example 2.12 Assume the scenario of Example 2.11 where N = 5 units arrive each day and n = 2 are sampled without replacement. A sample on five days yields the following number of defectives: (0, 1, 0, 0, 0) that have an average of $\bar{x} = 0.20$. The estimate of the ratio of defectives is below:

$$\frac{\widehat{D}}{N} = 0.20/2 = 0.10$$

Hence, 10% of the units are estimated to be defective.

Example 2.13 A park manager wants to estimate the number of grown turtles in a small lake in the park. Over a series of days, he traps D = 10 turtles, and tags each and releases them back into the lake. On subsequent days, he traps n = 8 turtles and x = 2 of them have the tag attached. An estimate on the number of grown turtles, N, in the lake is obtained as below:

Since,

$$E(x) = nD/N$$

$$\begin{aligned}\widehat{N} &= nD/x \\ &= 8 \times 10/2 = 40\end{aligned}$$

2.9 Summary

Six discrete distributions are summarized: discrete uniform, binomial, geometric, Pascal, Poisson and hyper geometric. The probability function, parameters and admissible region are described for each. Estimates of the parameter values are obtained, for each of the distributions, by use of sample data. Examples are also presented to guide the user on the characteristics and calculations needed for each distribution.

References

1. Law, A. M., & Kelton, W. D. (2000). *Simulation modeling and analysis*. Boston: McGraw Hill.
2. Hasting, N. A. J., & Peacock, J. B. (1974). *Statistical distributions*. New York: Wiley & Sons.
3. Hines, W. W., Montgomery, L. D. C., Goldsman, D. M., & Burror, C. M. (2003). *Probability and statistics for engineers*. New York: Wiley & Sons.

Chapter 3
Standard Normal

3.1 Introduction

The normal distribution is perhaps the most used distribution in business, engineering and scientific research and analysis of data. The distribution is fully described by two parameters, the mean and the standard deviation. A related distribution is the standard normal that has a mean of zero and standard deviation equal to one. Almost all statistical books have table measures listed on the standard normal. An easy conversion from the normal variable to the standard normal variable and vice versa is available. Since there is no closed-form solution to the cumulative probability of the standard normal, various approximations have been developed over the years. The Hasting's approximation is applied here and table listings in the chapter are derived from the same. The standard normal variable ranges between minus and plus infinity, but almost all of the probability falls within minus and plus 3.0. For ease of calculations in the chapter and the book, only the range of the standard normal between minus and plus 3 is used. For an application in a latter chapter, (Bivariate Normal), the standard normal distribution is converted to a discrete distribution, for which the variable changes from continuous to a discrete; and table values of the discrete normal are listed here for later use.

3.2 Gaussian Distribution

The normal distribution is often referred as the Gaussian distribution, named after Johann C. F. Gauss, a German mathematician, who first formulated and published the distribution in 1809 [1]. The normal distribution is a widely used probability distribution and applies in all type of disciplines. The normal distribution also evolves when the number of independent samples of any shaped variable increases,

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_3

as stated by the central-limit theorem. Further, the distribution is symmetrical and is bell shaped.

The standard normal distribution is a derivation of the normal distribution. It has a mean of zero and a standard deviation of one, and can be converted from and to any normal distribution. This is the distribution whose measures are published in almost all statistical and probability books.

3.3 Some Relations on the Standard Normal Distribution

z = standard normal variable
k = a particular value of z

$f(z) = \left(1/\sqrt{2\pi}\right)e^{-z^2/2}$ = probability density of z

$F(k) = P(z \leq k) = \int_{-\infty}^{k} f(z)dz$ = cumulative probability of z = k

$H(k) = P(z > k) = 1 - F(k)$ = complementary probability of z = k

$$\int_{-\infty}^{k} zf(z)dz = -f(k)$$

$$\int_{k}^{\infty} zf(z)dz = f(k)$$

$$\int_{-\infty}^{\infty} zf(z)dz = 0$$

$$\int_{-\infty}^{k} z^2 f(z)dz = -kf(k) + F(k)$$

$$\int_{k}^{\infty} z^2 f(z)dz = kf(k) + H(k)$$

$$\int_{-\infty}^{\infty} z^2 f(z)dz = 1$$

Note since z is a continuous variable, $F(k) = P(z \leq k) = P(z < k)$.

3.4 Normal Distribution

The random variable of the normal distribution is denoted as x and has mean μ and standard deviation σ. The common notation on x and the normal distribution is below,

$$x \sim N(\mu, \sigma^2)$$

The approximate low (L) and high (H)) limits on x are the following:

$$L = \mu - 3\sigma$$
$$H = \mu + 3\sigma$$

The probability that x falls between L and H is near unity, since:

$$P(L \leq x \leq H) \approx 0.998$$

3.5 Standard Normal

The random variable of the standard normal distribution is denoted as z and has mean zero and standard deviation one. The notation for z is listed below:

$$z \sim N(0, 1)$$

The way to convert x (normal distribution) to z (standard normal distribution), and vice versa is shown below:

$$z = (x - \mu)/\sigma$$
$$x = \mu + z\sigma$$

The approximate low (L) and high (H) limits on z are listed below:

$$L = -3$$
$$H = +3$$

and the probability of z within these limits is almost unity, since,

$$P(L \leq z \leq H) \approx 0.998$$

For simplicity in this book, only the values of z ranging from −3.0 to +3.0 are in consideration. See Fig. 3.1.

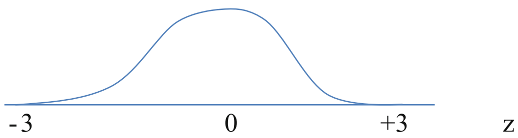

Fig. 3.1 The standard normal distribution

3.6 Hastings Approximations

Because there is no closed-form solution for the cumulative distribution F(z); various researchers have developed estimate algorithms for the cumulative probability of the standard normal. Two approximations are used in this book and both are due to C. Hastings, Jr. [2, 3].

Approximation of F(z) from z For a given z, to find F(z), the following Hastings routine is run.

1. $d_1 = 0.0498673470$
   $d_2 = 0.0211410061$
   $d_3 = 0.0032776263$
   $d_4 = 0.0000380036$
   $d_5 = 0.0000488906$
   $d_6 = 0.0000053830$
2. If $z \geq 0$: $w = z$
   If $z < 0$: $w = -z$
3. $F = 1–0.5[1 + d_1w + d_2w^2 + d_3w^3 + d_4w^4 + d_5w^5 + d_6w^6]^{-16}$
4. if $z \geq 0$: $F(z) = F$
   If $z < 0$: $F(z) = 1 – F$

   Return F(z).

Approximation of z from F(z) Another useful approximation also comes from Hastings, and gives a routine that yields a random z from a value of F(z). The routine is listed below.

1. $c_0 = 2.515517$
   $c_1 = 0.802853$
   $c_2 = 0.010328$
   $d_1 = 1.432788$
   $d_2 = 0.189269$
   $d_3 = 0.001308$
2. $H(z) = 1 – F(z)$
   If $H(z) \leq 0.5$: $H = H(z)$
   If $H(z) > 0.5$: $H = 1 – H(z)$
3. $t = \sqrt{\ln\left(1/H^2\right)}$ where ln = natural logarithm.
4. $w = t – [c_0 + c_1t + c_2t^2]/[1 + d_1t + d_2t^2 + d_3t^3]$
5. If $H(z) \leq 0.5$: $z = w$
   If $H(z) > 0.5$: $z = -w$.

   Return z.

3.7 Table Values of the Standard Normal

Table 3.1 lists values of k, F(k), H(k), and f(k), from the standard normal distribution that fall in the range: [−3.0, (0.1), +3.0]. In a similar way, Table 3.2 contains values of F(z), z, H(z) and f(z) for the range of F(z): [0.01, (0.01), 0.99]. Figure 3.2 depicts F(k), f(k) and k for standard normal variable z.

Example 3.1 Suppose an analyst has experiment results where x is normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 10$, i.e., $x \sim N(100, 10^2)$. Assume the analyst wants to find the probability of x less or equal to 115. To accomplish, the following three steps are applied:

Table 3.1 The standard normal statistics sorted by k, with cumulative distribution, F(k), complementary distribution, H(k), and probability density, f(k)

k	F(k)	H(k)	f(k)	k	F(k)	H(k)	f(k)
−3.0	0.001	0.999	0.004	0.0	0.500	0.500	0.399
−2.9	0.002	0.998	0.006	0.1	0.540	0.460	0.397
−2.8	0.003	0.997	0.008	0.2	0.579	0.421	0.391
−2.7	0.003	0.997	0.010	0.3	0.618	0.382	0.381
−2.6	0.005	0.995	0.014	0.4	0.655	0.345	0.368
−2.5	0.006	0.994	0.018	0.5	0.691	0.309	0.352
−2.4	0.008	0.992	0.022	0.6	0.726	0.274	0.333
−2.3	0.011	0.989	0.028	0.7	0.758	0.242	0.312
−2.2	0.014	0.986	0.035	0.8	0.788	0.212	0.290
−2.1	0.018	0.982	0.044	0.9	0.816	0.184	0.266
−2.0	0.023	0.977	0.054	1.0	0.841	0.159	0.242
−1.9	0.029	0.971	0.066	1.1	0.864	0.136	0.218
−1.8	0.036	0.964	0.079	1.2	0.885	0.115	0.194
−1.7	0.045	0.955	0.094	1.3	0.903	0.097	0.171
−1.6	0.055	0.945	0.111	1.4	0.919	0.081	0.150
−1.5	0.067	0.933	0.130	1.5	0.933	0.067	0.130
−1.4	0.081	0.919	0.150	1.6	0.945	0.055	0.111
−1.3	0.097	0.903	0.171	1.7	0.955	0.045	0.094
−1.2	0.115	0.885	0.194	1.8	0.964	0.036	0.079
−1.1	0.136	0.864	0.218	1.9	0.971	0.029	0.066
−1.0	0.159	0.841	0.242	2.0	0.977	0.023	0.054
−0.9	0.184	0.816	0.266	2.1	0.982	0.018	0.044
−0.8	0.212	0.788	0.290	2.2	0.986	0.014	0.035
−0.7	0.242	0.758	0.312	2.3	0.989	0.011	0.028
−0.6	0.274	0.726	0.333	2.4	0.992	0.008	0.022
−0.5	0.309	0.691	0.352	2.5	0.994	0.006	0.018
−0.4	0.345	0.655	0.368	2.6	0.995	0.005	0.014
−0.3	0.382	0.618	0.381	2.7	0.997	0.003	0.010
−0.2	0.421	0.579	0.391	2.8	0.997	0.003	0.008
−0.1	0.460	0.540	0.397	2.9	0.998	0.002	0.006
				3.0	0.999	0.001	0.004

Table 3.2 Standard normal distribution sorted by cumulative distribution F(z); with variable z; complement probability H(z); and probability density f(z)

F(z)	z	H(z)	f(z)
0.01	−2.327	0.990	0.027
0.02	−2.054	0.980	0.048
0.03	−1.881	0.970	0.068
0.04	−1.751	0.960	0.086
0.05	−1.645	0.950	0.103
0.06	−1.555	0.940	0.119
0.07	−1.476	0.930	0.134
0.08	−1.405	0.920	0.149
0.09	−1.341	0.910	0.162
0.10	−1.282	0.900	0.175
0.11	−1.227	0.890	0.188
0.12	−1.175	0.880	0.200
0.13	−1.126	0.870	0.212
0.14	−1.080	0.860	0.223
0.15	−1.036	0.850	0.233
0.16	−0.994	0.840	0.243
0.17	−0.954	0.830	0.253
0.18	−0.915	0.820	0.262
0.19	−0.878	0.810	0.271
0.20	−0.841	0.800	0.280
0.21	−0.806	0.790	0.288
0.22	−0.772	0.780	0.296
0.23	−0.739	0.770	0.304
0.24	−0.706	0.760	0.311
0.25	−0.674	0.750	0.318
0.26	−0.643	0.740	0.324
0.27	−0.612	0.730	0.331
0.28	−0.582	0.720	0.337
0.29	−0.553	0.710	0.342
0.30	−0.524	0.700	0.348
0.31	−0.495	0.690	0.353
0.32	−0.467	0.680	0.358
0.33	−0.439	0.670	0.362
0.34	−0.412	0.660	0.366
0.35	−0.385	0.650	0.370
0.36	−0.358	0.640	0.374
0.37	−0.331	0.630	0.378
0.38	−0.305	0.620	0.381
0.39	−0.279	0.610	0.384
0.40	−0.253	0.600	0.386
0.41	−0.227	0.590	0.389
0.42	−0.202	0.580	0.391

(continued)

Table 3.2 (continued)

F(z)	z	H(z)	f(z)
0.43	−0.176	0.570	0.393
0.44	−0.151	0.560	0.394
0.45	−0.125	0.550	0.396
0.46	−0.100	0.540	0.397
0.47	−0.075	0.530	0.398
0.48	−0.050	0.520	0.398
0.49	−0.025	0.510	0.399
0.50	0.000	0.500	0.399
0.51	0.025	0.490	0.399
0.52	0.050	0.480	0.398
0.53	0.075	0.470	0.398
0.54	0.100	0.460	0.397
0.55	0.125	0.450	0.396
0.56	0.151	0.440	0.394
0.57	0.176	0.430	0.393
0.58	0.202	0.420	0.391
0.59	0.227	0.410	0.389
0.60	0.253	0.400	0.386
0.61	0.279	0.390	0.384
0.62	0.305	0.380	0.381
0.63	0.331	0.370	0.378
0.64	0.358	0.360	0.374
0.65	0.385	0.350	0.370
0.66	0.412	0.340	0.366
0.67	0.439	0.330	0.362
0.68	0.467	0.320	0.358
0.69	0.495	0.310	0.353
0.70	0.524	0.300	0.348
0.71	0.553	0.290	0.342
0.72	0.582	0.280	0.337
0.73	0.612	0.270	0.331
0.74	0.643	0.260	0.324
0.75	0.674	0.250	0.318
0.76	0.706	0.240	0.311
0.77	0.739	0.230	0.304
0.78	0.772	0.220	0.296
0.79	0.806	0.210	0.288
0.80	0.841	0.200	0.280
0.81	0.878	0.190	0.271
0.82	0.915	0.180	0.262
0.83	0.954	0.170	0.253
0.84	0.994	0.160	0.243

(continued)

Table 3.2 (continued)

F(z)	z	H(z)	f(z)
0.85	1.036	0.150	0.233
0.86	1.080	0.140	0.223
0.87	1.126	0.130	0.212
0.88	1.175	0.120	0.200
0.89	1.227	0.110	0.188
0.90	1.282	0.100	0.175
0.91	1.341	0.090	0.162
0.92	1.405	0.080	0.149
0.93	1.476	0.070	0.134
0.94	1.555	0.060	0.119
0.95	1.645	0.050	0.103
0.96	1.751	0.040	0.086
0.97	1.881	0.030	0.068
0.98	2.054	0.020	0.048
0.99	2.327	0.010	0.027

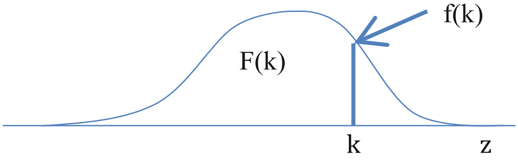

Fig. 3.2. F(k) with f(k), k and z from the standard normal distribution

1. Convert x to z as: $z = (x - \mu)/\sigma = (115\text{–}100)/10 = 1.5$.
2. Search Table 3.1 to find: $F(1.5) = 0.933 = P(z \leq 1.5)$.
3. Since $P(x \leq 115) = P(z < 1.5)$, the probability sought is 0.933.

Example 3.2 Consider the data from Example 3.1 again where $x \sim N(100, 10^2)$. Assume the analyst is now seeking the probability of x falling between 80 and 90. To find this probability, i.e. $p(80 < x < 90)$, the following four steps are taken:

1. $z_L = (x - \mu)/\sigma = (80\text{–}100)/10 = -2.0$
2. $z_H = (x - \mu)/\sigma = (90\text{–}100)/10 = -1.0$
3. Table 3.1 yields: $F(-2.0) = 0.023$ and $F(-1.0) = 0.159$
4. $P(80 < x < 90) = F(z_H) - F(z_L) = F(-1.0) - F(-2.0) = 0.159 - 0.023 = 0.136$

Example 3.3 Suppose a situation where x is normally distributed as follows: $x \sim N(5,1^2)$, and a researcher needs to find the value of x_o, where the probability of x exceeding x_o is 0.05, i.e., $P(x > x_o) = 0.05$. In this situation, the following three steps are taken:

1. Table 3.2 is searched to find H(1.645) = 0.05.
2. The corresponding value of x is $x_o = \mu + z\sigma = 5 + 1.645 \times 1 = 6.645$
3. Hence, P(x > 6.645) = 0.05.

Example 3.4 For the random variable, $x \sim N(8,2^2)$, find the mid (L,H) where p (L < x < H) = 0.50. To obtain, the three steps are below:

1. Table 3.2 shows F(−0.674) = 0.25 and F(0.674) = 0.75.
2. Hence, L = 8–0.674 × 2 = 6.652 and H = 8 + 0.674 × 2 = 9.348.
3. Thereby, P(6.652 < x < 9.348) = 0.50.

3.8 Discrete Normal Distribution

Consider an adaptation of the standard normal distribution where only a finite set of discrete values are permitted. Let these values be the following: (−3.0, −2.9, ..., 2.9, 3.0), where the total set is denoted as: k = [−3.0, (0.1), 3.0]. Altogether, there are 61 discrete possible values. For simplicity, this distribution is here named the discrete normal distribution. It is needed in Chap. 7 (Bivariate Normal) to estimate the probabilities associated with the bivariate normal distribution.

The probability of discrete variable k, P(k), is obtained as follows:

$$\mathrm{P(k) = f(k)}/\sum_{z=-3.0}^{3.0} f(z) \quad \text{for k} = [-3.0, (0.1), 3.0]$$

where f(k) is the probability density of z = k from the standard normal distribution. The cumulative probability of z less or equal to k, P(z ≤ k), is denoted as F(k). Table 3.3 lists the complete set of this adapted discrete normal distribution, with values of k, P(k) and F(k).

Example 3.5 Consider the discrete normal distribution, and note from Table 3.3 that the most likely value of k is 0.0, since P(0.0) = 0.040 is the largest of all probabilities. Note also, the expected number of random trials to obtain k = 0.0 is 1/P(0.0) = 1/0.0400 = 25.

Example 3.6 Consider once more the discrete normal distribution. The least likely values of k to occur by chance are k = −3.0 and k = 3.0 since the probability of each is the smallest at P(−3.0) = P(3.0) = 0.0004.

3.9 Summary

The standard normal distribution with variable z is a special case of the normal distribution with variable x, and conversion from one to the other is easily done. The Hasting’s approximation of the cumulative distribution for the standard normal is

Table 3.3 Discrete normal distribution with discrete normal variable, k; probability, P(k); and cumulative probability, F(k)

k	P(k)	F(k)
−3.0	0.0004	0.0004
−2.9	0.0006	0.0010
−2.8	0.0008	0.0018
−2.7	0.0010	0.0029
−2.6	0.0014	0.0042
−2.5	0.0018	0.0060
−2.4	0.0022	0.0082
−2.3	0.0028	0.0111
−2.2	0.0036	0.0146
−2.1	0.0044	0.0190
−2.0	0.0054	0.0245
−1.9	0.0066	0.0310
−1.8	0.0079	0.0389
−1.7	0.0094	0.0484
−1.6	0.0111	0.0595
−1.5	0.0130	0.0725
−1.4	0.0150	0.0875
−1.3	0.0172	0.1047
−1.2	0.0195	0.1241
−1.1	0.0218	0.1460
−1.0	0.0243	0.1702
−0.9	0.0267	0.1969
−0.8	0.0290	0.2259
−0.7	0.0313	0.2572
−0.6	0.0334	0.2906
−0.5	0.0353	0.3259
−0.4	0.0369	0.3628
−0.3	0.0382	0.4010
−0.2	0.0392	0.4402
−0.1	0.0398	0.4800
0.0	0.0400	0.5200
0.1	0.0398	0.5598
0.2	0.0392	0.5990
0.3	0.0382	0.6372
0.4	0.0369	0.6741
0.5	0.0353	0.7094
0.6	0.0334	0.7428
0.7	0.0313	0.7741
0.8	0.0290	0.8031
0.9	0.0267	0.8298
1.0	0.0243	0.8541
1.1	0.0218	0.8759

(continued)

Table 3.3 (continued)

k	P(k)	F(k)
1.2	0.0195	0.8954
1.3	0.0172	0.9125
1.4	0.0150	0.9275
1.5	0.0130	0.9405
1.6	0.0111	0.9516
1.7	0.0094	0.9611
1.8	0.0079	0.9690
1.9	0.0066	0.9756
2.0	0.0054	0.9810
2.1	0.0044	0.9854
2.2	0.0036	0.9889
2.3	0.0028	0.9918
2.4	0.0022	0.9940
2.5	0.0018	0.9958
2.6	0.0014	0.9971
2.7	0.0010	0.9982
2.8	0.0008	0.9990
2.9	0.0006	0.9996
3.0	0.0004	1.0000

described and is used to generate the tables of the chapter. The variable z ranges from minus to plus infinity, but almost all the probability is from -3.0 to $+3.0$, and for simplicity, this is the range used in the chapter. One table lists statistical measure for k, a particular value of z, ranging from -3.0 to $+3.0$. Another table is for the cumulative probability, F(z), ranging from 0.01 to 0.99. For use in a subsequent chapter, an adaption of the standard normal to a discrete normal is developed.

References

1. Gauss, J. C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium.*
2. Hastings, C., Jr. (1955). *Approximation for digital computers*. Princeton: Princeton University Press.
3. Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs and tables* (pp. 931–937). Washington, DC: U.S. Department of Commerce.

Chapter 4
Partial Expectation

4.1 Introduction

This chapter concerns a particular value of the standard normal, k, called the left-location parameter, and the average of the difference of all standard normal z values larger than k. This is called the partial expectation of z greater than k, and is denoted as $E(z > k)$. A table of $E(z > k)$ for $k = -3.0$ to $+3.0$ is listed. Another partial described in this chapter is when the location parameter is on the right-hand side, and of interest is the average of the difference of all standard normal z values smaller than k. This is called the partial expectation of z less than k, and is denoted as $E(z < k)$. A table of $E(z < k)$ for $k = -3.0$ to $+3.0$ is listed. These measures are of particular interest in inventory management when determining when to order new stock for an item and how much. The partial expectation is used to compute the minimum amount of safety stock for an item to control the percent fill. The percent fill is the portion of total demand that is immediately filled from stock available. Another use in inventory management is an adjustment to the forecast of an item when advance demand becomes available. Advance demand occurs when a customer orders stock that is not to be delivered until a future date. Several examples are presented to guide the user on the use of the partial expectation.

In 1959, and 1962, Robert G. Brown showed how the partial expectation of the standard normal distribution can be used to compute the safety stock for an inventory item [1, 2]. In 1980, N. Thomopoulos developed and published tables on the partial expectation; and demonstrated their application in inventory management for computing the safety stock of a stocked item; and also for adjusting a previous forecast on a stock item when advance demands arrive, [3].

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_4

4.2 Partial Expectation

The partial expectation is a measure concerning a portion of the standard normal distribution; either on the right-hand side or on the left-hand side of the standard normal. To illustrate, assume the standard normal variable, z, and a particular value of k, where of interest is the average of (z – k) for all values of z larger than k. This average is called the partial expectation of z greater than k; and k is the left-location parameter. When the value of k is on the right-hand side of z, the measure of interest is the average of (z – k) for all values of z below k; and hence, this is the partial expectation of z smaller than k; where k is now the right location parameter. An important use of the partial expectation is in inventory management where it is needed to compute the size of the safety stock for each item held in stock. It is also used to adjust forecasts when advance demand information is available.

4.3 Left Location Parameter

Consider the standard normal distribution where a location parameter, k, is stated and only the values of z larger than k are considered. Hence (z > k) and of interest is the expected value of z larger than k. In this way, k is the left-most value of z. This is called the partial expectation of z greater than k and is denoted as E(z > k) as shown in Fig. 4.1. Below lists an identity of the expectation:

$$\mathrm{E(z > k)} = \int_k^\infty (z-k)f(z)dz = \mathrm{f(k) - kH(k)}$$

Recall, f(k) is the probability density of the standard normal when z = k, and H (k) is the probability of z larger than k.

Fig. 4.1 The partial expectation E(z > k) with k and z from the standard normal distribution

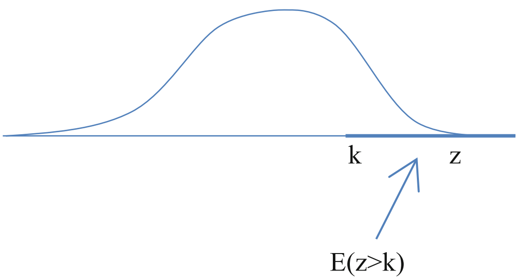

Table 4.1 Partial expectation E(z > k) sorted by the left location parameter k

k	E(z > k)	k	E(z > k)
−3.0	3.000	0.0	0.399
−2.9	2.901	0.1	0.351
−2.8	2.801	0.2	0.307
−2.7	2.701	0.3	0.267
−2.6	2.601	0.4	0.230
−2.5	2.502	0.5	0.198
−2.4	2.403	0.6	0.169
−2.3	2.304	0.7	0.143
−2.2	2.205	0.8	0.120
−2.1	2.106	0.9	0.100
−2.0	2.008	1.0	0.083
−1.9	1.911	1.1	0.069
−1.8	1.814	1.2	0.056
−1.7	1.718	1.3	0.046
−1.6	1.623	1.4	0.037
−1.5	1.529	1.5	0.029
−1.4	1.437	1.6	0.023
−1.3	1.346	1.7	0.018
−1.2	1.256	1.8	0.014
−1.1	1.169	1.9	0.011
−1.0	1.083	2.0	0.008
−0.9	1.000	2.1	0.006
−0.8	0.920	2.2	0.005
−0.7	0.843	2.3	0.004
−0.6	0.769	2.4	0.003
−0.5	0.698	2.5	0.002
−0.4	0.630	2.6	0.001
−0.3	0.567	2.7	0.001
−0.2	0.507	2.8	0.001
−0.1	0.451	2.9	0.001
		3.0	0.000

Table Entries

Table 4.1 is a list of the values of E(z > k) for the set of location parameters, k, with the range: [−3.0, (0.1), +3.0]. Note where E(z > k) starts with 3.000 at k = −3.0; drops to 0.399 at k = 0.0; and finally to 0.000 at k = +3.0.

4.4 Inventory Management

The partial expectation with a left location parameter plays an important role in inventory management in deciding when to replenish the inventory on each item in stock and how much. It is used in the computations to determine the size of the safety stock, SS, for every item held. The safety stock is needed in the event the demand exceeds the forecast over the lead-time. The lead-time is the duration that begins when a replenish order is sent to the supplier and ends when the stock is received. The demand over the lead-time is assumed to follow a normal distribution. The percent-fill (PF), is measured for a duration called the order-cycle (OC). The OC is the time between two replenishments of stock. The percent-fill becomes:

PF = (demand filled in OC)/(total demand in OC)

The demand is the customer orders that arrive to be immediately filled by the stock currently on-hand. The management sets the PF desired and a mathematical method determines the order point, OP, and order level, OL needed to accomplish. This method requires the following data:

F = monthly forecasts.
L = lead-time (month)
σ = one month standard deviation of forecast error
Q = order quantity
PF = desired percent-fill

The forecast for the lead-time is F_L, and the associated standard deviation over the lead-time, σ_L, are obtained by,

$$F_L = L \times F$$

$$\sigma_L = \sqrt{L}\sigma$$

The safety stock is determined from a safety factor, k, as below:

$$SS = k\sigma_L$$

To find the safety factor, a time interval is needed that allows the computations to take place. The time duration covering the order cycle, OC, is selected, and the PF for this duration is set as the desired percent-fill. The order cycle is the time interval between two receipts of new stock. For this length of time, the percent-fill is:

$$\begin{aligned} PF &= \text{(demand filled in OC)/(total demand in OC)} \\ &= 1 - \text{(demand short in OC)/(total demand in OC)} \\ &= 1 - E(z > k)\sigma_L/Q \end{aligned}$$

Note, Q is the amount of replenish stock in the order cycle and represents the expected demand in the OC. $E(z > k)\sigma_L$ is the expected demand exceeding the OP during the order cycle, and therefore is a measure of the stock that is short in the OC. So now, the partial expectation becomes,

$$E(z > k) = (1 - PF)Q/\sigma_L$$

The safety factor, k, that corresponds to E(z > k) is obtained from Table 4.1. Note from the table where k > 0 only when E(z > k) < 0.40. To avoid a negative safety stock, k is set to zero when E(z > k) $\geq$ 0.40, where no safety stock is needed.

With k now obtained, the safety stock is computed by,

$$SS = k\sigma_L$$

The order point, OP, and order level, OL, become,

$$OP = F_L + SS$$

$$OL = OP + Q$$

Each day, the sum on-hand inventory (OH) and on-order inventory (OO) is compared to the OP and if (OH + OO) $\leq$ OP, a new buy quantity is needed and becomes,

$$buy = OL - (OH + OO)$$

In this way, the inventory replenishments for the item is controlled to yield the percent-fill, PF, desired by the management.

