Paulo Eduardo Oliveira

Asymptotics for Associated Random Variables

Springer

Asymptotics for Associated Random Variables

Paulo Eduardo Oliveira

Asymptotics for Associated Random Variables

Springer

Paulo Eduardo Oliveira
Department of Mathematics, CMUC
University of Coimbra
Coimbra, Portugal

ISBN 978-3-642-25531-1 e-ISBN 978-3-642-25532-8
DOI 10.1007/978-3-642-25532-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011945842

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Júlia, Sofia and Raúl for their patience…

To Fernanda and Graciano who started off everything…

Preface

The control of dependence between random variables has always been an object of interest and concern to probabilists and statisticians. Several ways to control this dependence have been introduced, and this book concerns the notion of association of random variables. Association and some other positive dependence notions were introduced in the mid 1960s. The interest on these dependence notions came from models where monotone transformations were concerned. Association and generally positive dependence received little attention of the probabilistic and statistics community, but the interest increased in more recent years. Therefore, a rather complete body of theory was constructed covering the traditional probabilistic topics and, eventually, studying statistics based on dependent samples. Although this increased interest, characterizations and results remained essentially scattered in the literature published in different journals. So, it was time to bring together the bulk of these results, presenting the theory in a unified way, explaining relations and implications of the results. Such a challenge may be taken in, at least, two directions: either going to the more subtle and at the peak of the wave results or introducing the notions from a more elementary approach. In this book this later choice is taken. In this way, the attention of the reader will not be diverted from the essential point, which is the peculiarities of positive dependence and the way to get around the difficulties due this dependence structure. This does not mean that advanced or recent results are not included. On the contrary, the text is organized in a manner such that, starting from this elementary approach, and progression is made towards recent results on the asymptotics of sequences of associated random variables. This book is addressed to researchers in probability and statistics, with a special concern on people interested in kernel estimation methods. It will also be of interest to graduate students in those areas. The book could also be used as a reference on association on a course covering dependent variables and their asymptotics.

After presenting the notion of association of random variables, together with a few variations on this definition, an account of inequalities hold for this dependence structure is given. Many of these inequalities are extended versions of counterparts that are well known for independent random variables, while others are really specific to this dependence. Most of these inequalities were developed as a means to prove or characterize extension of the classical results to associated variables. These inequalities are presented as a chapter in order to have most of the basic tools available once and for all. The role that the covariance structure plays while controlling asymptotic results for associated random variables will become more explicit. Again, throwing these into a separate chapter would contribute to leave the concentration of the reader directed in the appropriate direction when dealing with the proofs of the later results. With these tools in hand, the text concentrates on the convergence, almost sure or in distribution, and for this later with a special interest on functional results, of sequences of associated random variables. At each of these chapters we include a reference to the asymptotics of kernel estimators based on associated samples.

Writing this book comes a result of work developed through many years during which I had the opportunity to discuss and collaborate with a few colleagues. From these, I would like to leave a special acknowledgement to Pierre Jacob and Charles Suquet for the collaboration, many discussions and friendship throughout the years. Finally, I wish to express my gratitude to my colleague Carlos Tenreiro who helped improving an earlier version of this text.

Coimbra, Portugal Paulo Eduardo Oliveira

Contents

Chapter 1
Positive Dependence

Abstract This is an introductory chapter where the notions of positive dependence, with a particular interest in association, will be introduced. A few useful alternative characterizations of association will be studied. Basic properties are proved showing that, although some vagueness in its definition, association is reach enough to be an interesting structure. We include examples and applications that are illustrative of the scope and usefulness of this dependence notion. Association is natively a dependence structure on random variables, so we will discuss its extension to some more abstract spaces, highlighting the connections with the order structure of the underlying space. The chapter is concluded with a reference to other types of positive dependence and their relations. A brief discussion on negative dependence notions concludes the chapter.

1.1 Introduction

Some ideas of what is now known as positive dependence were explored in the literature for particular and relevant examples, without explicit referral to any particular dependence structure. As an example, statistical procedures were used involving pairs of dependent random variables trying to detect if large values of one variable tend to be associated with large values of the other variable, as is the case for the classical tests based on correlation ranks, or Kendall's t-statistic, as discussed for example in Hoeffding [46], Lehmann [57] or Blum, Kiefer and Rosenblatt [18]. Most of such procedures explore implicitly some notion of positive dependence, meaning with this that some kind of measuring if large values of one variable tend to be associated to large values of another variable, or, to put it in a different way, if random variables tend to increase or decrease simultaneously. This was explicitly noticed by Lehmann [58], who was the first reference to attempt a formalization of these kind of dependence notions and to explore a few of its properties, trying to extend the scope of applicability of the behaviour of the above-mentioned tests beyond the particular distributions considered. Lehmann [58] was essentially interested in the distribution of pairs of random variables and introduced a dependence notion well adapted for this framework, noticing immediately some consequences about the covariance between the random variables. So, it was only natural that positive dependence notions appeared in a bivariate context. Later, this idea was