Example 4.1 Suppose a part where the management wants a percent fill of PF = 0.95. Assume a horizontal forecast applies and F = 10 per future month, the standard deviation of the one-month forecast error is σ = 5, the lead-time is L = 2 months, and the order quantity is Q = 20 pieces. For this situation, the following computations take place.

$$F_L = 2 \times 10 = 20$$

$$\sigma_L = \sqrt{2} \times 5 = 7.07$$

$$E(z > k) = (1 - 0.95)20/7.07 = 0.14$$

Table 4.1 shows: E(z > 0.7) ≈ 0.14

Thereby, the safety factor to use is k = 0.7, and the safety stock becomes:

$$SS = 0.7 \times 7.07 = 4.95 \approx 5$$

Finally, the order point and order level on this part are obtained as below:

$$OP = FL + SS = 20 + 5 = 25$$

$$OL = OP + Q = 25 + 20 = 45$$

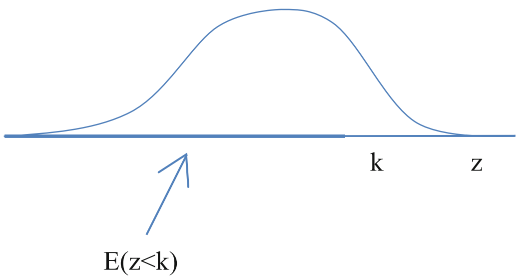

Fig. 4.2 The partial expectation E(z < k) with k and z from the standard normal distribution

4.5 Right Location Parameter

Suppose a scenario where the standard normal distribution applies and a location parameter, k, is specified and only the values of z smaller than k are observed. Hence (z < k) and of interest is the expected value of z smaller than k. In this way, k is the right-most value of z. The analysis here is called the partial expectation of (z < k). Below is the expectation related to the size of the difference of z smaller than k; and a depiction is in Fig. 4.2.

$$\mathrm{E}(\mathrm{z} < \mathrm{k}) = \int_{\infty}^{-k} (z-k)f(z)dz = -\mathrm{f}(\mathrm{k}) - \mathrm{kF}(\mathrm{k})$$

As before, f(k) is the probability density of the standard normal when $z = k$, and F(k) is the probability of z less than k.

Table 4.2 is a list of the values of E(z < k) for the set of location parameters, k, with the range: [−3.0, (0.1), +3.0]. Note where E(z < k) starts with 0.000 at $k = -3.0$; decreases to −0.399 at $k = 0.0$; and finally to −3.000 at $k = +3.0$.

Example 4.2 The daily water intake at a water supply system in an urban community during August is normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 10$. In the event the intake is low, a reserve supply of water is needed to satisfy the daily demands. A call for the use of a reserve is triggered should the daily intake fall to 90 or less. For this reason, the management of the system wants to know the average shortage when the supply is 90 or less.

To find this amount, the four steps listed below are followed:

1. The associated location parameter from the standard normal distribution, k, is obtained as shown here:

$$\mathrm{k} = (90 - \mu)/\sigma = (90 - 100)/10 = -1.0$$

2. Now, use Table 4.2 with $k = -1.0$ to find the corresponding partial expectation:

$$\mathrm{E}(\mathrm{z} < \mathrm{k}) = \mathrm{E}(\mathrm{z} < -1.0) = -0.083.$$

Table 4.2 Partial expectation E(z < k) sorted by location parameter k

k	E(z < k)	k	E(z < k)
−3.0	0.000	0.0	−0.399
−2.9	−0.001	0.1	−0.451
−2.8	−0.001	0.2	−0.507
−2.7	−0.001	0.3	−0.567
−2.6	−0.001	0.4	−0.630
−2.5	−0.002	0.5	−0.698
−2.4	−0.003	0.6	−0.769
−2.3	−0.004	0.7	−0.843
−2.2	−0.005	0.8	−0.920
−2.1	−0.006	0.9	−1.000
−2.0	−0.008	1.0	−1.083
−1.9	−0.011	1.1	−1.169
−1.8	−0.014	1.2	−1.256
−1.7	−0.018	1.3	−1.346
−1.6	−0.023	1.4	−1.437
−1.5	−0.029	1.5	−1.529
−1.4	−0.037	1.6	−1.623
−1.3	−0.046	1.7	−1.718
−1.2	−0.056	1.8	−1.814
−1.1	−0.069	1.9	−1.911
−1.0	−0.083	2.0	−2.008
−0.9	−0.100	2.1	−2.106
−0.8	−0.120	2.2	−2.205
−0.7	−0.143	2.3	−2.304
−0.6	−0.169	2.4	−2.403
−0.5	−0.198	2.5	−2.502
−0.4	−0.230	2.6	−2.601
−0.3	−0.267	2.7	−2.701
−0.2	−0.307	2.8	−2.801
−0.1	−0.351	2.9	−2.901
		3.0	−3.000

3. The corresponding partial expectation for the normal distribution becomes:

$$E(x < 90) = E(z < k)\sigma = -0.083 \times 10 = -0.83.$$

4. Hence, the average shortage when x is less than 90 is 0.83.

Example 4.3 A dealer stocks service parts for cars needing maintenance and repair. One of the parts has the following data: F = 10 is the average forecast per month;

$\sigma = 3$ is the standard deviation of F; L = 0.5 month is the lead-time duration; OP = 6 is the order point; and OL = 11 is the order level. To measure the percent fill for this part, the following eight steps are taken:

1. $F_L = L \times F = 5$
2. $\sigma_L = \sqrt{L}\,\sigma = 2.12$
3. $SS = OP - F_L = 1$
4. $k = SS/\sigma_L = 0.47$
5. Via Table 4.1, $E(z > k) \approx 0.21$
6. $E(\text{short in OC}) = E(z > k)\sigma_L = 0.445$
7. $Q = OL - OP = 5$
8. $PF = 1 - E(z > k)\sigma_L/Q = 0.91$

Example 4.4 For the part in Example 4.3, find the OP and OL that yields a percent-fill of PF = 0.95. To accomplish, the following five steps are taken:

1. $E(z > k) = (1 - PF)Q/\sigma_L = (1–0.95)5/2.12 = 0.118$
2. Via Table 4.1, $k \approx 0.8$
3. $SS = k\sigma_L = 1.70$
4. $OP = FL + SS = 6.7 \sim 7$
5. $OL = OP + Q = 12$

4.6 Advance Demand

Another application of the partial expectation in inventory management is the forecast adjustments due to advance demands. On some occasions a demand for an item is known one or more months prior to the date the customer wants possession of the item. This is known as an advance demand. For example, in April, the customer orders the quantity prior to its need in June. The demand is part of the regular demand in June and not an addition to it. The forecast for the future months are already generated from the flow of demands of the past, and with this extra advance demand knowledge, an adjustment to the forecast can be applied.

Assume the current month is t = N and $F(\tau)$ = forecast for the τ th future month with σ the corresponding standard deviation. Further, let x_o = advance demand for future month τ. Of need now is an adjustment to the forecast for the τ-th future month. It is assumed the demand for the future month comes from the normal distribution with parameters, mean = $F(\tau)$ and standard deviation = σ. Note where,

$$k = [x_o - F(\tau)]/\sigma$$

and the partial expectation of the standard normal where (z > k) is:

$$E(z > k)$$

The corresponding partial expectation for the demand, x, larger than x_o, is:

$$E(x > x_o) = E(z > k)\sigma$$

Hence, the forecast adjustment for future month τ becomes,

$$F_a(\tau) = x_o + E(z > k)\sigma$$

Example 4.5 Suppose a part where the forecast for future month τ is $F(\tau) = 90$ and the standard deviation is $\sigma = 30$. Assume the sum of advance demands is $x_o = 70$ pieces, and an adjustment to the forecast is needed. In this scenario,

$$\begin{aligned} k &= [70 - 90]/30 \\ &= -0.67 \end{aligned}$$

and with interpolation, Table 4.1 gives the partial expectation $E(z > -0.67) \approx 0.82$. Thereby, the adjusted forecast is,

$$\begin{aligned} F_a(\tau) &= x_o + E(z > -0.67)\ \sigma \\ &= 70 + 0.82 \times 30 \\ &= 94.6 \end{aligned}$$

Example 4.6 Consider the same part as Example 4.5, ($F(\tau) = 90$, $\sigma = 30$), but now assume the advance demand is $x_o = 30$. In this scenario,

$$\begin{aligned} k &= [30 - 90]/30 \\ &= -2.00 \end{aligned}$$

and from Table 4.1, $E(z > -2) = 2.008$. Thereby, the adjusted forecast is,

$$\begin{aligned} F_a(\tau) &= x_o + E(z > -2)\ \sigma \\ &= 30 + 2.008 \times 30 \\ &\approx 90.0 \end{aligned}$$

Note, there is no change in the forecast since the demand is on the low end of the normal range.

Example 4.7 Suppose once more the same part as Example 4.5, ($F(\tau) = 90$, $\sigma = 30$), where now the advance demand is $x_o = 180$ pieces. In this scenario,

$$\begin{aligned} k &= [180 - 100]/30 \\ &= 3.00 \end{aligned}$$

and Table 4.1 gives, $E(z > k) = 0.00$. The adjusted forecast now becomes,

$$\begin{aligned} F_a(\tau) &= x_o + E(z > k)\sigma \\ &= 180 + 0.00 \times 30 \\ &= 180.0 \end{aligned}$$

Note when the advance demand is three or more standard errors above the forecast, F(τ), the adjusted forecast is the same as the advance demand, x_o.

4.7 Summary

A particular value of z = k, from the standard normal, is selected as a left-location parameter and the average of the difference from all values of z larger than k is computed and called the partial expectation of z greater than k. The mathematical notation is $E(z > k)$, and table values of $E(z > k)$ are listed for $k = -3.0$ to $+3.0$. In the same way, a right location parameter, k, is selected and the difference from all z values smaller than k is computed and noted as $E(z < k)$. Table values for $E(z < k)$ are listed for $k = -3.0$ to $+3.0$. The partial expectation is used in inventory management to generate the safety stock for each item in the inventory, and also is used to adjust a forecast when advance demands become available for an item.

References

1. Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw Hill.
2. Brown, R. G. (1962). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs: Prentice Hall.
3. Thomopouos, N. T. (1980). *Applied forecasting methods*. Englewood Cliffs: Prentice Hall.

Chapter 5
Left Truncated Normal

5.1 Introduction

The left-truncated normal distribution, (LTN), takes on many shapes from normal to exponential-like. It has one parameter, k, the left-location parameter of the standard normal distribution. The shape of the distribution follows the normal for all standard normal z values larger than k, and thereby the distribution skews to the right. The statistical measures of this distribution are readily computed and include the mean, standard deviation, coefficient-of-variation, cumulative probability, and all percent-points needed. Three sets of table values are listed in the chapter allowing the user easy access and use to the distribution choice. One of the tables includes a range of percent-points denoted as tα where the probability of t less or equal to tα is α. The percent-points start with a value of zero and this is located at the left-location parameter, and since all subsequent values are larger, they thereby are all positive quantities. In addition to the tables, plots of the distribution are provided to observe the various shapes and relation to k. When sample data is available, the average, standard deviation and coefficient-of variation are easily computed, and the analyst applies these to identify the left-truncated normal distribution that best fits the sample data. This allows the analyst to estimate any probabilities needed on the sample data, without having to always assume the normal distribution. Examples are provided to help the user on the application of the distribution. In Chap. 7 (Truncated Normal Spread Ratio), another statistic is introduced that further aids the analyst to identify the type of distribution (left-truncated, right-truncated, normal) that best fits sample data, and also provides an estimate of the low limit for the left-truncated normal, and the high-limit for the right truncated normal. This chapter pertains to a left-truncated normal, while the next chapter describes the right-truncated normal.

In 1980, Thomopoulos, [1] developed tables on the left-truncated normal distribution and described how to apply them to generate the safety stock for items in the inventory. In 2001, Johnson studied the characteristics of the truncated normal

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_5

distribution, [2]; and in 2002, Johnson and Thomopoulos, [3], provide tables on the left-truncated normal. A new set of tables is generated in this chapter.

5.2 Left-Location Parameter

The left-truncated normal distribution derives from the standard normal distribution where a location parameter, k, is set and includes the z values above k. Hence, only $z > k$ are included in the distribution. For stability, a new variable, t, is called where $t = (z - k)$ and $t \geq 0$. The probability density of t is denoted as g(t) and the cumulative probability distribution is G(t), and both are obtained as below:

$$g(t) = f(z)/H(k)$$
$$G(t) = [F(z) - F(k)]/H(k)$$

where f(z) is the probability density at z for the standard normal, and H(k) is the complementary cumulative distribution of the standard normal when $z = k$.

5.3 Mathematical Equations

Some of the mathematical relationships pertaining to the left-truncated normal distribution are listed below:

$$f(z) = \left(1/\sqrt{2\pi}\right)e^{-z^2/2} \quad = \text{probability density of z}$$

$$F(k) = P(z \leq k) = \int_{-\infty}^{k} f(z)dz \quad = \text{cumulative probability of } z = k$$

$$H(k) = P(z > k) = 1 - F(k) \quad = \text{complementary probability of } z = k$$

$$\int_{k}^{\infty} zf(z)dz = f(k)$$

$$E(z > k) = \int_{k}^{\infty} (z - k)f(z)dz = f(k) - kH(k)$$

$$E\left[(z > k)^2\right] = \int_{k}^{\infty} (z - k)^2 f(z)dz = -kf(k) + H(k)\left(1 + k^2\right)$$

$$E(t)_k = E(z > k)/H(k) \quad = \text{expected value of t given k}$$
$$E(t^2)_k = E[(z > k)^2]/H(k) \quad = \text{expected value of } t^2 \text{ given k}$$
$$V(t)_k = E(t^2)_k - E(t)_k{}^2 \quad = \text{variance of t given k}$$

$\mu_t(k) = E(t)_k$ = mean of t given k
$\sigma_t(k) = \sqrt{V(t)_k}$ = standard deviation of t given k

Note since z is a continuous variable, $F(k) = P(z \leq k) = P(z < k)$.

5.4 Table Entries

Table 5.1 records the statistical values for the left-truncated normal variable, t, when the location parameter, k, ranges as [−3.0, (0.1), 3.0]. The statistical values listed are the mean, $\mu_t(k)$; standard deviation, $\sigma_t(k)$; and coefficient of variation, $cov_t(k)$. Note at k = −3.0, the coefficient of variation is near 0.33, signifying the distribution is like a normal distribution. At k = +3.0, the coefficient of variation is approaching 1.00, where the shape is like and exponential distribution. See Fig. 5.1 that depicts the shape of the probability density of t for various location parameters, k.

The way to find the Table 5.1 entries $\mu_t(k)$, $\sigma_t(k)$ and $cov_t(k)$ for a location parameter k is described below: Table 3.1 is searched to find f(k) and H(k) for the selected value of k. Next the partial expectations of (z > k) are computed as follows:

$$E(z > k) = f(k) - kH(k)$$
$$E\left[(z > k)^2\right] = H(k)\left(1 + k^2\right) - kf(k)$$

This now allows deriving the expected values for the left-truncated normal variable, t, as below:

$$E(t)_k = E(z > k)/H(k)$$
$$E(t^2)_k = E\left[(z > k)^2\right]/H(k)$$
$$V(t)_k = E(t^2)_k - E(t)_k{}^2$$

Thereby,

$$\mu_t(k) = E(t)_k$$
$$\sigma_t(k) = \sqrt{V(t)_k}$$
$$cov_t(k) = \sigma_t(k)/\mu_t(k)$$

Example 5.1 Note in Table 5.1, when the location parameter is k = 1.0: $\mu_t(k)$ 0.525, $\sigma_t(k) = 0.446$, and $cov_t(k) = 0.850$. The computations to obtain these results are shown in the four steps below:

1. Table 3.1 is called when k = 1.0 to find:

$$f(k) = 0.242$$
$$H(k) = 0.159.$$

Table 5.1 Left truncated normal distribution sorted by location parameter k; with mean $\mu_t(k)$; standard deviation $\sigma_t(k)$; and coefficient of variation $cov_t(k)$

k	$\mu_t(k)$	$\sigma_t(k)$	$cov_t(k)$
−3.0	3.004	0.993	0.331
−2.9	2.906	0.991	0.341
−2.8	2.808	0.989	0.352
−2.7	2.710	0.986	0.364
−2.6	2.614	0.982	0.376
−2.5	2.518	0.978	0.388
−2.4	2.423	0.972	0.401
−2.3	2.329	0.966	0.415
−2.2	2.236	0.959	0.429
−2.1	2.145	0.951	0.443
−2.0	2.055	0.942	0.458
−1.9	1.968	0.931	0.473
−1.8	1.882	0.920	0.489
−1.7	1.798	0.907	0.504
−1.6	1.717	0.894	0.520
−1.5	1.639	0.879	0.536
−1.4	1.563	0.863	0.552
−1.3	1.490	0.847	0.569
−1.2	1.419	0.830	0.585
−1.1	1.352	0.812	0.601
−1.0	1.288	0.794	0.616
−0.9	1.226	0.775	0.632
−0.8	1.168	0.756	0.647
−0.7	1.112	0.736	0.662
−0.6	1.059	0.717	0.677
−0.5	1.009	0.697	0.691
−0.4	0.962	0.678	0.705
−0.3	0.917	0.659	0.718
−0.2	0.875	0.640	0.731
−0.1	0.835	0.621	0.744
0.0	0.798	0.603	0.756
0.1	0.763	0.585	0.767
0.2	0.729	0.568	0.778
0.3	0.698	0.551	0.789
0.4	0.669	0.534	0.799
0.5	0.641	0.518	0.808
0.6	0.615	0.503	0.817
0.7	0.590	0.488	0.826
0.8	0.567	0.473	0.834
0.9	0.546	0.460	0.842
1.0	0.525	0.446	0.850
1.1	0.506	0.433	0.857

(continued)

Table 5.1 (continued)

k	$\mu_t(k)$	$\sigma_t(k)$	$cov_t(k)$
1.2	0.488	0.421	0.863
1.3	0.470	0.409	0.870
1.4	0.454	0.398	0.876
1.5	0.439	0.387	0.882
1.6	0.424	0.376	0.887
1.7	0.410	0.366	0.892
1.8	0.397	0.356	0.897
1.9	0.385	0.347	0.902
2.0	0.373	0.338	0.906
2.1	0.362	0.330	0.910
2.2	0.351	0.321	0.914
2.3	0.341	0.313	0.918
2.4	0.332	0.306	0.921
2.5	0.323	0.298	0.924
2.6	0.314	0.291	0.926
2.7	0.306	0.284	0.928
2.8	0.298	0.277	0.929
2.9	0.291	0.270	0.931
3.0	0.283	0.264	0.933

2. The partial expectations are now computed:

$$E(z > k) = f(k) - kH(k) \quad = 0.242 - 1.0 \times 0.159 = 0.083$$
$$E[(z > k)^2] = H(k)(1 + k^2) - kf(k) \quad = 0.159(1 + 1^2) - 1 \times 0.242 = 0.076$$

3. The expected values for variable t are the following:

$$E(t)_k = E(z > k)/H(k) \quad = 0.083/0.159 = 0.522$$
$$E(t^2)_k = E[(z > k)^2]/H(k) \quad = 0.076/0.159 = 0.478$$
$$V(t)_k = E(t^2)_k - E(t)_k^2 \quad = 0.478 - 0.522^2 = 0.206$$

4. Finally, the mean, standard deviation and coefficient of variation for t are:

$$\mu_t(k) = E(t)_k \quad = 0.522$$
$$\sigma_t(k) = \sqrt{V(t)_k} \quad = 0.453$$
$$cov_t(k) = \sigma_t(k)/\mu_t(k) \quad = 0.868$$

The difference in the table entries from the computations is due to rounding.

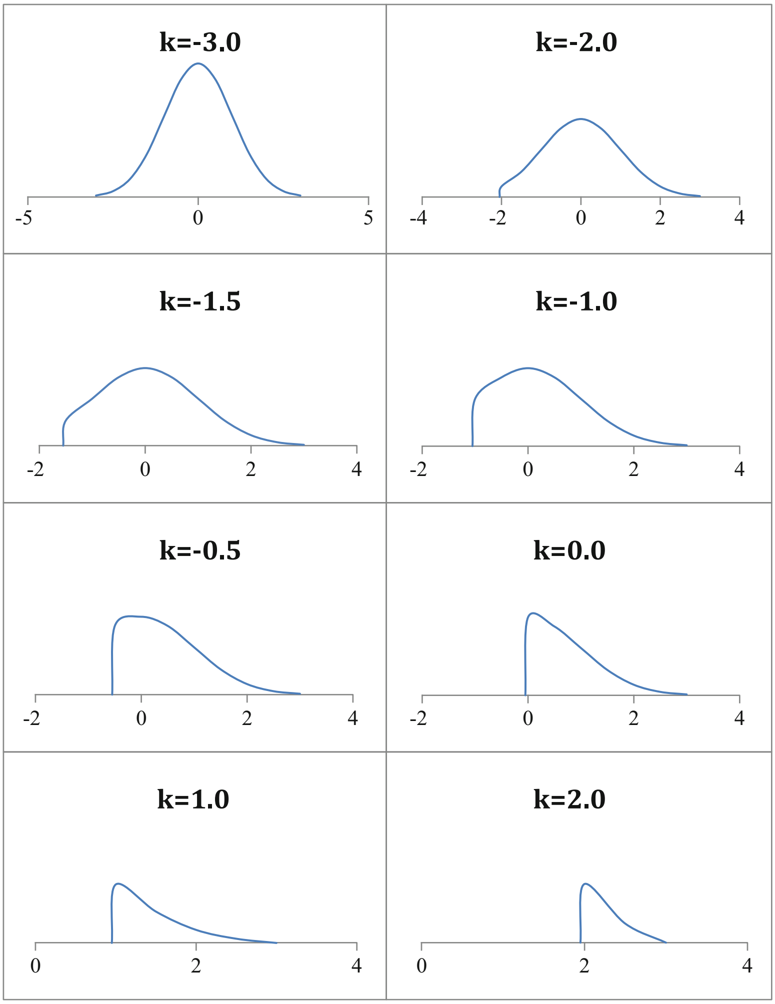

Fig. 5.1 Depiction of left-truncated normal when k = −3 to +2

5.5 More Tables

Table 5.2 contains the statistical values for left-truncated normal variable, t, when the cumulative probability of the location parameter k, F(k), ranges as [0.01, (0.01), 0.99]. The statistical values listed are the mean, $\mu_t(k)$; standard deviation, $\sigma_t(k)$; and coefficient of variation, $cov_t(k)$. Note at F(k) = 0.01, the coefficient of variation is near 0.411, signifying the distribution is similar to a normal distribution. At F (k) = 0.99, the coefficient of variation is approaching 1.00, where the shape is like an exponential distribution.

Table 5.2 Left truncated normal distribution sorted by standard normal cumulative probability F(k) of location parameter k; with mean $\mu_t(k)$; standard deviation $\sigma_t(k)$; and coefficient of variation $cov_t(k)$

F(k)	k	$\mu_t(k)$	$\sigma_t(k)$	$cov_t(k)$
0.01	−2.327	2.354	0.968	0.411
0.02	−2.054	2.104	0.947	0.450
0.03	−1.881	1.951	0.929	0.476
0.04	−1.751	1.841	0.914	0.496
0.05	−1.645	1.754	0.900	0.513
0.06	−1.555	1.682	0.887	0.527
0.07	−1.476	1.620	0.875	0.540
0.08	−1.405	1.567	0.864	0.552
0.09	−1.341	1.519	0.854	0.562
0.10	−1.282	1.477	0.844	0.571
0.11	−1.227	1.438	0.834	0.580
0.12	−1.175	1.402	0.825	0.589
0.13	−1.126	1.370	0.817	0.596
0.14	−1.080	1.339	0.808	0.604
0.15	−1.036	1.311	0.800	0.611
0.16	−0.994	1.284	0.792	0.617
0.17	−0.954	1.259	0.785	0.623
0.18	−0.915	1.235	0.778	0.629
0.19	−0.878	1.213	0.770	0.635
0.20	−0.841	1.191	0.764	0.641
0.21	−0.806	1.171	0.757	0.646
0.22	−0.772	1.152	0.750	0.651
0.23	−0.739	1.133	0.744	0.656
0.24	−0.706	1.115	0.737	0.661
0.25	−0.674	1.098	0.731	0.666
0.26	−0.643	1.081	0.725	0.671
0.27	−0.612	1.066	0.719	0.675
0.28	−0.582	1.050	0.713	0.679
0.29	−0.553	1.035	0.708	0.684
0.30	−0.524	1.021	0.702	0.688
0.31	−0.495	1.007	0.696	0.692
0.32	−0.467	0.993	0.691	0.696
0.33	−0.439	0.980	0.686	0.700
0.34	−0.412	0.967	0.680	0.703
0.35	−0.385	0.955	0.675	0.707
0.36	−0.358	0.943	0.670	0.711
0.37	−0.331	0.931	0.665	0.714
0.38	−0.305	0.919	0.660	0.718
0.39	−0.279	0.908	0.655	0.721
0.40	−0.253	0.897	0.650	0.725
0.41	−0.227	0.886	0.645	0.728

(continued)

Table 5.2 (continued)

F(k)	k	$\mu_t(k)$	$\sigma_t(k)$	$cov_t(k)$
0.42	−0.202	0.876	0.640	0.731
0.43	−0.176	0.865	0.635	0.734
0.44	−0.151	0.855	0.631	0.738
0.45	−0.125	0.845	0.626	0.741
0.46	−0.100	0.835	0.621	0.744
0.47	−0.075	0.826	0.617	0.747
0.48	−0.050	0.816	0.612	0.750
0.49	−0.025	0.807	0.607	0.753
0.50	0.000	0.798	0.603	0.756
0.51	0.025	0.789	0.598	0.758
0.52	0.050	0.780	0.594	0.761
0.53	0.075	0.771	0.589	0.764
0.54	0.100	0.763	0.585	0.767
0.55	0.125	0.754	0.580	0.769
0.56	0.151	0.746	0.576	0.772
0.57	0.176	0.737	0.571	0.775
0.58	0.202	0.729	0.567	0.777
0.59	0.227	0.721	0.562	0.780
0.60	0.253	0.713	0.558	0.782
0.61	0.279	0.705	0.553	0.785
0.62	0.305	0.697	0.549	0.788
0.63	0.331	0.689	0.545	0.790
0.64	0.358	0.681	0.540	0.793
0.65	0.385	0.674	0.536	0.795
0.66	0.412	0.666	0.531	0.798
0.67	0.439	0.658	0.527	0.801
0.68	0.467	0.650	0.522	0.803
0.69	0.495	0.643	0.518	0.806
0.70	0.524	0.635	0.513	0.808
0.71	0.553	0.628	0.509	0.811
0.72	0.582	0.620	0.504	0.814
0.73	0.612	0.612	0.500	0.816
0.74	0.643	0.605	0.495	0.819
0.75	0.674	0.597	0.491	0.822
0.76	0.706	0.590	0.486	0.824
0.77	0.739	0.582	0.481	0.827
0.78	0.772	0.574	0.477	0.830
0.79	0.806	0.566	0.472	0.833
0.80	0.841	0.559	0.467	0.836
0.81	0.878	0.551	0.462	0.839
0.82	0.915	0.543	0.457	0.842
0.83	0.954	0.535	0.452	0.845

(continued)

Table 5.2 (continued)

F(k)	k	$\mu_t(k)$	$\sigma_t(k)$	$cov_t(k)$
0.84	0.994	0.526	0.447	0.849
0.85	1.036	0.518	0.441	0.852
0.86	1.080	0.509	0.436	0.856
0.87	1.126	0.501	0.430	0.860
0.88	1.175	0.492	0.425	0.864
0.89	1.227	0.482	0.419	0.868
0.90	1.282	0.473	0.413	0.873
0.91	1.341	0.463	0.406	0.878
0.92	1.405	0.452	0.400	0.883
0.93	1.476	0.441	0.392	0.890
0.94	1.555	0.429	0.385	0.897
0.95	1.645	0.416	0.377	0.905
0.96	1.751	0.402	0.367	0.914
0.97	1.881	0.385	0.357	0.927
0.98	2.054	0.365	0.344	0.944
0.99	2.327	0.336	0.326	0.970

The way to find the Table 5.2 entries k, $\mu_t(k)$, $\sigma_t(k)$ and $cov_t(k)$ for the cumulative probability F(k) is described below: First, Table 3.2 is searched to find the value of k that is associated with F(k). Next, Table 3.1 yields f(k) and H(k) for the value of k. Then, the partial expectations of (z > k) are computed as follows:

$$E(z > k) = f(k) - kH(k)$$
$$E\left[(z > k)^2\right] = H(k)\left(1 + k^2\right) - kf(k)$$

Finally, the expected values for the left-truncated normal variable, t, are below:

$$E(t)_k = E(z > k)/H(k)$$
$$E(t^2)_k = E\left[(z > k)^2\right]/H(k)$$
$$V(t)_k = E(t^2)_k - E(t)_k{}^2$$

Thereby,

$$\mu_t(k) = E(t)_k$$
$$\sigma_t(k) = \sqrt{V(t)_k}$$
$$cov_t(k) = \sigma_t(k)/\mu_t(k)$$

Example 5.2 Note in Table 5.2, when the cumulative probability of the location parameter is F(k) = 0.40: $\mu_t(k) = 0.897$, $\sigma_t(k) = 0.650$, and $cov_t(k) = 0.725$. The computations to obtain these results are shown in the five steps below:

1. Table 3.2 is called when F(k) = 0.40 to find k = −0.253.

2. The values of f(k) and H(k) are obtained by:

$$f(k) = \left(1/\sqrt{2\pi}\right)e^{-k^2/2} = 0.386$$
$$H(k) = 1 - F(k) = 0.60$$

3. The partial expectations are now computed:

$E(z > k) = f(k) - kH(k)$	$= 0.386 + 0.253 \times 0.600 = 0.538.$
$E[(z > k)^2] = H(k)(1 + k^2) - kf(k)$	$= 0.60[1 + (-0.253)^2] + 0.253 \times 0.386 = 0.736$

4. The expected values for variable t are the following:

$E(t)_k = E(z > k)/H(k)$	$= 0.538/0.60 = 0.897$
$E(t^2)_k = E[(z > k)^2]/H(k)$	$= 0.736/0.60 = 1.227$
$V(t)_k = E(t^2)_k - E(t)_k{}^2$	$= 1.227 - 0.897^2 = 0.424$

5. Finally, the mean, standard deviation and coefficient of variation for t are:

$\mu_t(k) = E(t)_k$	$= 0.897$
$\sigma_t(k) = \sqrt{V(t)_k}$	$= 0.651$
$cov_t(k) = \sigma_t(k)/\mu_t(k)$	$= 0.726$

The difference in the table results from the computations is due to rounding.

5.6 Left Truncated Distribution

Table 5.3 gives selected values of the t variable for the left-truncated normal distribution when the location parameter, k, ranges as: [−3.0, (0.1), 3.0], and the cumulative probability distribution, G(t) spans from 0.01 to 0.99. Recall, t begins at zero and is computed as shown below.

Given k and G(t), then:

$$G(t) = [F(z) - F(k)]/H(k)$$
$$[1 - G(t)] = H(z)/H(k),$$

where Table 3.1 yields H(k).

Hence,

$$H(z) = [1 - G(t)]H(k)$$

and Table 3.2 gives the associated value of z. Finally,

$$t = (z - k)$$

is obtained.

Table 5.3 Left truncated normal distribution sorted by location parameter k; listing variable t for cumulative probabilities G(t) from 0.01 to 0.99

k\G(t)	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
–3.0	0.72	1.37	1.73	2.16	2.48	2.75	3.00	3.25	3.53	3.84	4.28	4.65	5.33
–2.9	0.64	1.27	1.63	2.06	2.38	2.65	2.90	3.15	3.43	3.74	4.18	4.55	5.23
–2.8	0.56	1.18	1.53	1.97	2.28	2.55	2.80	3.06	3.33	3.64	4.08	4.45	5.13
–2.7	0.49	1.09	1.44	1.87	2.18	2.45	2.70	2.96	3.23	3.54	3.98	4.35	5.03
–2.6	0.42	1.00	1.34	1.77	2.09	2.35	2.61	2.86	3.13	3.44	3.88	4.25	4.93
–2.5	0.36	0.91	1.25	1.68	1.99	2.26	2.51	2.76	3.03	3.35	3.79	4.15	4.83
–2.4	0.31	0.83	1.16	1.58	1.89	2.16	2.41	2.66	2.93	3.25	3.69	4.05	4.73
–2.3	0.26	0.75	1.07	1.49	1.80	2.06	2.31	2.56	2.83	3.15	3.59	3.95	4.63
–2.2	0.22	0.67	0.99	1.40	1.70	1.97	2.22	2.47	2.74	3.05	3.49	3.85	4.53
–2.1	0.18	0.60	0.91	1.31	1.61	1.87	2.12	2.37	2.64	2.95	3.39	3.75	4.43
–2.0	0.15	0.54	0.83	1.22	1.52	1.78	2.03	2.28	2.54	2.86	3.29	3.66	4.34
–1.9	0.13	0.48	0.75	1.14	1.43	1.69	1.94	2.18	2.45	2.76	3.20	3.56	4.24
–1.8	0.11	0.42	0.68	1.06	1.35	1.60	1.84	2.09	2.36	2.67	3.10	3.46	4.14
–1.7	0.09	0.37	0.62	0.98	1.26	1.52	1.76	2.00	2.26	2.57	3.01	3.37	4.04
–1.6	0.08	0.33	0.56	0.91	1.18	1.43	1.67	1.91	2.17	2.48	2.91	3.27	3.95
–1.5	0.07	0.29	0.51	0.84	1.11	1.35	1.58	1.82	2.08	2.39	2.82	3.18	3.85
–1.4	0.06	0.26	0.46	0.77	1.03	1.27	1.50	1.74	2.00	2.30	2.73	3.09	3.76
–1.3	0.05	0.23	0.41	0.71	0.96	1.19	1.42	1.65	1.91	2.21	2.64	2.99	3.66
–1.2	0.04	0.20	0.37	0.65	0.90	1.12	1.34	1.57	1.83	2.13	2.55	2.90	3.57
–1.1	0.04	0.18	0.34	0.60	0.83	1.05	1.27	1.50	1.75	2.04	2.46	2.81	3.48
–1.0	0.03	0.16	0.30	0.55	0.78	0.99	1.20	1.42	1.67	1.96	2.38	2.73	3.39

(continued)

Table 5.3 (continued)

k\G(t)	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
−0.9	0.03	0.14	0.27	0.51	0.72	0.93	1.13	1.35	1.59	1.88	2.29	2.64	3.30
−0.8	0.03	0.13	0.25	0.47	0.67	0.87	1.07	1.28	1.52	1.80	2.21	2.56	3.21
−0.7	0.02	0.12	0.23	0.43	0.62	0.81	1.01	1.21	1.45	1.73	2.13	2.48	3.13
−0.6	0.02	0.11	0.21	0.40	0.58	0.76	0.95	1.15	1.38	1.66	2.06	2.40	3.04
−0.5	0.02	0.10	0.19	0.37	0.54	0.71	0.90	1.09	1.32	1.59	1.98	2.32	2.96
−0.4	0.02	0.09	0.17	0.34	0.50	0.67	0.85	1.04	1.25	1.52	1.91	2.24	2.88
−0.3	0.02	0.08	0.16	0.31	0.47	0.63	0.80	0.98	1.19	1.46	1.84	2.17	2.80
−0.2	0.02	0.07	0.15	0.29	0.44	0.59	0.75	0.93	1.14	1.40	1.77	2.10	2.72
−0.1	0.01	0.07	0.14	0.27	0.41	0.56	0.71	0.89	1.09	1.34	1.71	2.03	2.65
0.0	0.01	0.06	0.13	0.25	0.38	0.52	0.67	0.84	1.04	1.28	1.65	1.96	2.58
0.1	0.01	0.06	0.12	0.24	0.36	0.49	0.64	0.80	0.99	1.23	1.59	1.90	2.50
0.2	0.01	0.05	0.11	0.22	0.34	0.47	0.60	0.76	0.94	1.18	1.53	1.83	2.43
0.3	0.01	0.05	0.10	0.21	0.32	0.44	0.57	0.72	0.90	1.13	1.47	1.77	2.37
0.4	0.01	0.05	0.10	0.20	0.30	0.42	0.55	0.69	0.86	1.08	1.42	1.71	2.30
0.5	0.01	0.04	0.09	0.18	0.29	0.40	0.52	0.66	0.83	1.04	1.37	1.66	2.24
0.6	0.01	0.04	0.08	0.17	0.27	0.38	0.49	0.63	0.79	1.00	1.32	1.61	2.18
0.7	0.01	0.04	0.08	0.16	0.26	0.36	0.47	0.60	0.76	0.96	1.27	1.55	2.12
0.8	0.01	0.04	0.08	0.16	0.24	0.34	0.45	0.57	0.73	0.92	1.23	1.50	2.06
0.9	0.01	0.04	0.07	0.15	0.23	0.32	0.43	0.55	0.70	0.89	1.19	1.46	2.00
1.0	0.01	0.03	0.07	0.14	0.22	0.31	0.41	0.53	0.67	0.86	1.15	1.41	1.95
1.1	0.01	0.03	0.06	0.13	0.21	0.30	0.39	0.51	0.64	0.82	1.11	1.37	1.90

1.2	0.01	0.03	0.06	0.13	0.20	0.28	0.38	0.48	0.62	0.80	1.07	1.33	1.85
1.3	0.01	0.03	0.06	0.12	0.19	0.27	0.36	0.47	0.60	0.77	1.04	1.29	1.80
1.4	0.01	0.03	0.06	0.12	0.18	0.26	0.35	0.45	0.57	0.74	1.01	1.25	1.75
1.5	0.01	0.03	0.05	0.11	0.18	0.25	0.33	0.43	0.55	0.72	0.97	1.21	1.70
1.6	0.01	0.03	0.05	0.11	0.17	0.24	0.32	0.42	0.53	0.69	0.94	1.18	1.66
1.7	0.01	0.02	0.05	0.10	0.16	0.23	0.31	0.40	0.52	0.67	0.92	1.14	1.62
1.8	0.00	0.02	0.05	0.10	0.16	0.22	0.30	0.39	0.50	0.65	0.89	1.11	1.58
1.9	0.00	0.02	0.05	0.10	0.15	0.21	0.29	0.37	0.48	0.63	0.86	1.08	1.53
2.0	0.00	0.02	0.04	0.09	0.15	0.21	0.28	0.36	0.47	0.61	0.84	1.05	1.49
2.1	0.00	0.02	0.04	0.09	0.14	0.20	0.27	0.35	0.45	0.59	0.81	1.02	1.46
2.2	0.00	0.02	0.04	0.09	0.14	0.19	0.26	0.34	0.44	0.57	0.79	0.99	1.42
2.3	0.00	0.02	0.04	0.08	0.13	0.19	0.25	0.33	0.42	0.55	0.77	0.97	1.38
2.4	0.00	0.02	0.04	0.08	0.13	0.18	0.24	0.32	0.41	0.54	0.75	0.94	1.34
2.5	0.00	0.02	0.04	0.08	0.12	0.18	0.24	0.31	0.40	0.52	0.72	0.91	1.30
2.6	0.00	0.02	0.04	0.08	0.12	0.17	0.23	0.30	0.39	0.51	0.70	0.89	1.26
2.7	0.00	0.02	0.03	0.07	0.12	0.16	0.22	0.29	0.38	0.49	0.68	0.86	1.22
2.8	0.00	0.02	0.03	0.07	0.11	0.16	0.21	0.28	0.36	0.48	0.66	0.84	1.17
2.9	0.00	0.01	0.03	0.07	0.11	0.15	0.21	0.27	0.35	0.46	0.64	0.81	1.12
3.0	0.00	0.01	0.03	0.06	0.10	0.15	0.20	0.26	0.34	0.45	0.62	0.78	1.07

Example 5.3 Table 5.3 shows when the location parameter is $k = -1.0$ and the cumulative distribution of t is $G(t) = 0.50$, the value of the left-truncated variable is $t = 1.20$. The four steps below show how this value is obtained:

1. Table 3.1 yields $H(k) = H(-1.0) = 0.841$
2. Using, $H(z) = [1 - G(t)]H(k) = [1\text{–}0.50] \times 0.841 = 0.420$
3. Table 3.2 shows $H(0.202) = 0.420$, thereby, $z = 0.202$
4. $t = (z - k) = [0.202 - (-1.0)] = 1.20$

5.7 Application to Sample Data

To apply the LTN Tables 5.1 and 5.3, to sample data, an estimate of the sample coefficient-of-variation, cov, is needed. The cov must be computed at a point where the low limit, γ, is zero. Hence, the cov estimate is obtained as below:

if $\gamma = 0$: $\quad cov = s/\bar{x}$
if $\gamma \neq 0$: $\quad cov = s/(\bar{x} - \gamma)$

A way to estimate γ is provided in Chap. 7 (Truncated Normal Spread Ratio).

The sample data includes the following:
$\bar{x}$= sample mean
s = sample standard deviation
γ = low-limit
$cov = s/(\bar{x} - \gamma)$ = adjusted coefficient of variation

The table statistics includes the following:

k = left-location parameter,
$\mu_t(k)$ = mean of t at k
$\sigma_t(k)$ = standard deviation of t at k

The percent-point conversions concerning variables x and t are as follows:

$$t\alpha = \mu_t(k) + \sigma_t(k)\left[\left(x\alpha - \bar{x}\right)/s\right]$$

$$x\alpha = \bar{x} + s\left[\left(t\alpha - \mu_t(k)\right)/\sigma_t(k)\right]$$

where:

$t\alpha$ = α-percent-point of variable t
$x\alpha$ = α-percent-point of variable x

Example 5.4 Consider sample data where the variable, x, is zero or larger ($\gamma = 0$); the mean is $\bar{x} = 50$, and the standard deviation is $s = 30$. Assume the analyst is seeking the probability for $x \leq 20$. Because the coefficient-of-variation is $cov = 30/50 = 0.60$, the left-truncated normal distribution applies. To find $P(x \leq 20)$, the following four steps apply.

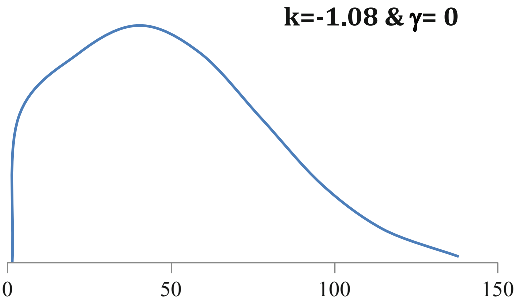

Fig. 5.2 Plot depicting distribution from Example 5.4

1. Table 5.2 is searched to locate the nearest left-location parameter, k, that has cov = 0.60, and this is k = −1.08. The associated mean and standard deviation of t are:

$$\mu_t(k) = 1.339$$
$$\sigma_t(k) = 0.808$$

2. The value, tα, that corresponds to xα = 20 is obtained as below:

$$t\alpha = \mu_t(k) + \sigma_t(k)\ \left[\left(x\alpha - \bar{x}\right)/s\right]$$
$$= 1.339 + 0.808[(20 - 50)/30] = 0.531$$

3. Table 5.3 is searched for the closest value of k = −1.08, (−1.1), and t = 0.531 (0.60) and finds the associate measure of G(t) = 0.20.
4. With interpolation, P(t ≤ 0.531) ≈ 0.17.
   Hence, P(x ≤ 20) ≈ 0.17.

   See Fig. 5.2.

Example 5.5 Consider the sample data of Example 5.4 again where $\bar{x} = 50$, s = 30, cov = 0.60, and recall, x ≥ 0. Suppose the analyst wants to find the median value of x, where $P(x \leq x_{0.50}) = 0.50$. To obtain, the following three steps are taken.

1. Table 5.2 is searched to locate the nearest left-location parameter, k, that has cov = 0.60. This is k = −1.08, and the variable is denoted as t. The associated mean and standard deviation of t are:

$$\mu_t(k) = 1.339$$
$$\sigma_t(k) = 0.808$$

2. Table 5.3 is searched with the closest k, (−1.1) and G(t) = 0.50 to find:

$$t_{0.50} = 1.27$$

3. The corresponding value of $x_{0.50}$ is now found as follows:

$$x\alpha = \bar{x} + s\ \left[(t\alpha - \mu_t(k))/\sigma_t(k) \qquad = 50 + 30[1.27 - 1.339)/0.808] = 47.4$$

   Hence, P(x ≤ 47.4) ≈ 0.50.

Example 5.6 Suppose sample data from a left-truncated normal distribution yields an average of 50, standard deviation 20, and the data has a low-limit of $\gamma = 10$. Of need here is to find the 90 percent-point of the data. The vital stats of the sample are as follows:

$$\gamma = 10$$
$$\bar{x} = 50$$
$$s = 20$$

To find $x_{0.90}$, the following four steps are followed:

1. The adjusted cov is computed as below:

$$cov = s/(\bar{x} - \gamma) = 20/(50 - 10) = 0.50$$

2. Table 5.1 is searched to locate the nearest left-location parameter, k, that has cov = 0.50, and this is k = −1.70. The associated mean and standard deviation of t are:

$$\mu_t(k) = 1.798$$
$$\sigma_t(k) = 0.907$$

3. Table 5.3 is searched with the closest k, (−1.7) and G(t) = 0.90 to find:

$$t_{0.90} = 3.01$$

4. The corresponding value of $x_{0.90}$ is found as follows:

$$x\alpha = \bar{x} + s[(t\alpha - \mu_t(k))/\sigma_t(k)$$
$$= 50 + 20[(3.01 - 1.798))/0.907] = 76.70$$

See Fig. 5.3 that depicts the distribution.

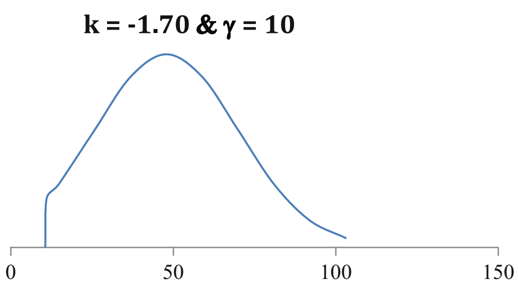

Fig. 5.3 Plot depicting distribution from Example 5.6

5.8 LTN for Inventory Control

The key function in inventory management is to determine when to replenish stock and how much for every item in each stock-holding location. To do this efficiently, a forecast, F, of every item is revised each month; along with an estimate of the standard deviation, s, of the forecast error. The coefficient of variation, cov, for each item is measured as cov = s/F. The mathematical methods that control the inventory are based on the forecast and the standard deviation, and mostly assumes the demand variation is normally distributed. But, with the author's experience, most of the items in the stock holding locations (distributions centers, dealers, stores) are not normally distributed; they are distributed as the left-truncated normal distribution with the low limit of $\gamma = 0$. Examples for an automotive service part distribution center and for a retailer are provided.

Automotive Service Parts Distribution Center

Below are statistics on the percent of parts by cov (cov, % parts) from a service parts automotive distribution center with over 100,000 part numbers and one-billion dollars in annual sales. The forecasts are revised each month and the replenish needs are determined every day. The parts with cov of 0.50 or less (38%) are the more high demand parts; while the parts with cov of 0.50 and higher (62%) are mainly the lower demand parts. For the parts with the higher cov, it is important to use the left-truncated normal with $\gamma = 0$, rather than the normal to compute the inventory need; otherwise, the system will not yield the level of service that is planned.

Cov	% parts
0.00–0.30	26
0.30–0.50	12
0.50–0.80	12
0.80–1.00	10
1.00 -	40
sum	100

Retail Products

Most items in a retail store are of the low demand type, as described in Thomopoulos, [4]. These are items that have an average monthly demand of two or less. Could be pricey items like: televisions, refrigerators, power tools, furniture, mattresses, silverware, china, with many models in competition. Other retail items also have many competitive models, like toasters, razors, toothpaste, hair cream, and lamps. The more models in competition, the less demand per individual model. Many other items are by style with a variety of sizes, like: shoes, sweaters, shirts, trousers, dresses, coats, and so on. Each size of the style is a stock-keeping-unit, (sku). In most scenarios, the demand for an individual sku is small and most are less than one per month. The monthly demands for low demand items are not normally distributed; instead, the Poisson or the left-truncated normal, with the low limit set at zero, should be used in the inventory control computations.

To illustrate why the demands are quite low in an inventory store, two examples from the shoe industry are sited. Consider a large shoe store with 500 styles and an average of 20 sizes per style, representing 10,000 skus. Assume for a normal month, the store sells 2000 pair. The average demand per sku is thereby 0.20 pair. For a more moderate store, the number of styles is 300 and the average number of sizes per style is 15. In this scenario, the store carries 4500 skus. If the average number of sales per month is 1500 pair, the average demand per sku is 0.33 pair.

5.9 Summary

A particular value of the standard normal, z = k, is selected as a left location parameter and all values of z larger than k are allowed in the left-truncated normal distribution. The mathematical equations for the mean, standard deviation, coefficient-of-variation, cumulative probability, and a variety of percent-points are developed. Table values for k ranging from −3.0 to +3.0 are listed. Another table of F (k) ranging for 0.01 to 0.99 is also listed. A third table lists a variety of percent-points (0.01 to 0.99) as k flows from −3.0 to +3.0. When an analyst has sample data and the coefficient of variation is computed, the analyst can estimate which value of k best fits the data; and thereby with the tables can compute all the probabilities needed for the sample data, without the need to always assume the data is normally distributed.

References

1. Thomopoulos, N. T. (1980). *Applied forecasting methods* (pp. 318–324). Englewood Cliffs: Prentice Hall.
2. Johnson, A. C. (2001). *On the truncated normal distribution*. Doctoral Dissertation. Stuart School of Business: Illinois Institute of Technology.
3. Johnson A. C., & Thomopoulos, N. T. (2002). Characteristics and tables of the left-truncated normal distribution. In *Proceedings of the Midwest Decision Sciences Institute* (pp. 133–139).
4. Thomopoulos, N. T. (2016). *Demand forecasting for inventory control*. New York: Springer.

Chapter 6
Right Truncated Normal

6.1 Introduction

The right-truncated normal distribution (RTN) takes on many shapes, from a normal to an inverted exponential-like. The distribution has one parameter, k, called the right-location parameter of the standard normal distribution. The distribution includes all the standard normal values smaller than k, and thereby the density skews towards the left-tail. The important statistics are readily computed and are the following: mean, standard deviation, coefficient-of-variation, cumulative probability, and a variety of percent-points. Table values are provided to allow the analyst to apply the distribution to sample data. Also included is a series of plots that show the analyst how the density is shaped with respect to the right-location parameter. The percent-points are all negative values since the distribution starts with zero at the right-truncated location parameter, and subsequently moves downward to the left. Hence, the average and coefficient-of-variation are also negative values. When sample data is available, the average, standard deviation and coefficient-of variation are easily computed, and the analyst can apply these to identify the right-truncated normal distribution that best fits the sample data. This also allows the analyst to estimate any probabilities needed on the sample data, without having to always assume the normal distribution. Examples are provided to help the user on the application of the distribution. In Chap. 7, (Truncated Normal Spread Ratio), another statistic is introduced and allows the analyst to easily identify the type of distribution (left-truncated, right-truncated, normal) that best fits sample data. This chapter pertains to a right-truncated normal distribution; while the prior chapter describes the left-truncated normal. Recall from Chap. 1, (Continuous Distributions), only the beta distribution offers a choice of shapes that skew to the left, and thereby this right-truncated normal distribution may well be a welcome alternative. Furthermore, the percent-points of the right-truncated normal are far easier to compute that those of the beta distribution.

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_6

In 2001, Johnson [1] studied the left and right truncated normal distributions. Also in 2001, Johnson and Thomopoulos [2] generated tables on the right-truncated normal distribution. A complete new set of tables is produced in this chapter.

6.2 Right Truncated Distribution

The right-truncated normal distribution derives from the standard normal distribution where a location parameter, k, is set and only the z values of the standard normal below k are allowed. Hence, only z < k are included in the distribution. For stability, a new variable, t, is called where t = (z – k) and t ≤ 0. So the density of t is denoted as g(t) and the cumulative distribution is G(t). Note the following:

$$g(t) = f(z)/F(k)$$
$$G(t) = F(z)/F(k)$$

where f(z) is the probability density at z for the standard normal, F(z) = cumulative probability of z, and F(k) is the cumulative probability of the standard normal when z = k.

6.3 Mathematical Equations

Some of the mathematical relationships pertaining to the right-truncated normal distribution are listed below:

$f(z) = \left(1/\sqrt{2\pi}\right)e^{-z^2/2}$ = probability density of z

$F(k) = P(z \le k) = \int_{-\infty}^{k} f(z)dz$ = cumulative probability of z = k

$H(k) = P(z > k) = 1 - F(k)$ = complementary probability of z = k

Some related integrals are below:

$$\int_{-\infty}^{k} zf(z)dz = -f(k)$$
$$\int_{-\infty}^{k} z^2 f(z)dz = -kf(k) + F(k)$$

Some relations on the partial expectation of (z < k)

$$E(z < k) = \int_{-\infty}^{k} (z-k)f(z)dz = -f(k) - kF(k)$$

$$E\left[(z<k)^2\right] = \int_{-\infty}^{k} (z-k)^2 f(z)dz = kf(k) + F(k)\left(1+k^2\right)$$

6.4 Variable t Range

Below are some relations on the variable t = (z – k) where z < k and t is negative.

$E(t)_k = E(z < k)/F(k)$ = expected value of t at k
$E(t^2)_k = E[(z < k)^2]/F(k)$ = expected value of t^2 at k
$V(t)_k = E(t^2)_k - E(t)_k^2$ = variance of t at k
$\mu_t(k) = E(t)_k$ = mean of t at k
$\sigma_t(k) = \sqrt{V(t)_k}$ = standard deviation of t at k

For this analysis, the limits on z are ±3, thereby the high and low limits on t are: [0,–(3 + k)]. Some samples on the range of t as related to k follow (Fig. 6.1):

At k = 3: t = (0,–6)
At k = 2: t = (0.–5)
At k = 1: t = (0,–4)
At k = 0: t = (0,–3)
At k = −1: t = (0,–2)
and so on.

6.5 Table Entries

Table 6.1 contains a list of statistical measures from the right-truncated normal distribution with right-location parameter, k, ranging as: [−3.0, (0.1), 3.0]. The random variable is denoted as t where t ranges from –(3 + k) to 0. For each k, the table gives the values of $\mu_t(k)$, $\sigma_t(k)$, and $cov_t(k)$.

The four steps below describe how the statistical measures are obtained for each value of k.

1. For a given k, Table 3.1 is called to find:

$$f(k) \text{ and } F(k).$$

2. Now compute the partial expectations:

$$E(z<k) = -f(k) - kF(k)$$
$$E\left[(z<k)^2\right] = kf(k) + F(k)\left(1+k^2\right)$$

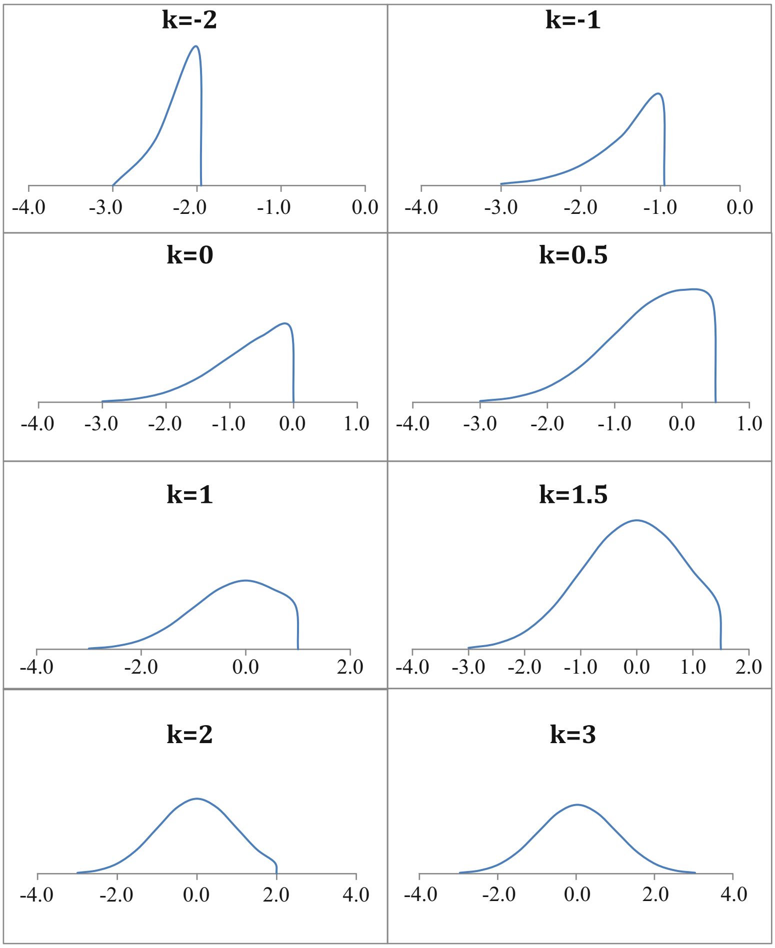

Fig. 6.1 Depiction of right-truncated normal for various values of k

3. For t = (z – k) with z < k, calculate the expectations of t:

$$E(t)_k = E(z < k)/F(k)$$
$$E(t^2)_k = E\left[(z < k)^2\right]/F(k)$$
$$V(t)_k = E(t^2)_k - E(t)_k{}^2$$

4. The mean, standard deviation and coefficient-of-variation of t are obtained as below:

$$\mu_t(k) = E(t)_k$$
$$\sigma_t(k) = \sqrt{V(t)k}$$

Table 6.1 Variable t of right truncated normal distribution sorted by location parameter k; with mean $\mu_t(k)$; standard deviation $\sigma_t(k)$; and the coefficient of variation cov

k	$\mu_t(k)$	$\sigma_t(k)$	cov	k	$\mu_t(k)$	$\sigma_t(k)$	cov
−3.0	−0.283	0.264	−0.933	0.0	−0.798	0.603	−0.756
−2.9	−0.291	0.270	−0.931	0.1	−0.835	0.621	−0.744
−2.8	−0.298	0.277	−0.929	0.2	−0.875	0.640	−0.731
−2.7	−0.306	0.284	−0.928	0.3	−0.917	0.659	−0.718
−2.6	−0.314	0.291	−0.926	0.4	−0.962	0.678	−0.705
−2.5	−0.323	0.298	−0.924	0.5	−1.009	0.697	−0.691
−2.4	−0.332	0.306	−0.921	0.6	−1.059	0.717	−0.677
−2.3	−0.341	0.313	−0.918	0.7	−1.112	0.736	−0.662
−2.2	−0.351	0.321	−0.914	0.8	−1.168	0.756	−0.647
−2.1	−0.362	0.330	−0.910	0.9	−1.226	0.775	−0.632
−2.0	−0.373	0.338	−0.906	1.0	−1.288	0.794	−0.616
−1.9	−0.385	0.347	−0.902	1.1	−1.352	0.812	−0.601
−1.8	−0.397	0.356	−0.897	1.2	−1.419	0.830	−0.585
−1.7	−0.410	0.366	−0.892	1.3	−1.490	0.847	−0.569
−1.6	−0.424	0.376	−0.887	1.4	−1.563	0.863	−0.552
−1.5	−0.439	0.387	−0.882	1.5	−1.639	0.879	−0.536
−1.4	−0.454	0.398	−0.876	1.6	−1.717	0.894	−0.520
−1.3	−0.470	0.409	−0.870	1.7	−1.798	0.907	−0.504
−1.2	−0.488	0.421	−0.863	1.8	−1.882	0.920	−0.489
−1.1	−0.506	0.433	−0.857	1.9	−1.968	0.931	−0.473
−1.0	−0.525	0.446	−0.850	2.0	−2.055	0.942	−0.458
−0.9	−0.546	0.460	−0.842	2.1	−2.145	0.951	−0.443
−0.8	−0.567	0.473	−0.834	2.2	−2.236	0.959	−0.429
−0.7	−0.590	0.488	−0.826	2.3	−2.329	0.966	−0.415
−0.6	−0.615	0.503	−0.817	2.4	−2.423	0.972	−0.401
−0.5	−0.641	0.518	−0.808	2.5	−2.518	0.978	−0.388
−0.4	−0.669	0.534	−0.799	2.6	−2.614	0.982	−0.376
−0.3	−0.698	0.551	−0.789	2.7	−2.710	0.986	−0.364
−0.2	−0.729	0.568	−0.778	2.8	−2.808	0.989	−0.352
−0.1	−0.763	0.585	−0.767	2.9	−2.906	0.991	−0.341
0.0	−0.798	0.603	−0.756	3.0	−3.004	0.993	−0.331

$$cov_t(k) = \sigma_t(k)/\mu_t(k)$$

Example 6.1 Note from Table 6.1, when k = 1.0: $\mu_t(k) = -1.288$, $\sigma_t(k) = 0.794$, and $cov_t(k) = -0.616$. Below shows how these results are obtained.