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8_1, © Springer-Verlag Berlin Heidelberg 2012

extended to multivariate distributions by Esary, Proschan and Walkup [36] to the notion of association. This new dependence structure showed to have more interesting properties and application scope still broad enough to attract the interest of probabilists and, somewhat later, of statisticians. The main ideas behind these positive dependence notions is the global tendency for increasingness or decreasingness of the family of random variables and the fact that approximation to independence is completely characterized by the covariance structure. As it will be proved later, association appears naturally in models that rely on monotone transformations, such us reliability and survival analysis, thus contributing for the interest in applications of these models and results derived. Most of the early results and applications where in this direction, as is described in Barlow and Proschan [5], one of the few references to explore in a more systematic way the positive dependence notions in the decade that followed the contributions by Lehmann [58] and Esary, Proschan and Walkup [36]. In the meanwhile, this dependence concept received some attention in statistical mechanics and percolation theory, being known as the FKG inequalities following the contribution by Fortuin, Kasteleyn and Ginibre [40]. Here the motivation for the interest in this kind of dependence appeared from properties satisfied by the Hamiltonian describing the interaction between two bodies in Ising ferromagnet spin models. The flow of literature eventually increased constructing a full body of results, either with a probabilistic flavour or with a more statistical application in mind.

Some other positive dependence notions were considered trying to respond to specific difficulties, or to extend the scope of applicability of some of the results or even trying to characterize independence from the covariances of the random variables. A few notions remained concentrated on bivariate distributions, while many others tried to deal with multivariate distributions, extending the more relevant properties. Contributions made by Block, Savits and Shaked [17], Joag-Dev [53] and Joag-Dev and Proschan [54] are examples of developments in this direction. Eventually some approaches using a more general framework appeared, trying to understand more deeply what these dependence structures really mean. Shaked [92, 93] tried to build a general approach to positive dependence notions, while Lindqvist [60] extended the approach to abstract space-valued random variables, highlighting the role of the order structure of the base space. Another direction of development reversed the direction of the inequalities, thus defining negative dependence notions. At first sight this seems to define an easier framework, as almost all the bounds known for independent variables still hold for negative dependence. Nevertheless, many of the asymptotic results do not follow directly from the independent case, so there was room for developing some theory. We will introduce and study a few results on this direction, but for the development of the asymptotic theory, we will remain concentrated in positive dependence notions. Many of the approaches and methodologies to be used throughout this text can be easily adapted to obtain characterizations about negatively dependent variables. Of course, in this later case one should expect to find better inequalities that, in a sense, would make proofs somewhat easier to carry. Another recent direction of extension emerged in more recent years, starting from Doukhan and Louhichi [34]. This was a consequence of some inequalities proved in the meanwhile for associated random variables that showed the importance of covariances in the characterization of independence. We chose not to include these later developments in this text, although these weak forms of dependence show a sufficiently rich structure allowing one to prove quite a few asymptotic results.

1.2 Basic Definitions and Examples

This section introduces the basic notions of positive dependence to be explored throughout this text and establishes a few of the fundamental properties. Before embarking in the main subject, we introduce some really basic notation that will be used throughout this text. We begin with the simplest kind of positive dependence, introduced by Lehmann [58].

Definition 1.1 Two random variables X and Y are said to be *positively quadrant dependent* (*PQD*) if, for all $x, y \in \mathbb{R}$,

$$H(x, y) = \mathbf{P}(X > x, Y > y) - \mathbf{P}(X > x)\mathbf{P}(Y > y) \geq 0.$$

In order to rewrite $H(x, y)$ in a convenient way, we denote, as usual, by $\mathbb{I}_A$ the indicator function of a set A, that is, the function such that $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$ if $x \notin A$. This will help us on rewriting $H(x, y)$ using distribution functions:

$$\begin{aligned} H(x, y) &= \mathbf{P}(X > x, Y > y) - \mathbf{P}(X > x)\mathbf{P}(Y > y) \\ &= \mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(X), \mathbb{I}_{(y,+\infty)}(Y)\big) \\ &= \mathrm{Cov}\big(1 - \mathbb{I}_{(x,+\infty)}(X), 1 - \mathbb{I}_{(y,+\infty)}(Y)\big) \\ &= \mathrm{Cov}\big(\mathbb{I}_{(-\infty,x]}(X), \mathbb{I}_{(-\infty,y]}(Y)\big) \\ &= \mathbf{P}(X \leq x, Y \leq y) - \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y). \end{aligned} \tag{1.1}$$