1. From Table 3.1 and k = 1.0,

$$f(1) = 0.242$$
$$F(1) = 0.841$$

2. The partial expectations of (z < k) become:

$$E(z < k) = -0.242 - 1.0 \times 0.841 = -1.083$$
$$E\left[(z < k)^2\right] = 1.0 \times 0.242 + 0.841\left(1 + 1^2\right) = 1.924$$

3. For t = (z-k), the expected values are below:

$$E(t)_k = -1.083/0.841 = -1.288$$
$$E(t^2)_k = 1.924/0.841 = 2.288$$
$$V(t)_k = 2.288 - (-1.288)^2 = 0.629$$

4. The mean, variance, and coefficient-of-variation of t are now computed:

$$\mu_t(k) = -1.288$$
$$\sigma_t(k) = 0.629^{0.5} = 0.791$$
$$cov = 0.791/-1.288 = -0.614$$

Any difference from the table values and the above are due to rounding.

6.6 More Tables

Table 6.2 contains selected values of the t variable for the right-truncated normal distribution when the location parameter, k, ranges as: [−3.0. (0.1), 3.0], and the cumulative distribution, G(t) spans from 0.01 to 0.99. Recall, t begins at –(3 + k) and is computed as shown below.

The four steps that follow describe how the percent-point values of t are obtained for each combination of k and G(t).

1. Table 3.2 gives the value of F(k) associated with k.
2. The cumulative distribution of t is:

$$G(t) = F(z)/F(k)$$

and thereby,

$$F(z) = G(t) \times F(k)$$

Table 6.2 Right truncated normal distribution sorted by location parameter k; listing variable t for cumulative probabilities G(t) from 0.01 to 0.99

k\G(t)	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
−3.0	−1.49	−0.86	−0.66	−0.47	−0.36	−0.27	−0.21	−0.16	−0.11	−0.07	−0.03	−0.02	−0.01
−2.9	−1.40	−0.87	−0.67	−0.48	−0.36	−0.28	−0.21	−0.16	−0.11	−0.07	−0.04	−0.02	−0.01
−2.8	−1.37	−0.88	−0.69	−0.49	−0.37	−0.29	−0.22	−0.16	−0.12	−0.07	−0.04	−0.02	0.00
−2.7	−1.36	−0.89	−0.70	−0.50	−0.38	−0.29	−0.23	−0.17	−0.12	−0.07	−0.04	−0.02	0.00
−2.6	−1.37	−0.91	−0.72	−0.51	−0.39	−0.30	−0.23	−0.17	−0.12	−0.08	−0.04	−0.02	0.00
−2.5	−1.38	−0.93	−0.73	−0.53	−0.40	−0.31	−0.24	−0.18	−0.13	−0.08	−0.04	−0.02	0.00
−2.4	−1.40	−0.95	−0.75	−0.54	−0.41	−0.32	−0.25	−0.18	−0.13	−0.08	−0.04	−0.02	0.00
−2.3	−1.43	−0.98	−0.77	−0.56	−0.43	−0.33	−0.25	−0.19	−0.13	−0.08	−0.04	−0.02	0.00
−2.2	−1.45	−1.00	−0.79	−0.57	−0.44	−0.34	−0.26	−0.19	−0.14	−0.09	−0.04	−0.02	0.00
−2.1	−1.48	−1.03	−0.82	−0.59	−0.45	−0.35	−0.27	−0.20	−0.14	−0.09	−0.04	−0.02	0.00
−2.0	−1.52	−1.05	−0.84	−0.61	−0.47	−0.36	−0.28	−0.21	−0.15	−0.09	−0.04	−0.02	0.00
−1.9	−1.55	−1.08	−0.86	−0.63	−0.48	−0.37	−0.29	−0.22	−0.15	−0.10	−0.05	−0.02	0.00
−1.8	−1.59	−1.11	−0.89	−0.65	−0.50	−0.39	−0.30	−0.22	−0.16	−0.10	−0.05	−0.02	−0.01
−1.7	−1.63	−1.15	−0.92	−0.67	−0.52	−0.40	−0.31	−0.23	−0.16	−0.10	−0.05	−0.02	−0.01
−1.6	−1.67	−1.18	−0.95	−0.69	−0.53	−0.42	−0.32	−0.24	−0.17	−0.11	−0.05	−0.03	−0.01
−1.5	−1.71	−1.21	−0.97	−0.72	−0.55	−0.43	−0.33	−0.25	−0.18	−0.11	−0.05	−0.03	−0.01
−1.4	−1.76	−1.25	−1.01	−0.74	−0.57	−0.45	−0.35	−0.26	−0.19	−0.12	−0.06	−0.03	−0.01
−1.3	−1.80	−1.29	−1.04	−0.77	−0.60	−0.47	−0.36	−0.27	−0.19	−0.12	−0.06	−0.03	−0.01
−1.2	−1.85	−1.33	−1.07	−0.80	−0.62	−0.49	−0.38	−0.28	−0.20	−0.13	−0.06	−0.03	−0.01
−1.1	−1.90	−1.37	−1.11	−0.83	−0.64	−0.51	−0.39	−0.30	−0.21	−0.13	−0.06	−0.03	−0.01
−1.0	−1.95	−1.41	−1.15	−0.86	−0.67	−0.53	−0.41	−0.31	−0.22	−0.14	−0.07	−0.03	−0.01
−0.9	−2.01	−1.46	−1.19	−0.89	−0.70	−0.55	−0.43	−0.32	−0.23	−0.15	−0.07	−0.04	−0.01
−0.8	−2.06	−1.51	−1.23	−0.92	−0.73	−0.57	−0.45	−0.34	−0.24	−0.16	−0.08	−0.04	−0.01
−0.7	−2.12	−1.55	−1.27	−0.96	−0.76	−0.60	−0.47	−0.36	−0.26	−0.16	−0.08	−0.04	−0.01
−0.6	−2.18	−1.61	−1.32	−1.00	−0.79	−0.63	−0.49	−0.38	−0.27	−0.17	−0.08	−0.04	−0.01

(continued)

Table 6.2 (continued)

k\G(t)	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99
−0.5	−2.24	−1.66	−1.37	−1.04	−0.83	−0.66	−0.52	−0.40	−0.29	−0.18	−0.09	−0.04	−0.01
−0.4	−2.30	−1.72	−1.42	−1.08	−0.86	−0.69	−0.55	−0.42	−0.30	−0.20	−0.10	−0.05	−0.01
−0.3	−2.37	−1.77	−1.47	−1.13	−0.90	−0.72	−0.57	−0.44	−0.32	−0.21	−0.10	−0.05	−0.01
−0.2	−2.44	−1.83	−1.53	−1.18	−0.94	−0.76	−0.60	−0.47	−0.34	−0.22	−0.11	−0.05	−0.01
−0.1	−2.51	−1.90	−1.59	−1.23	−0.99	−0.80	−0.64	−0.49	−0.36	−0.24	−0.12	−0.06	−0.01
0.0	−2.58	−1.96	−1.65	−1.28	−1.04	−0.84	−0.67	−0.52	−0.38	−0.25	−0.13	−0.06	−0.01
0.1	−2.65	−2.03	−1.71	−1.34	−1.09	−0.89	−0.71	−0.56	−0.41	−0.27	−0.14	−0.07	−0.01
0.2	−2.73	−2.10	−1.77	−1.40	−1.14	−0.93	−0.75	−0.59	−0.44	−0.29	−0.15	−0.07	−0.02
0.3	−2.80	−2.17	−1.84	−1.46	−1.19	−0.98	−0.80	−0.63	−0.47	−0.31	−0.16	−0.08	−0.02
0.4	−2.88	−2.24	−1.91	−1.52	−1.25	−1.04	−0.85	−0.67	−0.50	−0.34	−0.17	−0.09	−0.02
0.5	−2.96	−2.32	−1.98	−1.59	−1.32	−1.09	−0.90	−0.71	−0.54	−0.37	−0.19	−0.10	−0.02
0.6	−3.05	−2.40	−2.06	−1.66	−1.38	−1.15	−0.95	−0.76	−0.58	−0.40	−0.21	−0.11	−0.02
0.7	−3.13	−2.48	−2.13	−1.73	−1.45	−1.21	−1.01	−0.81	−0.62	−0.43	−0.23	−0.12	−0.02
0.8	−3.22	−2.56	−2.21	−1.80	−1.52	−1.28	−1.07	−0.87	−0.67	−0.47	−0.25	−0.13	−0.03
0.9	−3.30	−2.64	−2.29	−1.88	−1.59	−1.35	−1.13	−0.93	−0.72	−0.51	−0.27	−0.14	−0.03
1.0	−3.39	−2.73	−2.38	−1.96	−1.67	−1.42	−1.20	−0.99	−0.78	−0.55	−0.30	−0.16	−0.03
1.1	−3.48	−2.82	−2.46	−2.04	−1.75	−1.50	−1.27	−1.05	−0.83	−0.60	−0.34	−0.18	−0.04
1.2	−3.57	−2.90	−2.55	−2.13	−1.83	−1.57	−1.34	−1.12	−0.90	−0.65	−0.37	−0.20	−0.04
1.3	−3.67	−2.99	−2.64	−2.21	−1.91	−1.65	−1.42	−1.20	−0.96	−0.71	−0.41	−0.23	−0.05
1.4	−3.76	−3.09	−2.73	−2.30	−2.00	−1.74	−1.50	−1.27	−1.03	−0.77	−0.46	−0.26	−0.06
1.5	−3.85	−3.18	−2.82	−2.39	−2.08	−1.82	−1.58	−1.35	−1.11	−0.84	−0.51	−0.29	−0.07
1.6	−3.95	−3.27	−2.91	−2.48	−2.17	−1.91	−1.67	−1.43	−1.18	−0.91	−0.56	−0.33	−0.08
1.7	−4.04	−3.37	−3.01	−2.57	−2.26	−2.00	−1.76	−1.52	−1.26	−0.98	−0.62	−0.37	−0.09
1.8	−4.14	−3.46	−3.10	−2.67	−2.36	−2.09	−1.84	−1.60	−1.35	−1.06	−0.68	−0.42	−0.11
1.9	−4.24	−3.56	−3.20	−2.76	−2.45	−2.18	−1.94	−1.69	−1.43	−1.14	−0.75	−0.48	−0.13

2.0	−4.34	−3.66	−3.29	−2.86	−2.54	−2.28	−2.03	−1.78	−1.52	−1.22	−0.83	−0.54	−0.15
2.1	−4.43	−3.75	−3.39	−2.95	−2.64	−2.37	−2.12	−1.87	−1.61	−1.31	−0.91	−0.60	−0.18
2.2	−4.53	−3.85	−3.49	−3.05	−2.74	−2.47	−2.22	−1.97	−1.70	−1.40	−0.99	−0.67	−0.22
2.3	−4.63	−3.95	−3.59	−3.15	−2.83	−2.56	−2.31	−2.06	−1.80	−1.49	−1.07	−0.75	−0.26
2.4	−4.73	−4.05	−3.69	−3.25	−2.93	−2.66	−2.41	−2.16	−1.89	−1.58	−1.16	−0.83	−0.31
2.5	−4.83	−4.15	−3.79	−3.35	−3.03	−2.76	−2.51	−2.26	−1.99	−1.68	−1.25	−0.91	−0.36
2.6	−4.93	−4.25	−3.88	−3.44	−3.13	−2.86	−2.61	−2.35	−2.09	−1.77	−1.34	−1.00	−0.42
2.7	−5.03	−4.35	−3.98	−3.54	−3.23	−2.96	−2.70	−2.45	−2.18	−1.87	−1.44	−1.09	−0.49
2.8	−5.13	−4.45	−4.08	−3.64	−3.33	−3.06	−2.80	−2.55	−2.28	−1.97	−1.53	−1.18	−0.56
2.9	−5.23	−4.55	−4.18	−3.74	−3.43	−3.15	−2.90	−2.65	−2.38	−2.06	−1.63	−1.27	−0.64
3.0	−5.33	−4.65	−4.28	−3.84	−3.53	−3.25	−3.00	−2.75	−2.48	−2.16	−1.73	−1.37	−0.72

3. Table 3.1 gives the value of z related to F(z).
4. Hence,

$$t = (z - k)$$

Example 6.2 Table 6.2, of the right-truncated normal distribution, shows when the location parameter is k = 1.0, and the cumulative distribution is G(t) = 0.50, the percent-point is t = −1.20. The four steps below shows the computations.

1. Table 3.1 is called to give F(1) = 0.841 at k = 1.0
2. F(z) = 0.50 × 0.841 = 0.420
3. Table 3.2 shows z = −0.202 at F(z) = 0.420
4. Hence,

$$t = (z - k) = -0.202 - 1.0 = -1.202$$

The difference between the table values and the above computations is due to rounding.

6.7 Application to Sample Data

To apply the RTN Tables 6.1 and 6.2, to sample data, an estimate of the sample coefficient-of-variation, cov, is needed. The cov must be computed at a point where the high-limit, δ, is zero. Hence, the cov estimate is obtained as below:

If $\delta = 0$: $\text{cov} = s/\bar{x}$
If $\delta \neq 0$: $\text{cov} = s/(\bar{x} - \delta)$

A way to estimate the high-limit δ is provided in Chap. 7 (Truncated Normal Spread Ratio).

The sample data includes the following:$\bar{x}$= sample average
s = sample standard deviation
δ = high-limit
cov = $s/(\bar{x} - \delta)$ = adjusted coefficient of variation

The table statistics includes the following:

k = right-location parameter,
$\mu_t(k)$ = mean of t at k
$\sigma_t(k)$ = standard deviation of t at k

The percent-point conversions concerning variables x and t are as follows:

$$t\alpha = \mu_t(k) + \sigma_t(k)\left[\left(x\alpha - \bar{x}\right)/s\right]$$
$$x\alpha = \bar{x} + s[(t\alpha - \mu_t(k))/\sigma_t(k)]$$

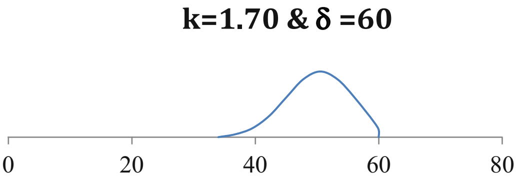

Fig. 6.2 Plot depicting the distribution of Example 6.3

where:

$t\alpha$ = α-percent-point of variable t
$x\alpha$ = α-percent-point of variable x

Example 6.3 A researcher has sample data from a right-truncated normal distribution where the mean is $\bar{x} = 50$, the standard deviation is $s = 5$, and high-limit is $\delta = 60$. The analyst wants to find the probability of x less than 55. To find this probability, the five steps below are followed.

1. The coefficient-of-variation is $cov = s/(\bar{x} - \delta) = 5/(50–60) = -0.50$.
2. Table 6.1 is searched to find $k \approx 1.70$ when $cov = -0.50$. Also, $\mu_t(k) = -1.798$ and $\sigma_t(k) = 0.907$.
3. Applying: $ta = \mu_t(k) + \sigma_t(k)\ [(xa - \bar{x})/s]$
   $= -1.798 + 0.907[(55 - 50/5)]$
   $= -0.891$
4. From Table 6.2 at $k = 1.70$ and $t = -0.891$: $\alpha \approx 0.83$.
5. Hence, $P(x < 55) \approx 0.83$.

See Fig. 6.2 for a depiction on the shape of the distribution.

Example 6.4 Suppose sample data from a right-truncated normal are available whose mean is $\bar{x} = 200$, the standard deviation is $s = 30$, and the right-location parameter is $\delta = 250$. Of interest is to find the probability of x less or equal to 210. To obtain, the four steps below are followed.

1. The adjusted coefficient-of-variation is $cov = s/(\bar{x} - \delta) = 30/(200–250) = -0.60$.
2. Table 6.1 is searched to find $k \approx 1.10$ when $cov = -0.60$. Also, $\mu_t(k) = -1.352$ and $\sigma_t(k) = 0.812$.
3. Using $x\alpha = 210$, the corresponding value of t is obtained as below:

$$\begin{aligned} t\alpha &= \mu_t(k) + \sigma_t(k)\ [(x\alpha - \bar{x})/s] \\ &= -1.352 + 0.812[(210 - 200)/30] \\ &= -1.08 \end{aligned}$$

Table 6.2 is searched with $t = -1.08$ and $k = 1.10$ to find $G(t) \approx 0.59$.
Hence, $P(x \leq 210) \approx 0.59$.
See Fig. 6.3 that depicts the shape of the distribution.

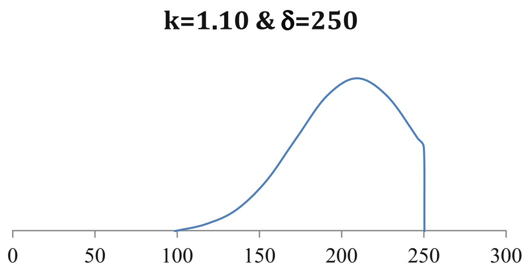

Fig. 6.3 Plot depicting the distribution shape of Example 6.4

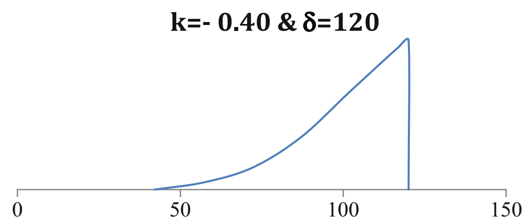

Fig. 6.4 Plot depicting the shape of the distribution for Example 6.5

Example 6.5 Suppose an analyst has sample data from a right-truncated normal distribution where $\bar{x} = 100$, s = 16 and the high-limit is δ = 120. Of interest is to find the 0.95-percent-point of x. To accomplish, the four steps below are followed:

1. The adjusted coefficient of variation is:

$$\text{cov} = 16/(100 - 120) = -0.80$$

2. Table 6.1 is searched with cov = −0.80 to find k ≈ −0.40.
   Note also, $\mu_t(k) = -0.669$ and $\sigma_t(k) = 0.534$.
3. Table 6.2 is searched with k = −0.40 and G(t) = 0.95 to find tα = −0.05.
4. The corresponding value of x is computed as below:

$$\begin{aligned} x\alpha &= \bar{x} + s\,[(t\alpha - \mu_t(k))/\sigma_t(k)] \\ &= 100 + 16[-0.05 - (-0.669)]/0.534] = 118.5 \end{aligned}$$

See Fig. 6.4.

6.8 Summary

A particular value of the standard normal, z = k, is selected as a right location parameter and all values of z smaller than k are allowed in a new distribution called the right-truncated normal. The mathematical equations for the mean, standard deviation, coefficient-of-variation, cumulative probability, and a variety of percent-points are developed. Table values for k ranging from −3.0 to +3.0 are listed. A second table lists a variety of percent-points (0.01–0.99) when k flows from −3.0 to +3.0. When an analyst has sample data and the coefficient of variation is computed, the analyst can estimate which value of k, right-location parameter, best fits the data, and thereby can compute all the probabilities needed for the sample data, without the need to always assume the data is from a normal distribution.

References

1. Johnson, A. C. (2001). *On the truncated normal distribution*. Doctoral Dissertation. Chicago: Stuart School of Business, Illinois Institute of Technology.
2. Johnson A. C., & Thomopoulos, N. T. (2001). *Tables on the left and right truncated normal distribution*. Proceedings of the Decision Sciences Institute, Chicago.

Chapter 7
Truncated Normal Spread Ratio

7.1 Introduction

A new statistic, called the spread ratio, is introduced that allows the analyst to identify which distribution best fits sample data, where the choice of distributions are the normal, left-truncated normal (LTN), and right-truncated normal (RTN). When the choice is the LTN or RTN, the location parameter, k, is also identified. The spread ratio for each distribution is computed using percent-points, $t_{0.01}$, $t_{0.99}$, and the mean of the distribution, and is a positive number. Using sample data, an estimate of the spread ratio is easily measured, and when the ratio is near one, the normal distribution fits the data best; when below one, the LTN is chosen; and when above one, the RTN is selected. For LTN, the low limit of the population data, γ, is estimated with use of the sample data and the tables provided on LTN. In the same way, if the RTN is chosen as the distribution, the high limit, δ, of the population data is easily estimated using sample data and the tables on RTN. Further, in either event, LTN or RTN, the analysis allows the researcher to estimate a value xα, where P (x ≤ xα) = α. Also when LTN or RTN, the analysis shows how to estimate the value of α for a given x\` where P(x ≤ x\`) = α.

7.2 The Spread Ratio

This chapter seeks a measure to identify the type of distribution to select for a collection of sample data. The choice is between the normal (N), left-truncated normal (LTN) and right-truncated normal (RTN). In the event of the latter two, the location limit is measured, γ for LTN, and δ for RTN. A new measure called the spread ratio, denoted as θ, is introduced here. This ratio is developed using the t distributions of the LTN and the RTN. The spread ratio, θ, is computed subsequently in the chapter. When θ is close to one, the N distribution is selected.

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_7

Otherwise, if θ is less than one, the LTN distribution is chosen; and if θ is greater than one, the RTN distribution is called.

When an analyst has sample data, denoted as $(x_1, \ldots, x_n)$, and the data is used to measure various stats of: $\bar{x}$, s, x(1) = min, x(n) = max; these stats are applied to estimate the spread ratio, θ. The estimated spread ratio is used to identify the type of distribution that is most likely: N, LTN or RTN. The stats also allow the analyst to estimate the low limit, γ for LTN, or the high limit, δ for RTN. Further, the analyst also can estimate the α-percent-point, xα, of the data, where $P(x \leq x\alpha) = \alpha$. Further, for any value x\` of variable x, the estimate of α is measured where $P(x \leq x\grave{}) = \alpha$.

7.3 LTN Distribution Measures

In Chap. 5 (Left Truncated Normal), various measures of the LTN distribution are defined, and some are used in this chapter. The LTN variable is labeled as $t = (z - k)$ and is greater or equal to zero because only the values of $z > k$ are in use. Recall, z is the variable from the standard normal distribution, and k is the left location parameter. The measures in use here are the following:

$t_{0.01}$ = 0.01-percent-point
$t_{.0.99}$ = 0.99-percent-point
$\mu_{t(k)}$ = mean of variable t with location parameter k
$\sigma_{t(k)}$ = standard deviation of variable t with location parameter k

Some new measures are defined in this chapter. These are the left-spread which is the width of the interval from $t_{0.01}$ to $\mu_{t(k)}$, and the right-spread that is the interval from $\mu_{t(k)}$ to $t_{0.99}$. The spread ratio, denoted as θ, is the ratio of the left-spread over the right-spread. A summary of these measures is below:

$$\left[\mu_{t(k)} - t_{0.01}\right] = \text{left} - \text{spread}$$

$$\left[t_{0.99} - \mu_{t(k)}\right] = \text{right} - \text{spread}$$

$$\theta = \left[\mu_{t(k)} - t_{0.01}/\left[t_{0.99} - \mu_{t(k)}\right] = \text{spread-ratio}\right.$$

7.4 LTN Table Entries

Table 7.1 is a list of various measures from the left-truncated normal distribution. The table is sorted by the left-location parameter, k, with a range of: [−3.0, (0.1), 3.0]. The table includes the percent-points, $t_{0.01}$ and $t_{0.99}$; the mean, $\mu_{t(k)}$, standard deviation, $\sigma_{t(k)}$, coefficient-of-variation, cov(k), and spread-ratio, θ. Note all values of θ are less than one, indicating the left-spread is smaller than the right-spread.

Table 7.1 Left truncated normal by location parameter k, percent-points $t_{0.01}$, $t_{0.99}$, mean, $\mu_t(k)$, standard deviation, $\sigma_t(k)$, coefficient-of-variation, cov(k), spread-ratio, θ

k	$t_{0.01}$	$t_{0.99}$	$\mu_{t(k)}$	$\sigma_{t(k)}$	cov(k)	θ
−3.0	0.720	5.327	3.004	0.993	0.33	0.98
−2.9	0.637	5.227	2.906	0.991	0.34	0.98
−2.8	0.559	5.127	2.808	0.989	0.35	0.97
−2.7	0.486	5.028	2.710	0.986	0.36	0.96
−2.6	0.419	4.928	2.614	0.982	0.38	0.95
−2.5	0.359	4.829	2.518	0.978	0.39	0.93
−2.4	0.305	4.729	2.423	0.972	0.40	0.92
−2.3	0.258	4.630	2.329	0.966	0.41	0.90
−2.2	0.218	4.532	2.236	0.959	0.43	0.88
−2.1	0.183	4.433	2.145	0.951	0.44	0.86
−2.0	0.155	4.335	2.055	0.942	0.46	0.83
−1.9	0.130	4.237	1.968	0.931	0.47	0.81
−1.8	0.110	4.140	1.882	0.920	0.49	0.78
−1.7	0.093	4.043	1.798	0.907	0.50	0.76
−1.6	0.080	3.947	1.717	0.894	0.52	0.73
−1.5	0.068	3.852	1.639	0.879	0.54	0.71
−1.4	0.059	3.758	1.563	0.863	0.55	0.69
−1.3	0.051	3.664	1.490	0.847	0.57	0.66
−1.2	0.044	3.572	1.419	0.830	0.58	0.64
−1.1	0.039	3.481	1.352	0.812	0.60	0.62
−1.0	0.034	3.390	1.288	0.794	0.62	0.60
−0.9	0.030	3.302	1.226	0.775	0.63	0.58
−0.8	0.027	3.214	1.168	0.756	0.65	0.56
−0.7	0.024	3.128	1.112	0.736	0.66	0.54
−0.6	0.022	3.044	1.059	0.717	0.68	0.52
−0.5	0.020	2.962	1.009	0.697	0.69	0.51
−0.4	0.018	2.881	0.962	0.678	0.70	0.49
−0.3	0.017	2.802	0.917	0.659	0.72	0.48
−0.2	0.015	2.724	0.875	0.640	0.73	0.47
−0.1	0.014	2.649	0.835	0.621	0.74	0.45
0.0	0.012	2.576	0.798	0.603	0.76	0.44
0.1	0.011	2.504	0.763	0.585	0.77	0.43
0.2	0.010	2.435	0.729	0.568	0.78	0.42
0.3	0.010	2.367	0.698	0.551	0.79	0.41
0.4	0.009	2.301	0.669	0.534	0.80	0.40
0.5	0.008	2.238	0.641	0.518	0.81	0.40
0.6	0.008	2.176	0.615	0.503	0.82	0.39
0.7	0.007	2.117	0.590	0.488	0.83	0.38
0.8	0.007	2.059	0.567	0.473	0.83	0.38
0.9	0.007	2.003	0.546	0.460	0.84	0.37
1.0	0.007	1.949	0.525	0.446	0.85	0.36
1.1	0.006	1.897	0.506	0.433	0.86	0.36

(continued)

Table 7.1 (continued)

k	$t_{0.01}$	$t_{0.99}$	$\mu_{t(k)}$	$\sigma_{t(k)}$	cov(k)	θ
1.2	0.006	1.846	0.488	0.421	0.86	0.35
1.3	0.006	1.797	0.470	0.409	0.87	0.35
1.4	0.006	1.750	0.454	0.398	0.88	0.35
1.5	0.005	1.704	0.439	0.387	0.88	0.34
1.6	0.005	1.660	0.424	0.376	0.89	0.34
1.7	0.005	1.617	0.410	0.366	0.89	0.34
1.8	0.005	1.575	0.397	0.356	0.90	0.33
1.9	0.005	1.534	0.385	0.347	0.90	0.33
2.0	0.004	1.495	0.373	0.338	0.91	0.33
2.1	0.004	1.456	0.362	0.330	0.91	0.33
2.2	0.004	1.417	0.351	0.321	0.91	0.33
2.3	0.004	1.379	0.341	0.313	0.92	0.33
2.4	0.004	1.340	0.332	0.306	0.92	0.33
2.5	0.003	1.301	0.323	0.298	0.92	0.33
2.6	0.003	1.260	0.314	0.291	0.93	0.33
2.7	0.003	1.218	0.306	0.284	0.93	0.33
2.8	0.002	1.173	0.298	0.277	0.93	0.34
2.9	0.002	1.124	0.291	0.270	0.93	0.35
3.0	0.001	1.070	0.283	0.264	0.93	0.36

When $k = -3.0$, the right and left spreads are almost the same size, similar to a normal distribution; and at $k = 3.0$, the left-spread is much smaller than the right-spread, like an exponential distribution.

7.5 RTN Distribution Measures

In Chap. 6 (Right Truncated Normal), various measures of the RTN distribution are defined, and some are repeated in this chapter. The RTN variable is defined as $t = (z - k)$ for $z < k$ and thereby t is less or equal to zero. As before, z is the variable from the standard normal distribution, and k is the right location parameter. The measures in use here are the following:

$t_{0.01}$ = 0.01-percent-point
$t_{0.99}$ = 0.99-percent-point
$\mu_{t(k)}$ = mean of variable t with location parameter k
$\sigma_{t(k)}$ = standard deviation of variable t with location parameter k

The spread ratio, θ, for the RTN is computed the same as given earlier for the LTN, and is repeated below:

$$\theta = \left[\mu_{t(k)} - t_{0.01}\middle/\left[t_{0.99} - \mu_{t(k)}\right]\right. = \text{spread-ratio}$$

7.6 RTN Table Entries

Table 7.2 is a list of various measures from the right-truncated normal distribution. The table is sorted by the left-location parameter, k, with a range of: [−3.0, (0.1), 3.0]. The table starts with k = −3.0 and includes the percent-points, $t_{0.01}$ and $t_{0.99}$; the mean, $\mu_{t(k)}$, standard deviation, $\sigma_{t(k)}$, coefficient–of-variation, cov(k), and spread-ratio, θ. Note all values of θ are greater than one, indicating the left-spread is larger than the right-spread. When k = 3.0, the right and left spreads are almost the same size, like a normal distribution. At k = −3.0, the left-spread is much larger than the right-spread, similar to an inverted exponential distribution.

7.7 Estimating the Distribution Type

Tables 7.1 and 7.2 can now be used to identify the type of distribution, normal or truncated, that most represents the stats gathered from a set of sample data. This scenario occurs when an analyst has n data points, (x_1, . . ., x_n), of a process and wishes to fit the data to such a distribution. With the data, the following statistics are listed below:

$\bar{x}$ = average
s = standard deviation
x(1) = minimum of sample
x(n) = maximum of sample

7.8 Selecting the Distribution Type

With the above stats, the sample estimate of the spread-ratio is computed as follows:

$$\widehat{\theta} = [\,\bar{x} - x(1)/[x(n) - \bar{x}]$$

Note, the sample spread ratio is always a value greater than zero, and when $\widehat{\theta}$ is reasonably close to 1.00, the normal distribution should be chosen. When below 1.00, the tilt is toward the LTN, and when above 1.00, it is towards the RTN. The author does not give a definitive rule on when $\widehat{\theta}$ deviates far enough away from 1.00 to not choose a normal distribution. A general rule on selecting the distribution type is given below:

If $\widehat{\theta} < 0.70$: select LTN
If $0.70 \leq \widehat{\theta} \leq 1.30$: select Normal
If $\widehat{\theta} > 1.30$: select RTN

Table 7.2 Right-truncated normal by location parameter k, percent-points $t_{0.01}$, $t_{0.99}$, mean, $\mu_t(k)$, standard deviation, $\sigma_t(k)$, coefficient-of-variation, cov(k), spread-ratio, θ

k	$t_{0.01}$	$t_{0.99}$	$\mu_{t(k)}$	$\sigma_{t(k)}$	cov(k)	θ
−3.0	−1.494	−0.006	−0.283	0.264	−0.93	4.36
−2.9	−1.397	−0.005	−0.291	0.270	−0.93	3.88
−2.8	−1.365	−0.005	−0.298	0.277	−0.93	3.64
−2.7	−1.359	−0.005	−0.306	0.284	−0.93	3.50
−2.6	−1.366	−0.005	−0.314	0.291	−0.93	3.40
−2.5	−1.381	−0.005	−0.323	0.298	−0.92	3.32
−2.4	−1.401	−0.005	−0.332	0.306	−0.92	3.27
−2.3	−1.426	−0.005	−0.341	0.313	−0.92	3.22
−2.2	−1.454	−0.005	−0.351	0.321	−0.91	3.18
−2.1	−1.485	−0.005	−0.362	0.330	−0.91	3.14
−2.0	−1.518	−0.005	−0.373	0.338	−0.91	3.11
−1.9	−1.553	−0.005	−0.385	0.347	−0.90	3.07
−1.8	−1.590	−0.005	−0.397	0.356	−0.90	3.04
−1.7	−1.629	−0.005	−0.410	0.366	−0.89	3.01
−1.6	−1.670	−0.005	−0.424	0.376	−0.89	2.98
−1.5	−1.713	−0.006	−0.439	0.387	−0.88	2.94
−1.4	−1.757	−0.006	−0.454	0.398	−0.88	2.91
−1.3	−1.803	−0.006	−0.470	0.409	−0.87	2.87
−1.2	−1.851	−0.006	−0.488	0.421	−0.86	2.83
−1.1	−1.901	−0.006	−0.506	0.433	−0.86	2.79
−1.0	−1.953	−0.007	−0.525	0.446	−0.85	2.75
−0.9	−2.006	−0.007	−0.546	0.460	−0.84	2.71
−0.8	−2.062	−0.007	−0.567	0.473	−0.83	2.67
−0.7	−2.119	−0.008	−0.590	0.488	−0.83	2.62
−0.6	−2.179	−0.008	−0.615	0.503	−0.82	2.58
−0.5	−2.240	−0.008	−0.641	0.518	−0.81	2.53
−0.4	−2.303	−0.009	−0.669	0.534	−0.80	2.48
−0.3	−2.369	−0.010	−0.698	0.551	−0.79	2.43
−0.2	−2.436	−0.010	−0.729	0.568	−0.78	2.37
−0.1	−2.506	−0.011	−0.763	0.585	−0.77	2.32
0.0	−2.577	−0.013	−0.798	0.603	−0.76	2.27
0.1	−2.650	−0.014	−0.835	0.621	−0.74	2.21
0.2	−2.726	−0.015	−0.875	0.640	−0.73	2.15
0.3	−2.803	−0.017	−0.917	0.659	−0.72	2.09
0.4	−2.882	−0.018	−0.962	0.678	−0.70	2.03
0.5	−2.963	−0.020	−1.009	0.697	−0.69	1.97
0.6	−3.045	−0.022	−1.059	0.717	−0.68	1.91
0.7	−3.129	−0.024	−1.112	0.736	−0.66	1.86
0.8	−3.215	−0.027	−1.168	0.756	−0.65	1.80
0.9	−3.303	−0.030	−1.226	0.775	−0.63	1.74
1.0	−3.391	−0.034	−1.288	0.794	−0.62	1.68
1.1	−3.481	−0.039	−1.352	0.812	−0.60	1.62

(continued)

Table 7.2 (continued)

k	$t_{0.01}$	$t_{0.99}$	$\mu_{t(k)}$	$\sigma_{t(k)}$	cov(k)	θ
1.2	−3.573	−0.044	−1.419	0.830	−0.58	1.57
1.3	−3.665	−0.051	−1.490	0.847	−0.57	1.51
1.4	−3.759	−0.059	−1.563	0.863	−0.55	1.46
1.5	−3.853	−0.068	−1.639	0.879	−0.54	1.41
1.6	−3.948	−0.080	−1.717	0.894	−0.52	1.36
1.7	−4.044	−0.094	−1.798	0.907	−0.50	1.32
1.8	−4.141	−0.110	−1.882	0.920	−0.49	1.28
1.9	−4.238	−0.131	−1.968	0.931	−0.47	1.24
2.0	−4.336	−0.155	−2.055	0.942	−0.46	1.20
2.1	−4.434	−0.184	−2.145	0.951	−0.44	1.17
2.2	−4.532	−0.218	−2.236	0.959	−0.43	1.14
2.3	−4.631	−0.259	−2.329	0.966	−0.41	1.11
2.4	−4.730	−0.305	−2.423	0.972	−0.40	1.09
2.5	−4.829	−0.359	−2.518	0.978	−0.39	1.07
2.6	−4.929	−0.419	−2.614	0.982	−0.38	1.06
2.7	−5.028	−0.486	−2.710	0.986	−0.36	1.04
2.8	−5.128	−0.559	−2.808	0.989	−0.35	1.03
2.9	−5.228	−0.638	−2.906	0.991	−0.34	1.02
3.0	−5.328	−0.721	−3.004	0.993	−0.33	1.02

Some may argue, the above selection rule should be adjusted depending on the sample size, n.

7.9 Estimating the Low and High Limits

When the LTN is selected, the low limit, denoted as γ, is estimated; and when the RTN is chosen, the high limit, denoted as δ, is estimated.

When LTN

In the event LTN is identified as the distribution to use, a next choice is to estimate the low limit of the data values, denoted as γ. In many situations, the low limit is known a-priori, and no estimate is needed. When γ is known, $x \geq \gamma$; and should $\gamma = 0$, then $x \geq 0$.

Estimate γ when LTN

When the low limit is not known and an estimate is needed, a way to approximate the value of γ is described below. From Table 7.1, gather the following measures that are associated with the value of the location parameter, k:

$t_{0.01}$ = 0.01-percent-point
$\mu_{t(k)}$ = mean
$\sigma_{t(k)}$ = standard deviation

Form the approximate relation between the sample data and the table measures as shown below:

$$(\gamma - \bar{x})/s = \left[t_{0.01} - \mu_{t(k)}\right]/\sigma_{t(k)}$$

Now use the above relation to estimate γ as follows:

$$\gamma' = \bar{x} + s\left[t_{0.01} - \mu_{t(k)}\right]/\sigma_{t(k)}$$
$$\widehat{\gamma} = \min[\gamma', x(1)]$$

When RTN

In the event RTN is identified as the distribution to use, a next choice is to estimate the high limit of the data values, denoted as δ. In many situations, the high limit is known a-priori. When δ is known, $x \le \delta$ applies.

Estimate δ when RTN

When the high limit is not known and an estimate is needed, a way to approximate the value of δ is described below. From Table 7.2, gather the following measures that are associated with the value of the location parameter, k:

$t_{0.01}$ = 0.01-percent-point
$t._{0.99}$ = 0.99-percent-point
$\mu_{t(k)}$ = mean
$\sigma_{t(k)}$ = standard deviation

Form the approximate relation between the sample data and the table measures as shown below:

$$\left(\delta - \bar{x}\right)/\mathrm{s} = \left[\mathrm{t}_{0.99} - \mu_{\mathrm{t(k)}}\right]/\sigma_{\mathrm{t(k)}}$$

Now use the above relation to estimate δ as follows:

$$\delta' = \bar{x} + \mathrm{s}\left[\mathrm{t}_{0.99} - \mu_{\mathrm{t(k)}}\right]/\sigma_{\mathrm{t(k)}}$$
$$\widehat{\delta} = \max[\delta', \mathrm{x(n)}]$$

When Normal

The normal distribution is selected when the sample spread-ratio, $\widehat{\theta}$, is close to 1.0. In this event, the related measures come from the standard normal distribution as described in Chap. 3 (Standard Normal).

Compute the Adjusted Coefficient of Variation

When the distribution choice for a set of sample data is LTN or RTN, and the respective limit, $\widehat{\gamma}$ or $\widehat{\delta}$, is estimated, it becomes possible to compute the adjusted coefficient of variation, cov, as follows:

If LTN: cov = $s/(\bar{x} - \widehat{\gamma})$
If RTN: cov = $s/(\bar{x} - \widehat{\delta})$

The analyst might compare the sample cov with its counterpart cov listed in Tables 7.1 or 7.2, to further verify the distribution type selected adequately represents the sample data. Note for example, it is possible to choose the wrong distribution since the spread ratios of Tables 7.1 and 7.2 are based exclusively on the truncated normal distributions. A case in point is the comparison of the normal and the continuous uniform distributions, where they both are symmetrical about their mean and the computed spread ratio for each is $\theta = 1.00$. Chapter 1 (Continuous Distributions), shows that the cov for a normal distribution is 0.33, and the cov for a continuous uniform distribution is 0.58.

7.10 Find xα Where P(x ≤ xα) = α

If the distribution chosen is LTN or RTN, the α-percent-point of variable x is estimated as below:

$$x\alpha = \bar{x} + s[(t\alpha - \mu_t(k))/\sigma_t(k)]$$

where tα is from Table 5.3 for the LTN, and from Table 6.2 for the RTN.

7.11 Find α Where P(x ≤ x`) = α

If the distribution chosen is LTN or RTN, the cumulative probability for a given x`, P (x ≤ x`) = α, is estimated using the two steps below:

1. $t\alpha = \mu_t(k) + \sigma_t(k)[(x` - \bar{x})/s]$
2. If LTN, find α from Table 5.3 at k and tα.
   If RTN, find α from Table 6.2 at k and tα.

Example 7.1 A sample yields the following statistics: $\bar{x} = 30$, s = 10, x(1) = 20 and x(n) = 50. The sample spread ratio becomes:

$$\hat{\theta} = (30 - 20)/(50 - 30) = 10/20 = 0.50.$$

Since $\hat{\theta} < 1.0$, the left-truncated normal is chosen, where Table 7.1 at $\theta = 0.50$ shows the location parameter as k = −0.45.

An estimate of the low limit, γ, is obtained as follows. Note at $\theta = 0.50$: $t_{0.01} = 0.019$, $\mu_{t(k)} = 0.98$ and $\sigma_{t(k)} = 0.69$. Thereby,

$$\gamma' = \bar{x} + s\left[t_{0.01} - \mu_{t(k)}\right]/\sigma_{t(k)}$$
$$= 30 + 10[(0.019 - 0.980)/0.69] = 16.1$$

$$\hat{\gamma} = \min[\gamma', x(1)]$$
$$= \min[16.1, 20] = 16.1$$

Figure 7.1 is a plot of the left-truncated normal distribution with k = −0.45 and the low limit estimated at $\hat{\gamma} = 16.1$.

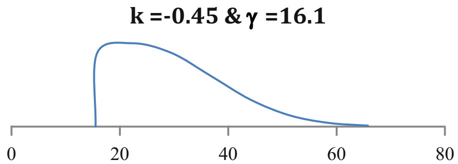

Fig. 7.1 Depiction of distribution from Example 7.1

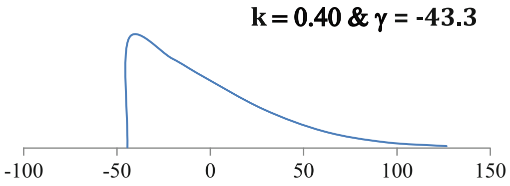

Fig. 7.2 Depiction of distribution from Example 7.2

The adjusted coefficient of variation is computed below:

$$\begin{aligned} \text{cov} &= s/\left(\bar{x} - \hat{\gamma}\right) \\ &= 10/(30 - 16.1) \\ &= 0.72 \end{aligned}$$

The corresponding cov from Table 7.1 with k = −0.45 is cov = 0.695.

Example 7.2 Suppose, n samples of a process gives the following: $\bar{x} = 0.0$, s = 35, x(1) = −40 and x(n) =100. The sample spread-ratio becomes:

$$\hat{\theta} = [0 - (-40)]/[100 - 0] = 40/100 = 0.40.$$

Since $\hat{\theta} < 1.0$, the left-truncated normal distribution applies; for which Table 7.1 is used to estimate the location parameter as, k = 0.40, showing the data is skewed to the right.

The associated parameters to estimate the low limit are: $t_{0.01} = 0.009$, $\mu_{t(k)} = 0.669$ and $\sigma_{t(k)} = 0.534$. Hence:

$$\gamma' = 0.00 + 35\,[(0.009 - 0.669)/0.534] = -43.3$$

$$\hat{\gamma} = \min[-43.3, -40] = -43.3$$

Figure 7.2 shows how the left-truncated normal distribution appears with the low limit estimated at $\hat{\gamma} = -43.3$.

The adjusted coefficient of variation is obtained as below:

$$\begin{aligned} \text{cov} &= s/\left(\bar{x} - \hat{\gamma}\right) \\ &= 35/(0.0 - (-43.3)) \\ &= 0.81 \end{aligned}$$

The corresponding cov from Table 7.1 with k = 0.40 is listed as cov = 0.80.

Example 7.3 A sample of size n yields the following: $\bar{x} = 300$, s = 20, x(1) = 250 and x(n) = 330, where the sample spread ratio becomes:

$$\hat{\theta} = (300 - 250)/(330 - 300) = 50/30 = 1.67.$$

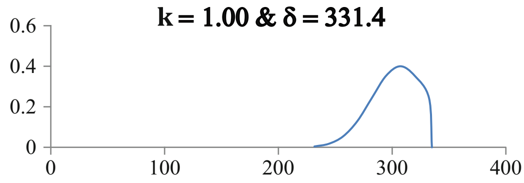

Fig. 7.3 Depiction of distribution from Example 7.3

Because, $\widehat{\theta} > 1.0$, the right-truncated normal is selected, and Table 7.2 is used to estimate the location parameter as k = 1.00.

An estimate of the high limit uses the following parameters when k = 1.00: $t_{0.99} = -0.034$, $\mu_{t(k)} = -1.288$, and $\sigma_{t(k)} = 0.794$. The high limit is now estimated below:

$$\begin{aligned}\delta' &= \bar{x} + s\left[t_{0.99} - \mu_{t(k)}\right]/\sigma_{t(k)} \\ &= 300 + 20[(-0.034 - (-1.288))/0.794] = 331.4\end{aligned}$$

$$\widehat{\delta} = \max[\delta', x(n)] = \max[331.4, 330] = 331.4$$

A plot for this data is presented in Fig. 7.3 where the right-truncated normal is selected and the high limit is estimated at $\widehat{\delta} = 331.4$.

The adjusted coefficient of variation is computed as follows:

$$\begin{aligned}\text{cov} &= s/\left(\bar{x} - \widehat{\delta}\right) \\ &= 20/(300 - 331.4) \\ &= -0.64\end{aligned}$$

The corresponding cov from Table 7.2 with k = 1.00 is listed as cov = −0.62.

Example 7.4 Suppose a sample yields the following: $\bar{x} = 110$, s = 44, x(1) = 50 and x(n) = 175. The sample spread ratio becomes:

$$\widehat{\theta} = (110 - 50)/(175 - 110) = 60/65 = 0.923.$$

Because, $\widehat{\theta} < 1.0$, the LTN is called and Table 7.1 shows the location parameter as k = −2.40 and cov ≈ 0.40, for which the data is similar to a normal distribution. Table 7.1 shows at k = −2.40, $\mu_t(k) = 2.423$, $\sigma_t(k) = 0.972$ and $t_{0.01} = 0.305$. The estimate of the low limit becomes:

$$\widehat{\gamma} = 110 + 44[(0.305 - 2.423)/0.972] = 14.12$$

A plot for this scenario is shown in Fig. 7.4.

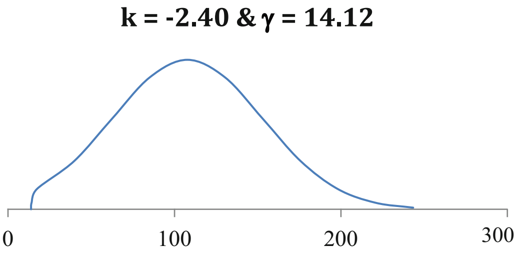

Fig. 7.4 Depiction of the distribution from Example 7.4

The adjusted coefficient of variation is computed below:

$$\begin{aligned} \text{cov} &= s/\left(\bar{x} - \hat{\gamma}\right) \\ &= 44/(110 - 14.12) \\ &= 0.46 \end{aligned}$$

The coefficient of variation from Table 7.1 with k = −2.40 is cov = 0.40.

Example 7.5 Assume the data of Example 7.1 where $\bar{x} = 30$, s = 10, x(1) = 20 and x(n) = 50, $\hat{\theta} = 0.50$, $\hat{\gamma} = 16.1$, and the LTN distribution is selected. Assume the analyst is seeking the 90-percent-point on x, where,

$$P(x \leq x_{0.90}) = 0.90$$

To find $x_{.90}$, the three steps below are taken:

1. Since the LTN is selected, Table 7.1 yields, with interpolation, the location parameter of k = −0.45, $\mu_t(k) = 0.98$ and $\sigma_t(k) = 0.69$.
2. Table 5.3 is now searched with k = −0.45 to find $t_{0.90} \approx 1.94$.
3. The 90-percent-point of x is computed as below:

$$\begin{aligned} x_{0.90} &= \bar{x} + s\left[\left(t0.90 - \mu_t(k)\right)/\sigma_t(k)\right] \\ &= 30 + 10[(1.94 - 0.98)/0.69] \\ &\approx 43.9 \end{aligned}$$

Example 7.6 Suppose the data of Example 7.3 where $\bar{x} = 300$, s = 20, x(1) = 250, x(n) = 330, $\hat{\theta} = 1.67$, k = 1.00, $\mu_t(k) = -1.288$, $\sigma_t(k) = 0.794$, and the RTN is selected with the high limit $\hat{\delta} = 331.4$. Assume the analyst is seeking the 10 percent percent-point on x, denoted as $x_{0.10}$.

To find $x_{0.10}$, the two steps below are run:

1. Table 6.2 is searched with k = 1.00 to find $t_{0.10} \approx -2.38$.
2. The 10-percent-point of x is obtained as below:

$$
\begin{aligned}
x_{0.10} &= \bar{x} + s\left[\left(t0.10 - \mu_t(k)\right)/\sigma_t(k)\right] \\
&= 300 + 20[(-2.38 - (-1.288))/0.794] \\
&= 272.5
\end{aligned}
$$

Example 7.7 Assume the data from Example 7.1 again, and suppose the analyst is seeking the cumulative probability when x\` = 40, i.e., $P(x \leq 40)$. Recall, $\bar{x}$ = 30, s = 10, k = −0.45, $\mu_t(k) = 0.98$, $\sigma_t(k) = 0.69$, where the LTN is selected. The value of variable t becomes:

$$
\begin{aligned}
t &= \mu_t(k) + \sigma_t(k)\left[\left(x\grave{} - \bar{x}\right)/s\right] \\
&= 0.98 + 0.69[(40 - 30)/10] \\
&= 1.67
\end{aligned}
$$

Table 5.3 is searched with k = −0.45 and t = 1.67 to find $\alpha \approx 0.84$. Hence, $P(x \leq 40) \approx 0.84$.

Example 7.8 Assume the data from Example 7.3 again, and suppose the analyst is seeking the cumulative probability when x\` = 280, i.e., $P(x \leq 280)$. Recall, $\bar{x}$ = 300, s = 20, k = 1.00, $\mu_t(k) = -1.288$, and $\sigma_t(k) = 0.794$ where the RTN is chosen. The value of variable t becomes:

$$
\begin{aligned}
t &= \mu_t(k) + \sigma_t(k)\left[\left(x\grave{} - \bar{x}\right)/s\right] \\
&= -1.288 + 0.794[(280 - 300)/20] \\
&= -2.08
\end{aligned}
$$

Table 6.3 is searched with k = 1.00 and t = −2.08 to find $\alpha \approx 0.17$; hence $P(x \leq 280) \approx 0.17$.

7.12 Summary

A new parameter, the spread ratio, is developed to identify the distribution that best fits sample data from: right truncated normal, left truncated normal, or normal. The value of the spread ratio is computed for each k of the left-truncated normal, and for every k of the right-truncated normal. When sample data is available, the analyst can easily estimate the spread ratio with the data and then determines which distribution best fits the sample data. The tables of this chapter and of Chaps. 5 and 6 allow the analyst to statistically analyze the data from the distribution chosen, without having to always assume the normal distribution.

Chapter 8
Bivariate Normal

8.1 Introduction

A common scenario in scientific studies occurs when two variables are jointly related in a statistical manner. It could be research on the height and weight by gender and ethnic group, where the weight of an adult is related to the height of a adult from the same gender and ethnic group. The bivariate normal distribution is the typical way to analyze data of this sort. The marginal distributions of the two variables are normally distributed, when their joint distribution is bivariate normal.

The distribution is defined by the marginal mean and standard deviation of each variable, and also the correlation between the two. A related distribution is the bivariate standard normal distribution, whose variables have a mean of zero and standard deviation of one. The correlation remains the same as the counterpart bivariate normal distribution and varies from -1.0 to $+1.0$. The mathematical relations pertaining to the marginal and conditional distributions of the standard normal are also described in the chapter. Because there is no closed-form solution to measure the probabilities, an approximation method is developed and applied in this chapter. Table entries of the joint cumulative probabilities are listed, one for each correlation ranging as: (-1.0, -0.9, . . ., 0.9, 1.0). For comparison sake, a single table is also structured to show how the joint probabilities vary by correlation. The chapter shows how to convert sample data from a bivariate normal to its counterpart bivariate standard normal, whereby joint probabilities are readily obtained. A series of examples are provided to guide the user on applications.

Over the years, a great many scholars have contributed to the literature concerning the bivariate normal distribution. A few associated with the author are listed here. In 1998, Montira Jantaravareerat studied the bivariate normal distribution and the cumulative probability distribution [1]; and also in 1998, Jantaravareerat and Thomopoulos describe how to estimate the cumulative joint probabilities for the bivariate normal distribution [2]. In 2001, Carol Lindee developed a method to estimate the cumulative probabilities for the multivariate normal distribution [3];

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_8

and also in 2001, Lindee with Thomopoulos formulated a method to estimate the cumulative joint probabilities for multivariate normal distributions [4]. This chapter develops a new method to approximate the cumulative probability of the bivariate normal distribution; from which the tables of the chapter are created.

8.2 Bivariate Normal Distribution

When two random variables, x_1 and x_2 are jointly related by the bivariate normal distribution their marginal distributions are normally distributed. The notation is:

$$(x_1, x_2) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$

where (μ_1, μ_2, σ_1, σ_2, ρ) are five parameters of the distribution; and where $(-\infty < x_1 < \infty)$, and $(-\infty < x_2 < \infty)$.

The probability density of the variables x_1 and x_2 is below:

$$f(x_1, x_2) = 1/\left[2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}\exp\left[-w/2\left(1-\rho^2\right)\right]\right.$$

where

$$w = \left[(x_1-\mu_1)/\sigma_1{}^2 + \left[(x_2-\mu_2)/\sigma_2\right]^2 - 2\rho[(x_1-\mu_1)\ (x_2-\mu_2)/\sigma_1\sigma_2\right]$$

Marginal Distributions

The marginal distributions of x_1 and x_2 are normally distributed as follows:

$$x_1 \sim N\left(\mu_1, \sigma_1{}^2\right)$$

and

$$x_2 \sim N\left(\mu_2, \sigma_2{}^2\right)$$

where, $E(x_1) = \mu_1$, and $E(x_2) = \mu_2$.