Example 1.2 Let X and Y be jointly distributed as $\mathbf{P}(X = 0, Y = 0) = p_1$, $\mathbf{P}(X = 0, Y = 1) = p_2$, $\mathbf{P}(X = 1, Y = 0) = p_3$ and $\mathbf{P}(X = 1, Y = 1) = p_4$, where $p_1 + p_2 + p_3 + p_4 = 1$. Using the representation of H with distribution functions, it is easily seen that X and Y are PQD if and only if $\mathbf{P}(X \leq 0, Y \leq 0) - \mathbf{P}(X \leq 0)\mathbf{P}(Y \leq 0) = p_1 - (p_1 + p_2)(p_1 + p_3) = p_1 p_4 - p_2 p_3 \geq 0$.

To describe the next example, let us introduce some notation:

1. Given $x_1, \ldots, x_n \in \mathbb{R}$, denote $x_1 \vee \cdots \vee x_n = \max(x_1, \ldots, x_n)$.
2. Given $x_1, \ldots, x_n \in \mathbb{R}$, denote $x_1 \wedge \cdots \wedge x_n = \min(x_1, \ldots, x_n)$.

Example 1.3 Let T_1, T_2, T_3 be independent random variables with common distribution function F and define $X = T_1 \vee T_2$, $Y = T_2 \vee T_3$. Then,

$$\mathbf{P}(X \leq x, Y \leq y) = F(x)F(y)F(x \wedge y), \qquad \mathbf{P}(X \leq x) = \mathbf{P}(Y \leq x) = F^2(x),$$

and, for every $x, y \in \mathbb{R}$,

$$\begin{aligned} H(x,y) &= F(x)F(y)F(x \wedge y) - F^2(x)F^2(y) \\ &= F(x)F(y)F(x \wedge y)\big(1 - F(x \vee y)\big) \geq 0. \end{aligned}$$

Thus, the random variables X and Y are PQD.

The following result expresses the covariance between two random variables using $H(x,y)$ and provides one of the key links towards the extension of this dependence notion.

Theorem 1.4 (Hoeffding formula) *Let X and Y be square-integrable random variables. Then*

$$\operatorname{Cov}(X,Y) = \int_{\mathbb{R}^2} H(x,y)\,dx\,dy. \tag{1.2}$$

Proof Let (X_1, Y_1) and (X_2, Y_2) be independent random vectors with the same distribution as (X, Y). Then, by simple computation,

$$\begin{aligned} &\mathrm{E}(X_1Y_1) - \mathrm{E}(X_1)\mathrm{E}(Y_1) \\ &\quad = \frac{1}{2}\mathrm{E}\big((X_1 - X_2)(Y_1 - Y_2)\big) \\ &\quad = \frac{1}{2}\mathrm{E}\bigg(\int_{\mathbb{R}^2}\big(\mathbb{I}_{(-\infty,X_1]}(x) - \mathbb{I}_{(-\infty,X_2]}(x)\big)\big(\mathbb{I}_{(-\infty,Y_1]}(y) - \mathbb{I}_{(-\infty,Y_2]}(y)\big)\,dx\,dy\bigg). \end{aligned}$$

As the random variables are assumed to be square integrable, we can use Fubini's theorem to interchange the expectation with the integration above, to find

$$\begin{aligned} &\mathrm{E}(X_1Y_1) - \mathrm{E}(X_1)\mathrm{E}(Y_1) \\ &\quad = \frac{1}{2}\int_{\mathbb{R}^2} \mathrm{E}\big(\mathbb{I}_{(-\infty,X_1]}(x)\mathbb{I}_{(-\infty,Y_1]}(y)\big) - \mathrm{E}\big(\mathbb{I}_{(-\infty,X_1]}(x)\mathbb{I}_{(-\infty,Y_2]}(y)\big) \\ &\qquad - \mathrm{E}\big(\mathbb{I}_{(-\infty,X_2]}(x)\mathbb{I}_{(-\infty,Y_1]}(y)\big) + \mathrm{E}\big(\mathbb{I}_{(-\infty,X_2]}(x)\mathbb{I}_{(-\infty,Y_2]}(y)\big)\,dx\,dy \\ &\quad = \int_{\mathbb{R}^2} \mathbf{P}(X > x, Y > y) - \mathbf{P}(X > x)\mathbf{P}(Y > y)\,dx\,dy. \end{aligned}$$ □

The following result on covariances is now obvious.

Corollary 1.5 *Let X and Y be PQD random variables. Then* $\operatorname{Cov}(X,Y) \geq 0$.

It is also evident that, for PQD variables, their covariance completely characterizes independence.