The correlation between x_1 and x_2 is computed as below:

$$\rho = \sigma_{12}/(\sigma_1\sigma_2)$$

where σ_{12} is the covariance between x_1 and x_2. The covariance is also denoted as C (x_1, x_2), and is computed as follows:

$$C(x_1, x_2) = E(x_1x_2) - E(x_1)E(x_2)$$

Conditional Distributions

When $x_1 = x_{1o}$, a particular value of x_1, the conditional mean of x_2 is:

$$\mu_{x2|x1o} = \mu_2 + \rho(\sigma_2/\sigma_1)(x_{1o} - \mu_1)$$

and the corresponding variance is

$$\sigma_{x2|x1o}{}^2 = \sigma_2{}^2\left(1 - \rho^2\right)$$

The conditional distribution of x_2 given x_{1o}, denoted as $x_2|x_{1o}$, is also normally distributed as,

$$x_2 \mid x_{1o} \sim N\left(\mu_{x2|x1o}, \sigma_{x2|x1o}{}^2\right)$$

In the same way, when $x_2 = x_{2o}$, a specific value of x_2, the conditional mean of x_1 is

$$\mu_{x1|x2o} = \mu_1 + \rho(\sigma_1/\sigma_2)(x_{2o} - \mu_2)$$

and

$$\sigma_{x1|x2o}{}^2 = \sigma_1{}^2\left(1 - \rho^2\right)$$

The associated conditional distribution of x_1 given x_{2o} is also normally distributed as,

$$x_1 \mid x_{2o} \sim N\left(\mu_{x1|x2o}, \sigma_{x1|x2o}{}^2\right)$$

8.3 Bivariate Standard Normal Distribution

The bivariate standard normal distribution is a basic case of the bivariate normal. The variables are denoted as z_1 and z_2, in place of x_1 and x_2. The means are zero; the standard deviations are equal to one; and the correlation could range from -1 to $+1$. The notation for the bivariate standard normal is below:

$$(z_1, z_2) \sim BVN(0, 0, 1, 1, \rho)$$

where,

$$\mu_1 = 0, \sigma_1 = 1, \mu_2 = 0, \sigma_2 = 1$$

Conditional Distribution of z_2

When $z_1 = k_1$:

$$\mu_{z2|k_1} = \rho k_1$$
$$\sigma_{z2|k1}{}^2 = (1 - \rho^2)$$

and the conditional distribution of z_2 given k_1 is also normal as below:

$$z_2 \mid k_1 \sim N\left(\mu_{z2|k1}, \sigma_{z2|k1}{}^2\right)$$

Conditional Distribution of z_1

When $z_2 = k_2$:

$$\mu_{z1|k_2} = \rho k_2$$
$$\sigma_{z1|k2}{}^2 = (1 - \rho^2)$$

The conditional distribution of z_1 given k_2 is below:

$$z_1 \mid k_2 \sim N\left(\mu_{z1|k2}, \sigma_{z1|k2}{}^2\right)$$

Cumulative Joint Probability

The cumulative joint probability of k_1 and k_2 is below:

$$F(k_1, k_2) = P(z_1 \leq k_1 \cap z_2 \leq k_2)$$

Because there is no closed form solution to the above, an approximation method is developed here and is described below.

Approximation of $F(k_1, k_2)$

The cumulative probability of $z_1 < k_1$ and $z_2 < k_2$ is mathematically obtained as below:

$$F(k_1, k_2) = \int_{-\infty}^{k_1} \int_{-\infty}^{k_2} f(z_1, z_2) dz_2 dz_1$$

But since there is no closed-form solution to the above, the steps listed below are applied to approximate the double integral. For simplicity, the range on the two standard variables are set to ∓ 3, instead of the conventional limits of $\mp\infty$. Thereby, $(-3 \le z_1 \le 3)$, and $(-3 \le z_2 \le 3)$ are used throughout. To begin the approximation, note the following:

$$\begin{aligned} F(k_1, k_2) &= \int_{-\infty}^{k_1} \left[\int_{-\infty}^{k_2} f(z_2|z_1) dz_2 \right] f(z_1) dz_1 \\ &= \int_{-\infty}^{k_1} F(k_2|z_1) f(z_1) dz_1 \end{aligned}$$

Applying the discrete normal distribution, described in Chap. 3 (Standard Normal), for discrete variable $z_1 = [-3.0, (0.1), k_1]$, the joint cumulative probability is approximated by:

$$F(k_1, k_2) \approx \sum_{z_1=-3.0}^{k_1-0.1} F(k_2|z_1) P(z_1) + 0.5F(k_2|k_1)P(k_1)$$

Recall, the discrete variable z_1 where,

$$P(z_1) \ge 0 \text{ and } \sum_{z1=-3.0}^{3.0} P(z1) = 1.0$$

Note, also,

$$F(k_2|\, z_1) = F\left[\left(k_2 - \mu_{z2|z1}\right) / \sigma_{z2|z1}\right]$$

where the marginal mean of z_2 given z_1 is,

$$\mu_{z2|z1} = \rho z_1$$

and the corresponding standard deviation is,

$$\sigma_{z2|z1} = \sqrt{(1 - \rho^2)}$$

Table Values of F(k_1, k_2)

Table 8.1 includes a series of 21 separate tables that list the approximated cumulative probability, F(k_1, k_2), for selected values of the correlation, ρ, and for (k_1, k_2). Each table pertains to a specific correlation value as follows: ρ = [−1.0, (0.1), 1.0]. The standard normal values are: k_1 = [−3.0, (0.5), 3.0], and k_2 = [−3.0, (0.5), 3.0].

Table 8.2 is a related table that summarizes and compares many of the results from Table 8.1. The range of correlations is: ρ = [−1.0, (0.2), 1.0]; and the standard normal values are: k_1 = [−3.0, (1.0), 3.0], and k_2 = [−3.0, (1.0), 3.0].

8.4 Some Basic Probabilities for (z_1, z_2) ~ BVN(0, 0, 1, 1, ρ)

Tables 8.1 and 8.2 allow the user to estimate various basic joint probabilities concerning the variables z_1 and z_2. Five of the common such probabilities are below:

$$
\begin{aligned}
P(z_1 \le k_1 \cap z_2 \le k_2) &= F(k_1, k_2)\\
P(z_1 > k_1 \cup z_2 > k_2) &= 1 - F(k_1, k_2)\\
P(z_1 > k_1 \cap z_2 > k_2) &= F(\infty, \infty) - F(k_1, \infty) - F(\infty, k_2) + F(k_1, k_2)\\
P(z_1 \le k_1 \cup z_2 \le k_2) &= F(k_1, \infty) + F(\infty, k_2) - F(k_1, k_2)\\
&= 1 - P(z_1 > k_1 \cap z_2 > k_2)\\
P(k_{1L} \le z_1 \le k_{1H} \cap k_{2L} \le z_2 \le k_{2H}) &= F(k_{1H}, k_{2H}) - F(k_{1H}, k_{2L}) - F(k_{1L}, k_{2H})\\
&\quad + F(k_{1L}, k_{2L})
\end{aligned}
$$

Example 8.1 Use Table 8.2 to find the probabilities listed below for the standard bivariate normal variables: (z_1, z_2) ~ BVN(0, 0, 1, 1, 0.8). In the calculations, use k = 3 in place of k = ∞, except at F(∞, ∞) = 1.00.

$$
\begin{aligned}
P(z_1 \le 1 \cap z_2 \le 1) &= F(1, 1)\\
&= 0.781\\
P(z_1 > 1 \cup z_2 > 1) &= 1 - F(1, 1)\\
&= 1 - 0.781\\
&= 0.219\\
P(z_1 > 1 \cap z_2 > 1) &= F(\infty, \infty) - F(1, 3) - F(3, 1) + F(1, 1)\\
&= 1.000 - 0.842 - 0.842 + 0.781\\
&= 0.097\\
P(z_1 \le 1 \cup z_2 \le 1) &= F(1, 3) + F(3, 1) - F(1, 1)\\
&= 0.842 + 0.842 - 0.781\\
&= 0.903\\
P(-1 \le z_1 \le 1 \cap -1 \le z_2 \le 1) &= F(1, 1) - F(1, -1) - F(-1, 1) + F(-1, -1)\\
&= 0.781 - 0.158 - 0.158 + 0.097\\
&= 0.562
\end{aligned}
$$

Table 8.1 Bivariate normal cumulative distribution $F(k_1, k_2)$ by correlation ρ

ρ =	−1												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0	0.02	0.06	0.15	0.30	0.49	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0	0.05	0.14	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0	0	0.09	0.24	0.43	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0	0	0.01	0.15	0.34	0.53	0.68	0.77	0.82	0.84	0.84
0.5	0	0	0	0	0.00	0.01	0.19	0.38	0.53	0.63	0.67	0.69	0.69
0	0	0	0	0	0.00	0.00	0.01	0.19	0.34	0.43	0.48	0.49	0.50
−0.5	0	0	0	0	0.00	0.00	0.00	0.01	0.15	0.24	0.29	0.30	0.31
−1	0	0	0	0	0.00	0.00	0.00	0.00	0.01	0.09	0.14	0.15	0.16
−1.5	0	0	0	0	0	0	0	0	0	0	0.04	0.06	0.07
−2	0	0	0	0	0	0	0	0	0	0	0	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	−0.9												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0	0.02	0.06	0.15	0.30	0.49	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0.01	0.05	0.14	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0	0.02	0.10	0.24	0.43	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0	0.01	0.04	0.16	0.34	0.53	0.68	0.78	0.82	0.84	0.84
0.5	0	0	0	0	0.01	0.06	0.20	0.38	0.53	0.63	0.67	0.69	0.69
0	0	0	0	0	0	0.01	0.07	0.20	0.34	0.43	0.48	0.49	0.50
−0.5	0	0	0	0	0	0	0.01	0.06	0.16	0.24	0.29	0.30	0.31
−1	0	0	0	0	0	0	0	0.01	0.04	0.10	0.14	0.15	0.16
−1.5	0	0	0	0	0	0	0	0	0.01	0.02	0.05	0.06	0.07
−2	0	0	0	0	0	0	0	0	0	0	0.01	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 8.1 (continued)

ρ =	−0.8												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.06	0.15	0.30	0.49	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0.01	0.05	0.14	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.01	0.03	0.11	0.25	0.43	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0	0.01	0.06	0.17	0.35	0.53	0.68	0.78	0.82	0.84	0.84
0.5	0	0	0	0	0.02	0.09	0.22	0.39	0.53	0.63	0.67	0.69	0.69
0	0	0	0	0	0.01	0.03	0.10	0.22	0.35	0.43	0.48	0.49	0.50
−0.5	0	0	0	0	0.00	0.01	0.03	0.09	0.17	0.25	0.29	0.30	0.31
−1	0	0	0	0	0.00	0.00	0.01	0.02	0.06	0.11	0.14	0.15	0.16
−1.5	0	0	0	0	0	0	0	0	0.01	0.03	0.05	0.06	0.07
−2	0	0	0	0	0	0	0	0	0	0.01	0.01	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	−0.7												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0	0.02	0.06	0.15	0.30	0.49	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0.02	0.05	0.14	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.01	0.04	0.11	0.25	0.44	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0	0.02	0.07	0.19	0.36	0.54	0.68	0.78	0.82	0.84	0.84
0.5	0	0	0	0.01	0.04	0.11	0.24	0.40	0.54	0.63	0.67	0.69	0.69
0	0	0	0	0	0.01	0.05	0.13	0.24	0.36	0.44	0.48	0.49	0.50
−0.5	0	0	0	0	0	0.02	0.05	0.11	0.19	0.25	0.29	0.30	0.31
−1	0	0	0	0	0	0	0.01	0.04	0.08	0.11	0.14	0.15	0.16
−1.5	0	0	0	0	0	0	0	0.01	0.02	0.04	0.05	0.06	0.07
−2	0	0	0	0	0	0	0	0	0	0.01	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

ρ =	−0.6												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.06	0.15	0.30	0.49	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0.02	0.06	0.14	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.01	0.04	0.12	0.26	0.44	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0.01	0.03	0.09	0.20	0.36	0.54	0.69	0.78	0.82	0.84	0.84
0.5	0	0	0	0.01	0.05	0.13	0.26	0.41	0.54	0.63	0.67	0.69	0.69
0	0	0	0	0.01	0.02	0.07	0.15	0.26	0.36	0.44	0.48	0.49	0.50
−0.5	0	0	0	0	0.01	0.03	0.07	0.13	0.20	0.26	0.29	0.30	0.31
−1	0	0	0	0	0.00	0.01	0.02	0.05	0.09	0.12	0.14	0.15	0.16
−1.5	0	0	0	0	0	0	0.01	0.01	0.03	0.04	0.06	0.06	0.07
−2	0	0	0	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	−0.5												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.06	0.15	0.30	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0	0.02	0.06	0.15	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.01	0.05	0.13	0.26	0.44	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0.01	0.03	0.10	0.21	0.37	0.55	0.69	0.78	0.82	0.84	0.84
0.5	0	0	0	0.02	0.06	0.15	0.27	0.42	0.55	0.63	0.67	0.69	0.69
0	0	0	0	0.01	0.03	0.08	0.17	0.27	0.37	0.44	0.48	0.49	0.50
−0.5	0	0	0	0	0.01	0.04	0.08	0.15	0.21	0.26	0.29	0.30	0.31
−1	0	0	0	0	0	0.01	0.03	0.06	0.10	0.13	0.15	0.15	0.16
−1.5	0	0	0	0	0	0	0.01	0.02	0.03	0.05	0.06	0.06	0.07
−2	0	0	0	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 8.1 (continued)

ρ =	−0.4												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.06	0.16	0.30	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.06	0.15	0.29	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.02	0.05	0.13	0.27	0.45	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0.01	0.04	0.11	0.22	0.38	0.55	0.69	0.78	0.82	0.84	0.84
0.5	0	0	0.01	0.03	0.07	0.16	0.29	0.43	0.55	0.63	0.67	0.69	0.69
0	0	0	0	0.01	0.04	0.10	0.18	0.29	0.38	0.45	0.48	0.49	0.50
−0.5	0	0	0	0.01	0.02	0.05	0.10	0.16	0.22	0.27	0.29	0.30	0.31
−1	0	0	0	0	0.01	0.02	0.04	0.07	0.11	0.13	0.15	0.15	0.16
−1.5	0	0	0	0	0	0.01	0.01	0.03	0.04	0.05	0.06	0.06	0.07
−2	0	0	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	−0.3												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.06	0.15	0.30	0.48	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0	0.02	0.06	0.14	0.27	0.45	0.63	0.78	0.87	0.91	0.93	0.93
1	0	0	0.01	0.04	0.11	0.23	0.39	0.56	0.69	0.78	0.82	0.84	0.84
0.5	0	0	0.01	0.03	0.08	0.17	0.30	0.44	0.56	0.63	0.67	0.69	0.69
0	0	0	0.01	0.02	0.05	0.11	0.2	0.30	0.39	0.45	0.48	0.50	0.50
−0.5	0	0	0	0.01	0.03	0.06	0.11	0.17	0.23	0.27	0.30	0.30	0.31
−1	0	0	0	0	0.01	0.03	0.05	0.08	0.11	0.14	0.15	0.16	0.16
−1.5	0	0	0	0	0	0.01	0.02	0.03	0.04	0.05	0.06	0.06	0.07
−2	0	0	0	0	0	0	0.01	0.01	0.01	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

ρ =	−0.2											
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.06	0.15	0.30	0.49	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0.01	0.02	0.06	0.14	0.28	0.46	0.64	0.78	0.87	0.91	0.93	0.93
1	0	0	0.02	0.05	0.12	0.24	0.40	0.57	0.70	0.78	0.82	0.84	0.84
0.5	0	0	0.01	0.04	0.09	0.19	0.32	0.45	0.57	0.64	0.67	0.69	0.69
0	0	0	0.01	0.02	0.06	0.13	0.22	0.32	0.40	0.46	0.48	0.50	0.50
−0.5	0	0	0	0.01	0.03	0.07	0.13	0.19	0.24	0.28	0.30	0.31	0.31
−1	0	0	0	0.01	0.01	0.03	0.06	0.09	0.12	0.14	0.15	0.16	0.16
−1.5	0	0	0	0	0.01	0.01	0.02	0.04	0.05	0.06	0.06	0.06	0.07
−2	0	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	−0.1												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.06	0.15	0.30	0.49	0.67	0.82	0.91	0.96	0.97	0.98
1.5	0	0.01	0.02	0.06	0.14	0.28	0.46	0.64	0.78	0.87	0.91	0.93	0.93
1	0	0	0.02	0.05	0.13	0.25	0.41	0.57	0.70	0.78	0.82	0.84	0.84
0.5	0	0	0.01	0.04	0.10	0.20	0.33	0.47	0.57	0.64	0.67	0.69	0.69
0	0	0	0.01	0.03	0.07	0.14	0.23	0.33	0.41	0.46	0.49	0.50	0.50
−0.5	0	0	0.01	0.02	0.04	0.08	0.14	0.20	0.25	0.28	0.30	0.31	0.31
−1	0	0	0	0.01	0.02	0.04	0.07	0.10	0.13	0.14	0.15	0.16	0.16
−1.5	0	0	0	0	0.01	0.02	0.03	0.04	0.05	0.06	0.06	0.07	0.07
−2	0	0	0	0	0	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 8.1 (continued)

ρ =	0											
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.30	0.49	0.68	0.82	0.91	0.96	0.97	0.98
1.5	0	0.01	0.02	0.06	0.15	0.29	0.47	0.65	0.79	0.87	0.91	0.93	0.93
1	0	0.01	0.02	0.06	0.13	0.26	0.42	0.58	0.71	0.79	0.82	0.84	0.84
0.5	0	0	0.02	0.05	0.11	0.21	0.35	0.48	0.58	0.65	0.68	0.69	0.69
0	0	0	0.01	0.03	0.08	0.15	0.25	0.35	0.42	0.47	0.49	0.50	0.50
−0.5	0	0	0.01	0.02	0.05	0.10	0.15	0.21	0.26	0.29	0.30	0.31	0.31
−1	0	0	0	0.01	0.03	0.05	0.08	0.11	0.13	0.15	0.15	0.16	0.16
−1.5	0	0	0	0	0.01	0.02	0.03	0.05	0.06	0.06	0.06	0.07	0.07
−2	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	0.1												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.30	0.49	0.68	0.82	0.91	0.96	0.97	0.98
1.5	0	0.01	0.02	0.06	0.15	0.29	0.47	0.65	0.79	0.87	0.91	0.93	0.93
1	0	0.01	0.02	0.06	0.14	0.27	0.43	0.59	0.71	0.79	0.82	0.84	0.84
0.5	0	0	0.02	0.05	0.12	0.23	0.36	0.49	0.59	0.65	0.68	0.69	0.69
0	0	0	0.01	0.04	0.09	0.17	0.27	0.36	0.43	0.47	0.49	0.50	0.50
−0.5	0	0	0.01	0.03	0.06	0.11	0.17	0.23	0.27	0.29	0.30	0.31	0.31
−1	0	0	0.01	0.01	0.03	0.06	0.09	0.12	0.14	0.15	0.16	0.16	0.16
−1.5	0	0	0	0.01	0.01	0.02	0.04	0.05	0.06	0.06	0.07	0.07	0.07
−2	0	0	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

ρ =	0.2												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.30	0.49	0.68	0.83	0.91	0.96	0.97	0.98
1.5	0	0.01	0.02	0.06	0.15	0.30	0.48	0.66	0.79	0.88	0.91	0.93	0.93
1	0	0.01	0.02	0.06	0.14	0.28	0.44	0.60	0.72	0.79	0.83	0.84	0.84
0.5	0	0.01	0.02	0.05	0.13	0.24	0.37	0.50	0.60	0.66	0.68	0.69	0.69
0	0	0	0.02	0.04	0.10	0.18	0.28	0.37	0.44	0.48	0.49	0.50	0.50
−0.5	0	0	0.01	0.03	0.07	0.12	0.18	0.24	0.28	0.30	0.30	0.31	0.31
−1	0	0	0.01	0.02	0.04	0.07	0.10	0.13	0.14	0.15	0.16	0.16	0.16
−1.5	0	0	0	0.01	0.02	0.03	0.04	0.05	0.06	0.06	0.07	0.07	0.07
−2	0	0	0	0	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	0.3												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.49	0.68	0.83	0.92	0.96	0.97	0.98
1.5	0	0.01	0.02	0.07	0.16	0.30	0.48	0.66	0.80	0.88	0.92	0.93	0.93
1	0	0.01	0.02	0.06	0.15	0.28	0.45	0.61	0.73	0.80	0.83	0.84	0.84
0.5	0	0.01	0.02	0.06	0.13	0.25	0.39	0.52	0.61	0.66	0.68	0.69	0.69
0	0	0	0.02	0.05	0.11	0.2	0.3	0.39	0.45	0.48	0.49	0.50	0.50
−0.5	0	0	0.01	0.04	0.08	0.13	0.2	0.25	0.28	0.30	0.31	0.31	0.31
−1	0	0	0.01	0.02	0.05	0.08	0.11	0.13	0.15	0.15	0.16	0.16	0.16
−1.5	0	0	0	0.01	0.02	0.04	0.05	0.06	0.06	0.06	0.07	0.07	0.07
−2	0	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 8.1 (continued)

ρ =	0.4												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.68	0.83	0.92	0.96	0.97	0.98
1.5	0	0.01	0.02	0.07	0.16	0.30	0.49	0.67	0.80	0.88	0.92	0.93	0.93
1	0	0.01	0.02	0.06	0.15	0.29	0.46	0.62	0.74	0.80	0.83	0.84	0.84
0.5	0	0.01	0.02	0.06	0.14	0.26	0.40	0.53	0.62	0.67	0.68	0.69	0.69
0	0	0.01	0.02	0.05	0.12	0.21	0.32	0.40	0.46	0.49	0.50	0.50	0.50
−0.5	0	0	0.02	0.04	0.09	0.15	0.21	0.26	0.29	0.30	0.31	0.31	0.31
−1	0	0	0.01	0.03	0.05	0.09	0.12	0.14	0.15	0.16	0.16	0.16	0.16
−1.5	0	0	0.01	0.01	0.03	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07
−2	0	0	0	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	0.5												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.83	0.92	0.96	0.97	0.98
1.5	0	0.01	0.02	0.07	0.16	0.30	0.49	0.67	0.81	0.89	0.92	0.93	0.93
1	0	0.01	0.02	0.07	0.15	0.30	0.47	0.63	0.75	0.81	0.83	0.84	0.84
0.5	0	0.01	0.02	0.06	0.15	0.27	0.42	0.55	0.63	0.67	0.69	0.69	0.69
0	0	0.01	0.02	0.06	0.13	0.23	0.33	0.42	0.47	0.49	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.05	0.10	0.16	0.23	0.27	0.30	0.30	0.31	0.31	0.31
−1	0	0	0.01	0.03	0.06	0.10	0.13	0.15	0.15	0.16	0.16	0.16	0.16
−1.5	0	0	0.01	0.02	0.03	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07
−2	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

ρ =	0.6												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.92	0.96	0.97	0.98
1.5	0	0.01	0.02	0.07	0.16	0.31	0.49	0.68	0.81	0.89	0.92	0.93	0.93
1	0	0.01	0.02	0.07	0.16	0.30	0.48	0.64	0.76	0.81	0.83	0.84	0.84
0.5	0	0.01	0.02	0.06	0.15	0.28	0.43	0.56	0.64	0.68	0.69	0.69	0.69
0	0	0.01	0.02	0.06	0.14	0.24	0.35	0.43	0.48	0.49	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.05	0.11	0.18	0.24	0.28	0.30	0.31	0.31	0.31	0.31
−1	0	0	0.02	0.04	0.07	0.11	0.14	0.15	0.16	0.16	0.16	0.16	0.16
−1.5	0	0	0.01	0.02	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07
−2	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	0.7												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.92	0.96	0.98	0.98
1.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.68	0.82	0.90	0.92	0.93	0.93
1	0	0.01	0.02	0.07	0.16	0.30	0.49	0.65	0.77	0.82	0.84	0.84	0.84
0.5	0	0.01	0.02	0.07	0.15	0.29	0.45	0.58	0.65	0.68	0.69	0.69	0.69
0	0	0.01	0.02	0.06	0.14	0.26	0.37	0.45	0.49	0.50	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.06	0.12	0.20	0.26	0.29	0.31	0.31	0.31	0.31	0.31
−1	0	0.01	0.02	0.04	0.08	0.12	0.14	0.15	0.16	0.16	0.16	0.16	0.16
−1.5	0	0	0.01	0.03	0.04	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07
−2	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

(continued)

Table 8.1 (continued)

ρ =	0.8												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.96	0.98	0.98
1.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.83	0.90	0.93	0.93	0.93
1	0	0.01	0.02	0.07	0.16	0.31	0.49	0.67	0.78	0.83	0.84	0.84	0.84
0.5	0	0.01	0.02	0.07	0.16	0.30	0.47	0.60	0.67	0.69	0.69	0.69	0.69
0	0	0.01	0.02	0.07	0.15	0.28	0.40	0.47	0.49	0.50	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.06	0.13	0.22	0.28	0.30	0.31	0.31	0.31	0.31	0.31
−1	0	0.01	0.02	0.05	0.10	0.13	0.15	0.16	0.16	0.16	0.16	0.16	0.16
−1.5	0	0	0.02	0.03	0.05	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07
−2	0	0	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0
ρ =	0.9												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.98	0.98
1.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.91	0.93	0.93	0.93
1	0	0.01	0.02	0.07	0.16	0.31	0.50	0.68	0.80	0.84	0.84	0.84	0.84
0.5	0	0.01	0.02	0.07	0.16	0.31	0.49	0.63	0.68	0.69	0.69	0.69	0.69
0	0	0.01	0.02	0.07	0.16	0.30	0.43	0.49	0.50	0.50	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.07	0.15	0.24	0.30	0.31	0.31	0.31	0.31	0.31	0.31
−1	0	0.01	0.02	0.06	0.11	0.15	0.16	0.16	0.16	0.16	0.16	0.16	0.16
−1.5	0	0.01	0.02	0.04	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
−2	0	0	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

ρ =	1												
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	0.99
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.98	0.98
1.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.93	0.93	0.93
1	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.84	0.84	0.84	0.84
0.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.69	0.69	0.69	0.69	0.69
0	0	0.01	0.02	0.07	0.16	0.31	0.50	0.50	0.50	0.50	0.50	0.50	0.50
−0.5	0	0.01	0.02	0.07	0.16	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
−1	0	0.01	0.02	0.07	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
−1.5	0	0.01	0.02	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
−2	0	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
−2.5	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
−3	0	0	0	0	0	0	0	0	0	0	0	0	0

8.5 Probabilities for $(x_1, x_2) \sim BVN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

When seeking to compute a probability pertaining to the bivariate normal variables, (x_1, x_2), the following conversion to their counterpart standard normal variables, (z_1, z_2), are applied as follows:

$$z_1 = (x_1 - \mu_1)/\sigma_1$$
$$z_2 = (x_2 - \mu_2)/\sigma_2$$

The standard normal variable values included in the probability examples are listed here: $(x_{1o}, x_{1L}, x_{1H}, x_{2o}, x_{2L}, x_{2H})$. The conversion to their counterpart standard normal values are shown below:

$$k_1 = (x_{1o} - \mu_1)/\sigma_1$$
$$k_{1L} = (x_{1L} - \mu_1)/\sigma_1$$
$$k_{1H} = (x_{1H} - \mu_1)/\sigma_1$$
$$k_2 = (x_{2o} - \mu_2)/\sigma_2$$
$$k_{2L} = (x_{2L} - \mu_2)/\sigma_2$$
$$k_{2H} = (x_{2H} - \mu_2)/\sigma_2$$

The way to derive a joint probability of (x_1, x_2) is by the associated standard normal variables (z_1, z_2), as shown with the examples below:

$$P(x_1 \leq x_{1o} \cap x_2 \leq x_{2o}) = P(z_1 \leq k_1 \cap z_2 \leq k_2)$$
$$P(x_1 > x_{1o} \cup x_2 > x_{2o}) = P(z_1 > k_1 \cup z_2 > k_2)$$
$$P(x_1 \geq x_{1o} \cap x_2 > x_{2o}) = P(z_1 > k_1 \cap z_2 > k_2)$$
$$P(x_1 \leq x_{1o} \cup x_2 \leq x_{2o}) = P(z_1 \leq k_1 \cup z_2 \leq k_2)$$
$$P(x_{1L} \leq x_1 \leq x_{1H} \cap x_{2L} \leq x_2 \leq x_{2H}) = P(k_{1L} \leq z_1 \leq k_{1H} \cap k_{2L} \leq z_2 \leq k_{2H})$$

Example 8.2 Assume $(x_1, x_2) \sim BVN(10, 20, 2, 3, 0.8)$. The variables (x_1, x_2) are converted to the counterpart standard bivariate normal variables, (z_1, z_2), as shown below:

$$z_1 = (x_1 - 10)/2$$
$$z_2 = (x_2 - 20)/3$$

Various probabilities concerning (x_1, x_2) are listed subsequently and have the following values: $x_{1o} = 12$, $x_{1L} = 8$, $x_{1H} = 12$, $x_{2o} = 23$, $x_{2L} = 17$ and $x_{2H} = 23$. The associated values for the standard bivariate normal are obtained as follows:

$$k_1 = (12 - 10)/2 = 1.0$$
$$k_{1L} = (8 - 10)/2 = -1.0$$
$$k_{1H} = (12 - 10)/2 = 1.0$$
$$k_2 = (23 - 20)/3 = 1.0$$
$$k_{2L} = (17 - 20)/3 = -1.0$$
$$k_{2H} = (23 - 20)/3 = 1.0$$

Table 8.2 Bivariate normal cumulative distribution $F(k_1, k_2)$ for selected values of k_1 and k_2 and $\rho = -1$ to 1

k1 /	k2	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
3	**3**	1.000	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.999	0.999	1.000
3	**2**	0.978	0.977	0.977	0.977	0.977	0.977	0.977	0.977	0.978	0.978	0.978
3	**1**	0.842	0.841	0.841	0.841	0.841	0.841	0.841	0.841	0.842	0.842	0.842
3	**0**	0.500	0.499	0.499	0.499	0.499	0.499	0.500	0.500	0.500	0.500	0.500
3	**−1**	0.158	0.157	0.157	0.157	0.158	0.158	0.158	0.158	0.158	0.158	0.158
3	**−2**	0.022	0.021	0.021	0.021	0.022	0.022	0.022	0.022	0.022	0.022	0.022
3	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	**3**	0.978	0.978	0.978	0.977	0.977	0.977	0.977	0.977	0.978	0.978	0.978
2	**2**	0.956	0.956	0.956	0.956	0.956	0.956	0.957	0.958	0.961	0.965	0.977
2	**1**	0.820	0.820	0.820	0.820	0.821	0.823	0.826	0.830	0.835	0.840	0.842
2	**0**	0.478	0.478	0.479	0.481	0.484	0.489	0.493	0.497	0.499	0.500	0.500
2	**−1**	0.136	0.138	0.143	0.147	0.151	0.154	0.157	0.158	0.158	0.158	0.158
2	**−2**	0.001	0.013	0.017	0.019	0.020	0.021	0.022	0.022	0.022	0.022	0.022
2	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	**3**	0.842	0.842	0.842	0.841	0.841	0.841	0.841	0.842	0.842	0.842	0.842
1	**2**	0.820	0.820	0.820	0.820	0.821	0.823	0.826	0.830	0.835	0.840	0.842
1	**1**	0.684	0.684	0.686	0.690	0.698	0.708	0.721	0.737	0.756	0.781	0.836
1	**0**	0.342	0.348	0.364	0.383	0.402	0.421	0.440	0.459	0.478	0.494	0.500
1	**−1**	0.006	0.061	0.086	0.105	0.120	0.133	0.143	0.151	0.156	0.158	0.158
1	**−2**	0.000	0.002	0.007	0.012	0.015	0.018	0.020	0.021	0.022	0.022	0.022
1	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0	**3**	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
0	**2**	0.478	0.478	0.479	0.482	0.485	0.489	0.493	0.497	0.499	0.500	0.500
0	**1**	0.342	0.348	0.364	0.383	0.402	0.421	0.440	0.459	0.478	0.494	0.500
0	**0**	0.010	0.103	0.148	0.185	0.218	0.250	0.282	0.315	0.352	0.397	0.500

(continued)

Table 8.2 (continued)

k1 /	k2	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
0	**−1**	0.000	0.006	0.022	0.041	0.060	0.079	0.098	0.117	0.136	0.152	0.158
0	**−2**	0.000	0.000	0.001	0.003	0.007	0.011	0.015	0.018	0.021	0.022	0.022
0	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
−1	**3**	0.158	0.158	0.158	0.158	0.158	0.159	0.159	0.158	0.158	0.158	0.158
−1	**2**	0.136	0.138	0.143	0.148	0.152	0.155	0.157	0.158	0.158	0.158	0.158
−1	**1**	0.006	0.061	0.086	0.105	0.121	0.134	0.144	0.152	0.156	0.158	0.158
−1	**0**	0.000	0.006	0.022	0.041	0.060	0.079	0.098	0.117	0.136	0.152	0.158
−1	**−1**	0.000	0.000	0.002	0.007	0.015	0.025	0.038	0.053	0.072	0.097	0.152
−1	**−2**	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.010	0.015	0.020	0.022
−1	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
−2	**3**	0.022	0.022	0.022	0.022	0.023	0.023	0.023	0.023	0.022	0.022	0.022
−2	**2**	0.001	0.013	0.017	0.020	0.021	0.022	0.023	0.023	0.022	0.022	0.022
−2	**1**	0.000	0.002	0.007	0.012	0.016	0.019	0.021	0.022	0.022	0.022	0.022
−2	**0**	0.000	0.000	0.001	0.003	0.007	0.011	0.016	0.019	0.021	0.022	0.022
−2	**−1**	0.000	0.000	0.000	0.000	0.002	0.004	0.007	0.011	0.015	0.020	0.022
−2	**−2**	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.009	0.020
−2	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
−3	**3**	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
−3	**2**	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.000
−3	**1**	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.000
−3	**0**	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.000
−3	**−1**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.000
−3	**−2**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000
−3	**−3**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The probabilities for (x_1, x_2) of interest are listed here along with the computed results:

$$P(x_1 \leq 12 \cap x_2 \leq 23\ ) = P(z_1 \leq 1 \cap z_2 \leq 1) = 0.781$$
$$P(x_1 > 12 \cup x_2 > 23) = P(z_1 > 1 \cup z_2 > 1) = 0.219$$
$$P(x_1 > 12 \cap x_2 > 23) = P(z_1 > 1 \cap z_2 > 1) = 0.097$$
$$P(x_1 \leq 12 \cup x_2 \leq 23) = P(z_1 \leq 1 \cup z_2 \leq 1) = 0.903$$
$$P(8 \leq x_1 \leq 12 \cap 17 \leq x_2 \leq 23) = P(-1 \leq z_1 \leq 1 \cap -1 \leq z_2 \leq 1) = 0.562$$

Example 8.3 Assume (x_1, x_2) ~ BVN(100, 50, 8, 5, 0.5) and $x_{1o} = 112$. Find the minimum x_2, denoted as x_{2o}, where $P(x_1 < 112 \cap x_2 < x_{2o}) \leq 0.90$.

Note $k_1 = (112–100)/8 = 1.5$, and using Table 8.1 with $\rho = 0.5$, the smallest k_2 that yields $F(1.5,\ k_2) = 0.90$ is, by interpolation, $k_2 \approx 1.67$. Hence, $x_{20} = (\mu_2 + k_2\sigma_2) = (50 + 1.67{\times}5) = 58.35$.

8.6 Summary

The joint bivariate probability distribution of two variables is listed along with its marginal and conditional distributions. A special case of this distribution is the standard bivariate normal, and the way to convert data from a bivariate normal to a standard bivariate normal is shown. An approximation method is developed to generate joint cumulative probabilities from the bivariate standard normal. Altogether, 21 tables are listed, one per correlation of: $\rho = (-1.0, -0.9, \ldots, 0.9, 1.0)$. Another table is listed that compares the cumulative probabilities from one correlation to another. When sample data is available, the analyst can readily use the tables to estimate a variety of joint probabilities concerning the sample data.

References

1. Jantaravareerat, M. (1998). *Approximation of the distribution function for the standard bivariate normal.* Doctoral Dissertation, Stuart School of Business, Illinois Institute of Technology.
2. Jantaravareerat, M., & Thomopoulos, N. (1998). Some new tables on the standard bivariate normal distribution. *Computing Science and Statistics, 30*, 179–184.
3. Lindee, C., (2001). *The multivariate standard normal distribution*. Doctoral Dissertation, Stuart School of Business, Illinois Institute of Technology.
4. Lindee, C., & Thomopoulos, N. (2001). Values for the cumulative distribution function of the standard multivariate normal distribution. *Proceedings of the Midwest Decision Sciences Institute, 48*, 600–609.

Chapter 9
Lognormal

9.1 Introduction

When the logarithm of a variable, x, yields a normal variable, y, x is distributed as a lognormal distribution. The reverse occurs when the exponent of y transforms back to the lognormal x. The lognormal distribution has probabilities that peak on the left of its range of values and has a tail that skews far to the right. The mean and variance of the lognormal are related to the mean and variance for the counterpart normal. Further, the parameters of the lognormal are the mean and variance of the counterpart normal. In the pursuit to develop percent-point values for the lognormal in this chapter, some mathematical maneuvering is needed. When the normal variable y has its mean shifted to zero, to produce another normal variable y`, the exponent of y` transfers back to a lognormal that becomes a standard lognormal variable where percent-point values can be computed. Table values for the standard lognormal are listed in the chapter. The way to start with sample data from a lognormal variable and convert to a standard lognormal variable is described so that analysts can readily apply the tables.

Francis Galton [1] is credited with being the first to formulate the lognormal distribution; and as such, the lognormal is sometimes referred as the Galton distribution. He was an Englishman who in the 1800s was active in a variety of disciplines, including: heredity, psychology and statistics. Since Galton, many researchers have enhanced the literature on applications on the lognormal distribution. In 2003, N. Thomopoulos and A.C. Johnson published tables on percent-point values of the standard lognormal distribution [2]. A new set of lognormal tables is developed in this chapter.

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_9

9.2 Lognormal Distribution

The lognormal distribution occurs when the logarithm of its variable, x, yields a normally distributed variable, y. In the reverse way, the exponential of normal y reverts back to the lognormal x.

9.3 Notation

For simplicity in this chapter, the following notation is used:

LN = lognormal
N = normal
w = raw LN variable
x = shifted LN variable
x` = standard LN variable
y = N variable
y` = zero-mean N variable

9.4 Lognormal

The lognormal distribution with variable $x \geq 0$, has a peak near zero and is skewed far to the right. This variable is related to a counterpart normal variable y, in the following way:

$$y = \ln(x)$$

where ln is the natural logarithm, and

$$x = e^y$$

The variable **y** is normally distributed with mean and variance, μ_y and σ_y^2, respectively, and x is lognormal with mean and variance, μ_x and σ_x^2. The designation for x and y are listed below:

$$x \sim LN\left(\mu_y, \sigma_y^2\right)$$
$$y \sim N\left(\mu_y, \sigma_y^2\right)$$

Note, the parameters to define the distribution of x, are the mean and variance of y. The parameters between x and y are related in the following way:

$$\mu_x = \exp\left[\mu_y + \sigma_y^2/2\right]$$
$$\sigma_x^2 = \exp\left[2\mu_y + \sigma_y^2\right]\left[\exp\left(\sigma_y^2\right) - 1\right]$$
$$\mu_y = \ln\left[\mu_x^2/\sqrt{\mu_x^2 + \sigma_x^2}\right]$$
$$\sigma_y^2 = \ln\left[1 + \sigma_x^2/\mu_x^2\right]$$

Lognormal Mode

The mode of the lognormal variable, x, denoted as $\widetilde{\mu}_x$, is obtained as below:

$$\widetilde{\mu}_x = \exp\left(\mu_y - \sigma_y^2\right)$$

Lognormal Median

The median of lognormal x is obtained as follows:

$$\mu_{0.5} = \exp\left(\mu_y\right)$$

9.5 Raw Lognormal Variable

Consider n observations from a lognormal distribution, denoted as $(w_1, \ldots, w_n)$. Assume the smallest value of w that can occur is denoted as γ, where,

$$w \geq \gamma.$$

The bound, γ, is called the low limit.

In the pursuit of seeking a standard lognormal distribution, the analysis to follow requires a lognormal variable that is always zero or larger. Oftentimes, $\gamma = 0$, and hence, $w \geq 0$ at the outset.

9.6 Shifted Lognormal Variable

In the event the smallest value of w is not zero, a shifted lognormal variable, x, is formed as below:

$$x = w - \gamma$$

where,

$$x \geq 0$$

When the low limit of w is $\gamma = 0$, then $x = w$. The mean and standard deviation of LN variable x is denoted as follows:

μ_x = mean
σ_x = standard deviation

9.7 Normal Variable

The LN variable, x, is converted to its counterpart normal variable, y, as below:

$$y = \ln(x)$$

9.8 Zero-Mean Normal Variable

In the quest to define a standard LN variable, it is necessary that the associated N variable have a mean of zero. To accommodate, a shifted normal variable, y`, is formed as below:

$$y` = y - \mu_y$$

So now, the mean of y` is:

$$\mu_{y`} = 0$$

and the standard deviation of y` is the same as for y, where:

$$\sigma_{y`} = \sigma_y$$

The normal designation becomes:

$$y` \sim N\left(0, \sigma_{y`}{}^2\right)$$

9.9 Standard LN Variable

The standard LN variable, x`, is obtained from the zero-mean normal variable, y`. Note the relation below that shows how to convert from y` to x`, and from y` to x:

$$x = e^{y`} e^{\mu_y} = x` e^{\mu_y}$$

Since $\mu_{y`} = 0$, the conversion from y` to x` is:

$$x` = e^{y`}$$

Also, the way to convert back from x` to y` is the following:

$$y` = \ln(x`)$$

The designation for the standard LN distribution is the following:

$$x` \sim LN\left(0, \sigma_{y`}{}^2\right)$$

where the parameters, $\mu_{y`} = 0$, $\sigma_{y`}{}^2$, are from the associated zero mean normal distribution.

Below shows how the mean and variance of x is related to the same of x`:

$$\mu_x = \mu_{x`} \exp\left(\mu_y\right)$$

$$\sigma_x{}^2 = \sigma_{x`}{}^2 \exp\left(2\mu_y\right)$$

9.10 Lognormal Table Entries

Table 9.1 contains selected statistics from the standard lognormal distribution. The table is sorted by the standard deviation of the zero-mean normal distribution, $\sigma_{y`}$, with a range of: [0.05, (0.05), 4.00]. The standard lognormal statistics are the following: three percent-points: $x`_{.01}$, $x`_{.50}$, $x`_{.99}$; the mode, $\widetilde{\mu}_{x`}$; and the mean, standard deviation and coefficient of variation, $\mu_{x`}$, $\sigma_{x`}$ and $cov_{x`}$, respectively. Recall, the notation for the variables are x` for the standard lognormal, and y` for the zero-mean normal.

Example 9.1 Table 9.1 shows the following percent-point results when $\sigma_{y`} = 1.80$: $x_{.01} = .02$, $x_{.50} = 1.00$ and $x_{.99} = 65.86$. The results are obtained as follows:

$$x_{.01} = \exp\left(z_{.01} \times \sigma_{y'}\right) = \exp(-2.327 \times 1.80) = 0.015$$

Table 9.1 Lognormal distribution sorted by the standard normal standard deviation, $\sigma_{y`}$; with lognormal: percent-point $x`_{.01}$; mode $\tilde{\mu}_{x`}$ ; median $x`_{.50}$; mean $\mu_{x`}$; percent-point $x`_{.99}$; standard deviation $\sigma_{x`}$; and coefficient of variation $cov_{x`}$

$\sigma_{y`}$	$x`_{.01}$	$\tilde{\mu}_{x`}$	$x`_{.50}$	μ_x	$x`_{.99}$	$\sigma_{x`}$	$cov_{x`}$
0.05	0.89	1.00	1.00	1.00	1.12	0.05	0.05
0.10	0.79	0.99	1.00	1.01	1.26	0.10	0.10
0.15	0.71	0.98	1.00	1.01	1.42	0.15	0.15
0.20	0.63	0.96	1.00	1.02	1.59	0.21	0.20
0.25	0.56	0.94	1.00	1.03	1.79	0.26	0.25
0.30	0.50	0.91	1.00	1.05	2.01	0.32	0.31
0.35	0.44	0.88	1.00	1.06	2.26	0.38	0.36
0.40	0.39	0.85	1.00	1.08	2.54	0.45	0.42
0.45	0.35	0.82	1.00	1.11	2.85	0.52	0.47
0.50	0.31	0.78	1.00	1.13	3.20	0.60	0.53
0.55	0.28	0.74	1.00	1.16	3.59	0.69	0.59
0.60	0.25	0.70	1.00	1.20	4.04	0.79	0.66
0.65	0.22	0.66	1.00	1.24	4.54	0.90	0.73
0.70	0.20	0.61	1.00	1.28	5.10	1.02	0.80
0.75	0.17	0.57	1.00	1.32	5.72	1.15	0.87
0.80	0.16	0.53	1.00	1.38	6.43	1.30	0.95
0.85	0.14	0.49	1.00	1.44	7.22	1.48	1.03
0.90	0.12	0.44	1.00	1.50	8.12	1.67	1.12
0.95	0.11	0.41	1.00	1.57	9.12	1.90	1.21
1.00	0.10	0.37	1.00	1.65	10.24	2.16	1.31
1.05	0.09	0.33	1.00	1.74	11.50	2.46	1.42
1.10	0.08	0.30	1.00	1.83	12.92	2.81	1.53
1.15	0.07	0.27	1.00	1.94	14.52	3.21	1.66
1.20	0.06	0.24	1.00	2.05	16.31	3.69	1.79
1.25	0.05	0.21	1.00	2.18	18.32	4.24	1.94
1.30	0.05	0.18	1.00	2.33	20.58	4.89	2.10
1.35	0.04	0.16	1.00	2.49	23.12	5.67	2.28
1.40	0.04	0.14	1.00	2.66	25.97	6.58	2.47
1.45	0.03	0.12	1.00	2.86	29.17	7.67	2.68
1.50	0.03	0.11	1.00	3.08	32.77	8.97	2.91
1.55	0.03	0.09	1.00	3.32	36.82	10.54	3.17
1.60	0.02	0.08	1.00	3.60	41.36	12.43	3.45
1.65	0.02	0.07	1.00	3.90	46.46	14.71	3.77
1.70	0.02	0.06	1.00	4.24	52.19	17.49	4.12
1.75	0.02	0.05	1.00	4.62	58.63	20.87	4.51
1.80	0.02	0.04	1.00	5.05	65.86	25.03	4.95
1.85	0.01	0.03	1.00	5.54	73.98	30.14	5.44
1.90	0.01	0.03	1.00	6.08	83.11	36.46	6.00
1.95	0.01	0.02	1.00	6.69	93.36	44.31	6.62
2.00	0.01	0.02	1.00	7.39	104.88	54.10	7.32
2.05	0.01	0.01	1.00	8.18	117.82	66.35	8.12

(continued)

Table 9.1 (continued)

$\sigma_{y`}$	$x`_{.01}$	$\tilde{\mu}_{x`}$	$x`_{.50}$	μ_x	$x`_{.99}$	$\sigma_{x`}$	$cov_{x`}$
2.10	0.01	0.01	1.00	9.07	132.35	81.77	9.01
2.15	0.01	0.01	1.00	10.09	148.68	101.25	10.04
2.20	0.01	0.01	1.00	11.25	167.02	125.97	11.20
2.25	0.01	0.01	1.00	12.57	187.62	157.48	12.53
2.30	0.00	0.01	1.00	14.08	210.76	197.84	14.05
2.35	0.00	0.00	1.00	15.82	236.76	249.76	15.79
2.40	0.00	0.00	1.00	17.81	265.97	316.85	17.79
2.45	0.00	0.00	1.00	20.11	298.78	403.94	20.09
2.50	0.00	0.00	1.00	22.76	335.64	517.51	22.74
2.55	0.00	0.00	1.00	25.82	377.04	666.31	25.80
2.60	0.00	0.00	1.00	29.37	423.55	862.14	29.35
2.65	0.00	0.00	1.00	33.49	475.80	1121.09	33.48
2.70	0.00	0.00	1.00	38.28	534.49	1465.07	38.27
2.75	0.00	0.00	1.00	43.87	600.42	1924.15	43.86
2.80	0.00	0.00	1.00	50.40	674.48	2539.70	50.39
2.85	0.00	0.00	1.00	58.05	757.69	3368.93	58.04
2.90	0.00	0.00	1.00	67.02	851.15	4491.26	67.01
2.95	0.00	0.00	1.00	77.58	956.15	6017.44	77.57
3.00	0.00	0.00	1.00	90.02	1074.09	8102.58	90.01
3.05	0.00	0.00	1.00	104.72	1206.59	10964.90	104.71
3.10	0.00	0.00	1.00	122.12	1355.43	14912.67	122.12
3.15	0.00	0.00	1.00	142.77	1522.63	20383.39	142.77
3.20	0.00	0.00	1.00	167.34	1710.45	28000.63	167.33
3.25	0.00	0.00	1.00	196.62	1921.45	38657.15	196.61
3.30	0.00	0.00	1.00	231.60	2158.47	53636.80	231.60
3.35	0.00	0.00	1.00	273.49	2424.73	74794.03	273.48
3.40	0.00	0.00	1.00	323.76	2723.83	104819.51	323.76
3.45	0.00	0.00	1.00	384.23	3059.83	147634.75	384.23
3.50	0.00	0.00	1.00	457.14	3437.28	208980.79	457.14
3.55	0.00	0.00	1.00	545.25	3861.29	297300.39	545.25
3.60	0.00	0.00	1.00	651.97	4337.60	425065.61	651.97
3.65	0.00	0.00	1.00	781.53	4872.67	610784.32	781.53
3.70	0.00	0.00	1.00	939.17	5473.74	882045.95	939.17
3.75	0.00	0.00	1.00	1131.44	6148.96	1,280,165	1131.44
3.80	0.00	0.00	1.00	1366.49	6907.47	1,867,292	1366.49
3.85	0.00	0.00	1.00	1654.49	7759.55	2,737,347	1654.49
3.90	0.00	0.00	1.00	2008.21	8716.74	4,032,915	2008.21
3.95	0.00	0.00	1.00	2443.65	9792.00	5,971,447	2443.65
4.00	0.00	0.00	1.00	2980.96	10999.90	8,886,110	2980.96

$$x_{.50} = \exp(z_{.50} \times \sigma_{y'}) = \exp(0.000 \times 1.80) = 1.000$$
$$x_{.99} = \exp(z_{.99} \times \sigma_{y'}) = \exp(2.327 \times 1.80) = 65.93$$

Table 3.1 provides $z_{.01}$, $z_{.50}$ and $z_{.99}$. Any differences in the above computations and table values are due to rounding.

Example 9.2 Table 9.1 shows the following parameter results when $\mu_{y`} = 0$, and $\sigma_{y`} = 1.80$: $\widetilde{\mu}_{x'}, = 0.04$, $\mu_{x`}, = 5.05$ and $\sigma_{x`} = 25.03$, and $cov_{x`} = 4.95$. The results are obtained as shown below:

$$\widetilde{\mu}_{x`} = \exp(-\sigma_{y`}^{2}) = \exp[-1.80^2)] = 0.04$$
$$\mu_{x`} = \exp(\sigma_{y`}^{2}/2) = \exp[1.80^2/2\ )] = 5.05$$
$$\sigma_{x`}^{2} = \exp[\sigma_{y}^{2}][\exp(\sigma_{y}^{2}) - 1] = \exp(1.80^2)[\exp(1.80^2) - 1] = (25.03)^2$$
$$cov_{x`} = \sigma_{x`}/\mu_{x`} = 25.03/5.05 = 4.95$$

Example 9.3 Assume an analyst has a variable x that is lognormal and has n data entries: $(x_1, \ldots, x_n)$. Of interest is to compute some statistical measures on the variable x. Note, the following four steps below:

1. Each sample of x is converted to a normal variable by y = ln(x) where ln is the natural log. So now, the converted data is $(y_1, \ldots, y_n)$. With this data, assume the mean and standard deviation of y are computed and are listed as below:

$$\mu_y = 0.6735$$
$$\sigma_y = 1.805$$

Note:

$$y \sim N(0.6735, 1.805^2)$$

2. The normal variable y is converted to the zero mean normal:

$$y` = (y - \mu_y)$$

Hence,

$$\mu_{y`} = 0$$
$$\sigma_{y`} = 1.805$$

and

$$y` \sim N(0, 1.805^2)$$

3. With the mean of y` set to zero, the standard lognormal variable, denoted as x`, is now computed by:

$$x` = e^{y`}$$

where,

$$\mu_{x`} = \exp\left[1.805^2/2\right] = 5.099$$
$$\sigma_{x`}{}^2 = \exp[1.8052]\left[\exp\left(1.805^2\right) - 1\right] = 25.495^2$$
$$\sigma_{x`} = 25.495$$

and,

$$x` \sim LN\left(0, 1.805^2\right).$$

4. To transform the standard lognormal variable x` back to the original lognormal variable x, the following is applied:

$$x = x`e^{\mu_y}$$

So now, the variable x is lognormal with:

$$\mu_x = 10$$
$$\sigma_x = 50$$

where,

$$x \sim LN\left(0.6735, 1.805^2\right).$$

With the conversions completed, it is now possible to measure some statistical relations of the original lognormal variable x by use of Tables 9.1 and 9.2. Note from Table 9.1 when $\mu_y = 0$ and $\sigma_y = 1.8$, the mean and standard deviation of the corresponding standard lognormal variable is $\mu_{x`} = 5.05$ and $\sigma_{x`} = 25.03$ with $cov_{x`} = 4.95$. Further, the percent-points for the standard lognormal at 0.01, 0.50 and 0.99 are 0.02, 1.00 and 65.86, respectively.

To convert the above results from the standard lognormal to the lognormal, first note:

$$e^{\mu_y} = e^{(0.6735)} = 1.961$$

Hence,

$$\mu_x = \mu_{x`} \times 1.961 = 5.099 \times 1.961 = 10.00$$

Table 9.2 Lognormal distribution sorted by (standard normal) standard deviation, $\sigma_{y'}$, with listing of standard lognormal percent-points for 0.01 to 0.99

$\sigma_{y'}$	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0.05	0.89	0.92	0.94	0.96	0.97	0.99	1.00	1.01	1.03	1.04	1.07	1.09	1.12
0.10	0.79	0.85	0.88	0.92	0.95	0.98	1.00	1.03	1.05	1.09	1.14	1.18	1.26
0.15	0.71	0.78	0.83	0.88	0.92	0.96	1.00	1.04	1.08	1.13	1.21	1.28	1.42
0.20	0.63	0.72	0.77	0.85	0.90	0.95	1.00	1.05	1.11	1.18	1.29	1.39	1.59
0.25	0.56	0.66	0.73	0.81	0.88	0.94	1.00	1.07	1.14	1.23	1.38	1.51	1.79
0.30	0.50	0.61	0.68	0.78	0.85	0.93	1.00	1.08	1.17	1.29	1.47	1.64	2.01
0.35	0.44	0.56	0.64	0.74	0.83	0.92	1.00	1.09	1.20	1.34	1.57	1.78	2.26
0.40	0.39	0.52	0.60	0.71	0.81	0.90	1.00	1.11	1.23	1.40	1.67	1.93	2.54
0.45	0.35	0.48	0.56	0.68	0.79	0.89	1.00	1.12	1.27	1.46	1.78	2.10	2.85
0.50	0.31	0.44	0.53	0.66	0.77	0.88	1.00	1.13	1.30	1.52	1.90	2.28	3.20
0.55	0.28	0.40	0.49	0.63	0.75	0.87	1.00	1.15	1.33	1.59	2.02	2.47	3.59
0.60	0.25	0.37	0.46	0.60	0.73	0.86	1.00	1.16	1.37	1.66	2.16	2.68	4.04
0.65	0.22	0.34	0.43	0.58	0.71	0.85	1.00	1.18	1.41	1.73	2.30	2.91	4.54
0.70	0.20	0.32	0.41	0.55	0.69	0.84	1.00	1.19	1.44	1.80	2.45	3.16	5.10
0.75	0.17	0.29	0.38	0.53	0.68	0.83	1.00	1.21	1.48	1.88	2.61	3.43	5.72
0.80	0.16	0.27	0.36	0.51	0.66	0.82	1.00	1.22	1.52	1.96	2.79	3.73	6.43
0.85	0.14	0.25	0.34	0.49	0.64	0.81	1.00	1.24	1.56	2.04	2.97	4.05	7.22
0.90	0.12	0.23	0.32	0.47	0.62	0.80	1.00	1.26	1.60	2.13	3.17	4.40	8.12
0.95	0.11	0.21	0.30	0.45	0.61	0.79	1.00	1.27	1.65	2.22	3.38	4.77	9.12
1.00	0.10	0.19	0.28	0.43	0.59	0.78	1.00	1.29	1.69	2.32	3.60	5.18	10.24
1.05	0.09	0.18	0.26	0.41	0.58	0.77	1.00	1.30	1.73	2.42	3.84	5.63	11.50
1.10	0.08	0.16	0.24	0.40	0.56	0.76	1.00	1.32	1.78	2.52	4.10	6.11	12.92
1.15	0.07	0.15	0.23	0.38	0.55	0.75	1.00	1.34	1.83	2.63	4.37	6.63	14.52
1.20	0.06	0.14	0.21	0.36	0.53	0.74	1.00	1.35	1.88	2.74	4.66	7.20	16.31
1.25	0.05	0.13	0.20	0.35	0.52	0.73	1.00	1.37	1.93	2.86	4.96	7.82	18.32
1.30	0.05	0.12	0.19	0.33	0.51	0.72	1.00	1.39	1.98	2.99	5.29	8.49	20.58

1.35	0.04	0.11	0.18	0.32	0.49	0.71	1.00	1.41	2.03	3.11	5.64	9.22	23.12
1.40	0.04	0.10	0.17	0.31	0.48	0.70	1.00	1.42	2.08	3.25	6.02	10.01	25.97
1.45	0.03	0.09	0.16	0.30	0.47	0.69	1.00	1.44	2.14	3.39	6.41	10.86	29.17
1.50	0.03	0.08	0.15	0.28	0.46	0.68	1.00	1.46	2.19	3.53	6.84	11.79	32.77
1.55	0.03	0.08	0.14	0.27	0.44	0.68	1.00	1.48	2.25	3.68	7.29	12.81	36.82
1.60	0.02	0.07	0.13	0.26	0.43	0.67	1.00	1.50	2.31	3.84	7.77	13.90	41.36
1.65	0.02	0.07	0.12	0.25	0.42	0.66	1.00	1.52	2.37	4.01	8.29	15.10	46.46
1.70	0.02	0.06	0.11	0.24	0.41	0.65	1.00	1.54	2.44	4.18	8.84	16.39	52.19
1.75	0.02	0.06	0.11	0.23	0.40	0.64	1.00	1.56	2.50	4.36	9.42	17.80	58.63
1.80	0.02	0.05	0.10	0.22	0.39	0.63	1.00	1.58	2.57	4.55	10.04	19.32	65.86
1.85	0.01	0.05	0.09	0.21	0.38	0.63	1.00	1.60	2.64	4.74	10.71	20.98	73.98
1.90	0.01	0.04	0.09	0.20	0.37	0.62	1.00	1.62	2.71	4.95	11.42	22.78	83.11
1.95	0.01	0.04	0.08	0.19	0.36	0.61	1.00	1.64	2.78	5.16	12.17	24.73	93.36
2.00	0.01	0.04	0.08	0.19	0.35	0.60	1.00	1.66	2.85	5.38	12.98	26.85	104.88
2.05	0.01	0.03	0.07	0.18	0.34	0.60	1.00	1.68	2.93	5.61	13.84	29.2	117.8
2.10	0.01	0.03	0.07	0.17	0.33	0.59	1.00	1.70	3.01	5.85	14.75	31.7	132.4
2.15	0.01	0.03	0.06	0.16	0.32	0.58	1.00	1.72	3.08	6.10	15.73	34.4	148.7
2.20	0.01	0.03	0.06	0.16	0.32	0.57	1.00	1.74	3.17	6.37	16.77	37.3	167.0
2.25	0.01	0.02	0.06	0.15	0.31	0.57	1.00	1.77	3.25	6.64	17.88	40.5	187.6
2.30	0.00	0.02	0.05	0.14	0.30	0.56	1.00	1.79	3.34	6.93	19.06	44.0	210.8
2.35	0.00	0.02	0.05	0.14	0.29	0.55	1.00	1.81	3.43	7.22	20.33	47.8	236.8
2.40	0.00	0.02	0.05	0.13	0.28	0.54	1.00	1.83	3.52	7.53	21.67	51.8	266.0
2.45	0.00	0.02	0.04	0.13	0.28	0.54	1.00	1.86	3.61	7.86	23.11	56.3	298.8
2.50	0.00	0.02	0.04	0.12	0.27	0.53	1.00	1.88	3.71	8.20	24.64	61.1	335.6
2.55	0.00	0.02	0.04	0.12	0.26	0.52	1.00	1.91	3.80	8.55	26.27	66.4	377.0
2.60	0.00	0.01	0.04	0.11	0.26	0.52	1.00	1.93	3.91	8.91	28.00	72.0	423.5
2.65	0.00	0.01	0.03	0.11	0.25	0.51	1.00	1.95	4.01	9.30	29.86	78.2	475.8
2.70	0.00	0.01	0.03	0.10	0.24	0.51	1.00	1.98	4.12	9.70	31.83	84.9	534.5

(continued)

Table 9.2 (continued)

$\sigma_{y'}$	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
2.75	0.00	0.01	0.03	0.10	0.24	0.50	1.00	2.00	4.22	10.11	33.94	92.2	600.4
2.80	0.00	0.01	0.03	0.09	0.23	0.49	1.00	2.03	4.34	10.55	36.19	100.1	674.5
2.85	0.00	0.01	0.03	0.09	0.22	0.49	1.00	2.06	4.45	11.00	38.58	108.7	757.7
2.90	0.00	0.01	0.02	0.09	0.22	0.48	1.00	2.08	4.57	11.47	41.13	118.0	851.2
2.95	0.00	0.01	0.02	0.08	0.21	0.47	1.00	2.11	4.69	11.97	43.86	128.1	956.1
3.00	0.00	0.01	0.02	0.08	0.21	0.47	1.00	2.14	4.82	12.48	46.76	139.1	1074.1
3.05	0.00	0.01	0.02	0.08	0.20	0.46	1.00	2.16	4.94	13.02	49.85	151.0	1206.6
3.10	0.00	0.01	0.02	0.07	0.20	0.46	1.00	2.19	5.07	13.58	53.15	164.0	1355.4
3.15	0.00	0.01	0.02	0.07	0.19	0.45	1.00	2.22	5.21	14.16	56.67	178.1	1522.6
3.20	0.00	0.01	0.02	0.07	0.19	0.45	1.00	2.25	5.35	14.77	60.42	193.3	1710.5
3.25	0.00	0.00	0.02	0.06	0.18	0.44	1.00	2.27	5.49	15.40	64.42	209.9	1921.4
3.30	0.00	0.00	0.01	0.06	0.18	0.43	1.00	2.30	5.64	16.07	68.68	227.9	2158.5
3.35	0.00	0.00	0.01	0.06	0.17	0.43	1.00	2.33	5.79	16.76	73.23	247.4	2424.7
3.40	0.00	0.00	0.01	0.06	0.17	0.42	1.00	2.36	5.94	17.48	78.08	268.6	2723.8
3.45	0.00	0.00	0.01	0.05	0.16	0.42	1.00	2.39	6.10	18.23	83.24	291.7	3059.8
3.50	0.00	0.00	0.01	0.05	0.16	0.41	1.00	2.42	6.26	19.01	88.75	316.7	3437.3
3.55	0.00	0.00	0.01	0.05	0.16	0.41	1.00	2.45	6.42	19.83	94.63	343.8	3861.3
3.60	0.00	0.00	0.01	0.05	0.15	0.40	1.00	2.49	6.59	20.68	100.9	373.3	4337.6
3.65	0.00	0.00	0.01	0.05	0.15	0.40	1.00	2.52	6.77	21.57	107.6	405.3	4872.7
3.70	0.00	0.00	0.01	0.04	0.14	0.39	1.00	2.55	6.95	22.49	114.7	440.1	5473.7
3.75	0.00	0.00	0.01	0.04	0.14	0.39	1.00	2.58	7.13	23.46	122.3	477.8	6149.0
3.80	0.00	0.00	0.01	0.04	0.14	0.38	1.00	2.61	7.32	24.47	130.4	518.8	6907.5
3.85	0.00	0.00	0.01	0.04	0.13	0.38	1.00	2.65	7.52	25.52	139.0	563.2	7759.6
3.90	0.00	0.00	0.01	0.04	0.13	0.37	1.00	2.68	7.72	26.62	148.2	611.5	8716.7
3.95	0.00	0.00	0.01	0.04	0.13	0.37	1.00	2.72	7.92	27.76	158.0	663.9	9792.0
4.00	0.00	0.00	0.01	0.03	0.12	0.36	1.00	2.75	8.13	28.95	168.5	720.9	10999.9

Also,

$$P(x` \leq 0.02) = P(x \leq 0.02 \times 1.961) = P(x \leq 0.392) = 0.01$$
$$P(x` \leq 1.00) = P(x \leq 1.00 \times 1.961) = P(x \leq 1.961) = 0.50$$
$$P(x` \leq 65.86) = P(x \leq 65.86 \times 1.961) = P(x \leq 129.15) = 0.99$$

The entries in Table 9.2 also pertain to the standard lognormal variable. When $\mu_{y`} = 0$ and $\sigma_{y`} = 1.80$, the percent-points are listed for 0.01 to 0.99. For example, with use of the table, the 90% tolerance interval of x` is:

$$P(0.05 \leq x` \leq 19.32) = 0.90$$

The corresponding 90% tolerance interval of x is obtained from the following computations:

$$(0.05 \times 1.961) = 0.098$$
$$(19.32 \times 1.961) = 37.887$$

and thereby,

$$P(0.098 \leq x \leq 37.887) = 0.90$$

9.11 Lognormal Distribution Table

Table 9.2 contains selected percent-points for the standard lognormal distribution. The table is sorted by the standard deviation, $\sigma_{y`}$, of the zero-mean normal distribution with a range of [0.05, (0.05), 4.00]. The selected percent-points are for cumulative probabilities, α = [.01, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95, .99]. The α-percent-point of x` is obtained by:

$$x`\alpha = \exp\left(z\alpha\sigma_{y`}\right)$$

where,

zα is from the standard normal of Table 3.1, and $P(z < z\alpha) = \alpha$.
$\sigma_{y`}$ = standard deviation from the zero-mean normal distribution.
x`α = α-percent-point from the standard lognormal distribution.

Note, where the median, denoted as $x_{.50}$, is equal to 1.00 throughout the table. Also observe where the spread from ($x_{.01}$ to $x_{.50}$), is much smaller than the counterpart spread from ($x_{.50}$ to $x_{.99}$), for most of the table. When $\sigma_{y`}$ is small, the spreads are fairly even, indicating, the lognormal distribution is similar to a normal distribution when $\sigma_{y`}$ is near 0.5. See Fig. 9.1.

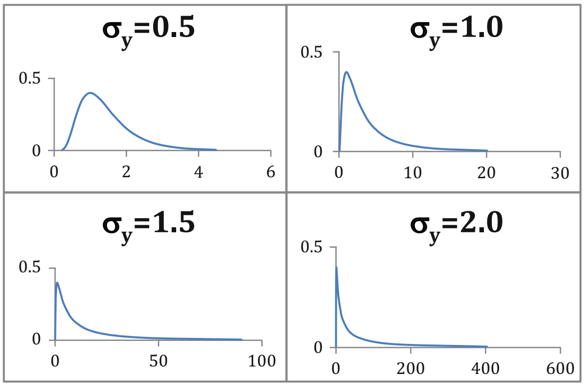

Fig. 9.1 Standard lognormal when $\sigma_y = 0.5$, 1.0, 1.5 and 2.0

Example 9.4 Table 9.2 gives the following results when $\sigma_{y`} = 1.00$: $x_{.05} = 0.19$ and $x_{.95} = 5.18$. The computations below show how these are obtained. First note from Table 3.2, where $z_{.05} = -1.645$ and $z_{.95} = 1.645$.

$$x`_{.05} = \exp\left(z_{.05} \times \sigma_{y`}\right) = \exp(-1.645 \times 1.00) = 0.19$$
$$x`_{.95} = \exp\left(z_{.95} \times \sigma_{y`}\right) = \exp(1.645 \times 1.00) = 5.181$$

9.12 Summary

The relation between the lognormal and normal distributions is described; along with the conversion between the means and variances of each. The development of a standard lognormal distribution is the focus of the chapter. The standard lognormal is related to a zero mean normal. This allows an analyst who has lognormal data to convert to a standard lognormal distribution for which statistical analysis is readily performed. The statistical measures of the standard lognormal are the following: mean, standard deviation, coefficient of variation, median, mode, and a variety of percent-points.

References

1. Galton, F. (1909). *Memories of my life*. New York: E.P. Dutton & Co.
2. Thomopoulos, N. T., & Johnson, A. C. (2003). Tables and characteristics of the standardized lognormal distribution. *Proceeding of the decision sciences institute, 103*, 1–6.

Chapter 10
Bivariate Lognormal

10.1 Introduction

When two variables are jointly related by the bivariate lognormal distribution, their marginal distributions are lognormal. At first, the distribution appears confounding due to the lognormal characteristics of the variables. By taking the log of each marginal distribution, a pair of normal marginal distributions evolve, and these are jointly related by the bivariate normal distribution described in Chap. 8 (Bivariate Normal). The bivariate normal is defined with the mean and standard deviation of each normal variable and by the correlation between them. These five parameters become the parameters for the counterpart bivariate lognormal distribution. The chapter shows how the mean and variance from the normal is transformed to the mean and variance for the lognormal. Also described is how to compute the correlation of the lognormal from the normal parameters. When sample data is distributed as bivariate lognormal, converting some data and applying the bivariate normal tables of Chap. 8 allow computation for a variety of joint probabilities.

Over the years, many scholars have provided mathematical formulation and applications on the lognormal bivariate distribution. Some applications familiar with the author are referred here. In 1984, N. Thomopoulos and Anatol Longinow showed how to compute the bivariate lognormal distribution for a structural engineering reliability problem [1]. During 2004, N. Thomopoulos and A.C. Johnson developed a way to compute the cumulative probabilities for a bivariate lognormal distribution [2].

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1_10

10.2 Bivariate Lognormal

When two variables, (x_1, x_2), are bivariate lognormal, their marginal distributions are lognormal. The associated bivariate normal variables are (y_1, y_2), where $y_i = \ln(x_i)$ $i = 1, 2$, and ln is the natural logarithm. The parameters for the lognormal distribution are those from the normal variables, y_1, y_2.

Notation

For clarification sake, it may be helpful to review the notation in use in this chapter, as listed below:

LN = lognormal
N = normal
x = LN variable
x` = standard LN variable
y = normal variable
y` = zero-mean N variable
z = standard N variable

Some Properties Between x and y

The lognormal variable, x, and the normal variable, y, are related as shown below:

$$y = \ln(x)$$
$$x = e^{y}$$

where ln is the natural log. Further, it is assumed $x \geq 0$.

The standard lognormal variable, x`, and the zero-mean normal variable, y`, are also related to each other. The zero-mean normal variable, y`, is obtained by shifting away from the mean of y as below:

$$y` = \left(y - \mu_y\right)$$

and thereby has a zero mean. The variable x` and y` are related as follows:

$$x` = e^{y`}$$
$$y` = \ln(x`)$$

Note also,

$$x = e^{y`} e^{\mu_y} = x` e^{\mu_y}$$

Mode of x and x`

The mode of lognormal x and x` are computed as below:

$$\text{mode}(x) = \exp\left(\mu_y - {\sigma_y}^2\right)$$
$$\text{mode}(x`) = \exp\left(-{\sigma_y}^2\right)$$

10.3 Lognormal and Normal Notation

The variable **y** is normally distributed with mean and variance, μ_y and ${\sigma_y}^2$, respectively, and x is lognormal with mean and variance, μ_x and ${\sigma_x}^2$. The notation for x and y are as below:

$$x \sim LN\left(\mu_y, {\sigma_y}^2\right)$$
$$y \sim N\left(\mu_y, {\sigma_y}^2\right)$$

Related Parameters

The parameters to define the distribution of x, are the mean and variance of y. The parameters between x and y are related in the following way:

$\mu_x = \exp\ [\mu_y + {\sigma_y}^2/2]$ = mean of x

${\sigma_x}^2 = \exp\ [2\mu_y + {\sigma_y}^2][\exp({\sigma_y}^2) - 1]$ = variance of x

$\mu_y = \ln\left[{\mu_x}^2/\sqrt{\mu_x^2 + \sigma_x^2}\right]$ = mean of y

${\sigma_y}^2 = \ln\ [1 + {\sigma_x}^2/{\mu_x}^2]$ = variance of y

10.4 Bivariate Lognormal Distribution

The bivariate lognormal is a joint distribution with variables, x_1 and x_2 when the marginal distributions are lognormal as below:

$$x_1 \sim LN\left(\mu_{y1}, {\sigma_{y1}}^2\right)$$

$$x_2 \sim LN\left(\mu_{y2}, \sigma_{y2}{}^2\right)$$

and where the counterpart normal variables, (y_1, y_2) are jointly related with a correlation of ρ_y.

Bivariate Lognormal Correlation

Law and Kelton [3] provide the following formulation for the bivariate lognormal correlation.

$\sigma_{y1y2} = E(y_1y_2) - E(y_1)E(y_2)$ = covariance between y_1 and y_2

$\rho_{y1y2} = \sigma_{y1y2}/\sigma_{y1}\sigma_{y2}$ = correlation between y_1 and y_2

$\rho_{x1x2} = \left[\exp\left(\sigma_{y1y2}\right) - 1/\left\{\left[\exp\left(\sigma_{y1}{}^2 - 1\right)\right]\left[\exp\left(\sigma_{y2}{}^2 - 1\right)\right]\right\}^{0.5}\right.$ = correlation between x_1 and x_2

Bivariate Lognormal Designation

The designation of the bivariate lognormal distribution is the following:

$$(x_1, x_2) \sim BVLN\left(\mu_{y1}, \mu_{y2}, \sigma_{y1}, \sigma_{y2}, \rho_y\right)$$

10.5 Bivariate Normal Distribution

The bivariate normal is a distribution with variables y_1, y_2 that are jointly related with a correlation ρ_y, and whose marginal distributions are normally distributed as shown below:

$$y_1 \sim N\left(\mu_{y1}, \sigma_{y1}{}^2\right)$$

$$y_2 \sim N\left(\mu_{y2}, \sigma_{y2}{}^2\right)$$

The common designation of the bivariate normal variable is,

$$(y_1, y_2) \sim BVN\left(\mu_{y1}, \mu_{y2}, \sigma_{y1}, \sigma_{y2}, \rho_y\right)$$

Recall,

$$y_1 = \ln(x_1)$$
$$y_2 = \ln(x_2)$$

10.6 Bivariate (Zero-Mean) Normal Distribution

The bivariate (zero-mean) normal distribution is the same as the bivariate normal distribution except the means of the jointly related variables, (y_1`, $y_2$`), are zero. The relations between (y_1`, $y_2$`) and (y_1, y_2) are below:

$$y_1\text{`} = \left(y_1 - \mu_{y1}\right)$$
$$y_2\text{`} = \left(y_2 - \mu_{y2}\right)$$

The standard deviations and the correlation between y_1` and $y_2$` are the same as that between y_1 and y_2. Hence,

$$\sigma_{y1\text{`}} = \sigma_{y1}$$
$$\sigma_{y2\text{`}} = \sigma_{y2}$$
$$\rho_{y\text{`}} = \rho_y$$

The common designation for the bivariate zero-mean normal distribution is below:

$$(y_1\text{`}, y_2\text{`}) \sim \text{BVN}\left(0, 0, \sigma_{y1}, \sigma_{y2}, \rho_y\right)$$

Bivariate (Standard) Normal Distribution

The bivariate (standard) normal distribution is the same as the bivariate (zero-mean) normal distribution except the standard deviations of the jointly related variables, (z_1, z_2), are equal to one. The relations between (z_1, z_2) and (y_1`, $y_2$`) are below:

$$z_1 = y_1\text{`}/\sigma_{y1}$$
$$z_2 = y_2\text{`}/\sigma_{y2}$$

The standard deviations and the correlation between z_1 and z_2 are the following:

$$\sigma_{z1} = 1.0$$
$$\sigma_{z2} = 1.0$$
$$\rho = \rho_z = \rho_y$$

and the notation for this distribution is below:

$$(z_1, z_2) \sim BVN(0, 0, 1, 1, \rho)$$

Example 10.1 An analyst has n paired data (x_{1i}, x_{2i}) for i = 1 to n whose range is quite large and suspects the data is distributed as a bivariate lognormal. To overcome the large range of values, the data are converted in the following way:

$$y_{1i} = \ln(x_{1i}) \text{ and } y_{2i} = \ln(x_{2i}) \text{for } i = 1 \text{ to } n$$

where ln is the natural log. Applying a statistical analysis to the paired data of (y_{1i}, y_{2i}), the average, $\bar{y}$, standard deviation, s, and correlation, r, statistics are computed and are as follows:

$$\bar{y}_1 = 5.0$$
$$\bar{y}_2 = 8.0$$
$$s_{y1} = 2.0$$
$$s_{y2} = 3.0$$
$$r_{y1y2} = 0.60$$

Since, (x_1, x_2) are assumed as bivariate lognormal, (y_1, y_2) are bivariate normal. Thereby,

$$(x_1, x_2) \sim BVLN(\ 5.8, 2, 3, 0.6)$$
$$(y_1, y_2) \sim BVN(\ 5, 8, 2, 3, 0.6)$$

10.7 Deriving F(x_1, x_2)

To find F(x_{1o}, x_{2o}) = P($x_1 \leq x_{1o} \cap x_2 \leq x_{2o}$), the four steps below are followed:

1. Derive the corresponding normal variables as below:

$$y_{1o} = \ln(x_{1o})$$
$$y_{2o} = \ln(x_{2o})$$

2. Compute the standard normal variables (k_1, k_2):

$$k_1 = \left[y_{1o} - \mu_{y1} \right] / \sigma_{y1}$$
$$k_2 = \left[y_{2o} - \mu_{y2} \right] / \sigma_{y2}$$

3. With interpolation, get F(k_1, k_2) from Table 8.1 with ρ_y.
4. Finally,

$$F(x_{1o}, x_{2o}) = F(k_1, k_2)$$

Example 10.2 As an exercise, use the results from Example 10.1, where (x_1, x_2) are bivariate lognormal with BVLN(5, 8, 2, 3, 0.6), and apply the equations between the bivariate normal and bivariate lognormal. The following results are found:

$$(y_1, y_2) \sim \text{BVN}(5, 8, 2, 3, 0.6)$$
$$(y_1`, y_2`) \sim \text{BVN}(0, 0, 2, 3, 0.6)$$
$$(z_1, z_2) \sim \text{BVN}(0, 0, 1, 1, 0.6)$$
$$\mu_{x1`} = e^{4/2} = 7.389$$
$$\mu_{x2`} = e^{9/2} = 90.017$$
$$\sigma_{x1`} = \left[e^4\left(e^4 - 1\right)\right]^{0.5} = 54.096$$
$$\sigma_{x2`} = \left[e^9\left(e^9 - 1\right)\right]^{0.5} = 8102.584$$
$$\sigma_{y1`y2`} = 0.6 \times 2 \times 3 = 3.6$$
$$\rho_{x1'x2'} = \left(e^{3.6} - 1\right) / \left[\left(e^4 - 1\right)\left(e^9 - 1\right)\right]^{0.5} = 0.054$$
$$\mu_{x1} = 7.389 \times e^5 = 1096.6$$
$$\mu_{x2} = 90.017 \times e^8 = 268,337.3$$
$$\sigma_{x1} = 54.096 \times \left[e^{10}\right]^{0.5} = 8028.5$$
$$\sigma_{x2} = 8102.584 \times \left[e^{16}\right]^{0.5} = 24153462.2)$$
$$\rho_{x1x2} = 0.054$$

Example 10.3 Assume the results from Examples 10.1 and 10.2 where the analyst wants to find the cumulated probability of F(x_{1o}, x_{2o}) = F(1000, 300,000). To obtain the results, the following four steps are taken:

1. Convert (x_{1o}, x_{2o}) to (y_{1o}, y_{2o}) as below:
   y = ln(x) and thereby,
   (y_{1o}, y_{2o}) = (6.91, 12.61)

2. Transfrm y to y\` = (y – μ_y), and obtain:
   (y_{1o}\`, $y_{2o}$\`) = (1.91, 4.61)
3. Convert y\` to standard normal variates by k = (y\`/σ_y) with results listed below:
   (k_1, k_2) = (0.95, 1.54)
4. Apply Table 8.1 with ρ = 0.6 and find $F_k(0.95, 1.54) \approx 0.81$, Thereby, F_x(1000, 300,000) ≈ 0.81.

Example 10.4 Table 10.1 shows that F(x_1, x_2) = F(1096.6, 13, 359.7) = 0.64 from the bivariate lognormal distribution where (x_1, x_2) ~ BVLN(5, 8, 2, 3, 0.6). The table also lists the same result for the bivariate normal distribution of F(k_1, k_2) = F(1.0, 0.5) = 0.64 where (k_1, k_2) ~ BVN(0, 0, 1, 1, 0.6). Below shows how the results are obtained.

At $k_1 = 1.0$: $y_1 = \mu_{y1} + k_1\sigma_{y1} = 5.0 + 1.0 \times 2.0 = 7.0$
$x_1 = e^{y1} = e^{7.0} = 1096.6$
At $k_2 = 0.5$: $y_2 = \mu_{y2} + k_2\sigma_{y2} = 8.0 + 0.5 \times 3.0 = 9.5$
$x_2 = e^{y2} = e^{9.5} = 13{,}359.7$

Table 10.1 gives the cumulative probability, F(x_1, x_2) for the bivariate lognormal variables where, (x_1, x_2) ~ BVLN(5, 8, 2, 3, 0.6). The table also shows how the probabilities are obtained from the standard bivariate lognormal variables when the correlation of the standard normal is ρ = 0.6. The latter results are taken from Chap. 8 when ρ = 0.6, and (k_1, k_2) ~ BVN(0, 0, 1, 1, 0.6).

Example 10.5 Assume the analyst using the data from Example 10.1 wants to find some representative values of (x_1,x_2) where the cumulative probability, F(x_1, x_2), is near 0.90. Using Table 10.1 with ρ = 0.6, note the following for F(k_1, k_2):

$$F(1.5, 2.0) = 0.92$$
$$F(2.0, 1.5) = 0.92$$
$$F(1.5, 1.5) = 0.89$$

Table 10.1 is a corresponding listing of F(x_1, x_2) when ρ = 0.6. Applying the conversion from k to y\` to y to x yields the following for (x_1, x_2):

At (k_1, k_2) = (1.5, 2.0): F(2980, 1,202,604) = 0.92
At (k_1, k_2) = (2.0, 1.5): F(8103, 268,337) = 0.92
At (k_1, k_2) = (1.5, 1.5): F(2980, 268,337) = 0.89

Noting the wide spread that the lognormal distribution gives, the analyst might select:
$(x_1, x_2) = (3{,}000, 300{,}000)$.

Table 10.1 $F(x_1, x_2)$ for selected values of $(x_1, x_2) \sim BVLN(5.0, 8.0, 2.0, 3.0, 0.6)$, and $F(k_1, k_2)$ for values of $(k_1, k_2) \sim BVN(0, 0, 1, 1, 0.6)$

ρ =	0.6													
k2/k1	−3	−2.5	−2	−1.5	−1	−0.5	0	0.5	1	1.5	2	2.5	3	×2
3	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.98	0.99	1.00	24,154,952
2.5	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.93	0.97	0.99	0.99	5,389,698
2	0	0.01	0.02	0.07	0.16	0.31	0.50	0.69	0.84	0.92	0.96	0.97	0.98	1,202,604
1.5	0	0.01	0.02	0.07	0.16	0.31	0.49	0.68	0.81	0.89	0.92	0.93	0.93	268,337
1	0	0.01	0.02	0.07	0.16	0.30	0.48	0.64	0.76	0.81	0.83	0.84	0.84	59,874
0.5	0	0.01	0.02	0.06	0.15	0.28	0.43	0.56	0.64	0.68	0.69	0.69	0.69	13,359
0	0	0.01	0.02	0.06	0.14	0.24	0.35	0.43	0.48	0.49	0.50	0.50	0.50	2981
−0.5	0	0.01	0.02	0.05	0.11	0.18	0.24	0.28	0.30	0.31	0.31	0.31	0.31	665
−1	0	0	0.02	0.04	0.07	0.11	0.14	0.15	0.16	0.16	0.16	0.16	0.16	148
−1.5	0	0	0.01	0.02	0.04	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07	33
−2	0	0	0	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	7
−2.5	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	2
−3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
×1	0	1	3	7	20	55	148	403	1097	2981	8103	22,026	59,874	

10.8 Summary

The relation between the lognormal and normal distributions is again refreshed to form the conversion between the bivariate lognormal and the bivariate normal. When an analyst has data from a bivariate lognormal distribution, some computational maneuvering transforms the distribution to a bivariate normal. The tables from Chap. 8 (Bivariate Normal) are then applied to compute the probabilities needed on the bivariate lognormal data. Examples are presented to guide the user on the application.

References

1. Thomopoulos, N. T., & Longinow, A. C. (1984). Bivariate lognormal probability distribution. *Journal of Structural Engineering, 110*, 3045–3049.
2. Thomopoulos, N. T., & Johnson, A. C. (2004). *Some Measures on the Standard Bivariate Lognormal Distribution*. Proceedings of the Decision Sciences Institute. pp 1721–1726.
3. Law, A. M., & Kelton, D. W. (2000). *Simulation, modeling and analysis*. Boston: McGraw Hill.

Recommended Reading

Aitchison, J., & Brown, J. (1969). *The lognormal distribution*. Cambridge: Cambridge University Press.

Beyer, W. H. (1968). *Handbook of Tables for Probability and Statistics*. Cleveland: Chemical Rubber Co..

Broadbent, S. (1956). Lognormal approximation to products and quotients. *Biometrika, 43*, 404–417.

Brown, R. G. (1959). *Smoothing, forecasting and prediction of discrete time series*. Engelwood Cliffs: Prentice Hall.

Crow, E., & Shinizu, K. (1988). *Lognormal distribution: Theory and applications*. New York: Marcel Dekker.

Hasting, N. A. J., & Peacock, J. B. (1974). *Statistical distributions*. New York: Wiley & Sons.

Hines, W. W., Montgomery, L. D. C., Goldsman, D. M., & Burror, C. M. (2003). *Probability and statistics for engineers*. New York: Wiley & Sons.

Law, A., & Kelton, W. (2000). *Simulation, modeling and analysis*. Boston: McGraw Hill.

Robson, D. S., & Whitlock, J. W. (1964). Estimation of a truncation point. *Biometrical, 51*, 33–39.

Schneider, J. (1986). *Truncated and censored samples from normal populations*. New York: Marcel Dekker.

Thomopoulos, N. T. (2013). *Essentials of Monte Carlo simulation*. New York: Springer.

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1

Index

N. T. Thomopoulos, *Probability Distributions*,
https://doi.org/10.1007/978-3-319-76042-1