Corollary 1.6 *Let X and Y be PQD random variables. X and Y are independent if and only if* $\operatorname{Cov}(X,Y) = 0$.

The following important example, as it gives a complete characterization for Gaussian random vectors, is now a simple consequence of the above results.

Example 1.7 Let (X, Y) be a random vector with bivariate normal distribution with correlation ρ. If X and Y are PQD, then $\mathrm{Cov}(X, Y) = \rho(\mathrm{Var}(X)\,\mathrm{Var}(Y))^{1/2} \geq 0$, that is, $\rho \geq 0$. The converse is also true but requires a more elaborate argument and is deferred to Theorem 1.35, where a stronger result is proved.

The following results show how we can proceed to construct pairs of variables that are PQD. The complete argument uses Corollary 1.5 in two steps.

Proposition 1.8 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \ldots, n$, X_i and Y_i are PQD. Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be such that, for each $i = 1, \ldots, n$, when considered as functions of the ith coordinate alone, they are both nondecreasing or both nonincreasing, and let $X = f(X_1, \ldots, X_n)$ and $Y = g(Y_1, \ldots, Y_n)$. Then* $\mathrm{Cov}(X, Y) \geq 0$.

Proof We proceed by induction on n. The case $n = 1$ is immediate: assuming f and g to be nondecreasing, we have

$$\begin{aligned} H(x, y) &= \mathbf{P}(X \leq x, Y \leq y) - \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y) \\ &= \mathbf{P}\big(X_1 \leq f^{\leftarrow}(x), Y_1 \leq g^{\leftarrow}(y)\big) - \mathbf{P}\big(X_1 \leq f^{\leftarrow}(x)\big)\mathbf{P}\big(Y_1 \leq g^{\leftarrow}(y)\big), \end{aligned}$$

where $f^{\leftarrow}$ and $g^{\leftarrow}$ are the generalized inverses of f and g, respectively (see Appendix C, page 181). Thus, $H(x, y) \geq 0$, so the variables X and Y are PQD, and from Corollary 1.5 it follows that $\mathrm{Cov}(X, Y) \geq 0$. Assume now that the proposition is true for $n - 1$ variables and define

$$f^*(x_2, \ldots, x_n) = \mathrm{E} f(X_1, x_2, \ldots, x_n) \quad \text{and} \quad g^*(x_2, \ldots, x_n) = \mathrm{E} g(Y_1, x_2, \ldots, x_n).$$

These functions of $n - 1$ variables share the monotonicity properties in each of the variables $x_2, \ldots, x_n$ as the functions f and g. Thus, by the induction hypothesis, the covariance with respect to the distributions of $(X_2, Y_2), \ldots, (X_n, Y_n)$ is nonnegative, that is,

$$\mathrm{Cov}_{(X_2,Y_2),\ldots,(X_n,Y_n)}\big(f^*(X_2, \ldots, X_n), g^*(Y_2, \ldots, Y_n)\big) \geq 0,$$

to emphasize the distributions with respect to which we are integrating. Now, taking into account the independence of (X_1, Y_1) from $(X_2, Y_2), \ldots, (X_n, Y_n)$, we may write, using the same indexing for expectation as done for the covariances,

$$\begin{aligned} &\mathrm{Cov}_{(X_1,Y_1),\ldots,(X_n,Y_n)}(X, Y) \\ &\quad = \mathrm{E}_{(X_1,Y_1),\ldots,(X_n,Y_n)}(XY) - \mathrm{E}_{(X_1,Y_1),\ldots,(X_n,Y_n)}(X)\mathrm{E}_{(X_1,Y_1),\ldots,(X_n,Y_n)}(Y) \\ &\quad = \mathrm{E}_{(X_1,Y_1)}\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(XY) \\ &\qquad - \mathrm{E}_{(X_1,Y_1)}\big(\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X)\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(Y)\big) \\ &\qquad + \mathrm{E}_{(X_1,Y_1)}\big(\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X)\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(Y)\big) \\ &\qquad - \mathrm{E}_{(X_1,Y_1)}\big(\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X)\big)\mathrm{E}_{(X_1,Y_1)}\big(\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(Y)\big) \\ &\quad = \mathrm{E}_{(X_1,Y_1)}\big(\mathrm{Cov}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X, Y)\big) \\ &\qquad + \mathrm{Cov}_{(X_1,Y_1)}\big(\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X), \mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(Y)\big). \end{aligned}$$

Now, from the induction hypothesis, the first term on the right is nonnegative. For the second, both $\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(X)$ and $\mathrm{E}_{(X_2,Y_2),\ldots,(X_n,Y_n)}(Y)$ are nondecreasing functions of (X_1, Y_1), and so, as for the case $n = 1$, they are PQD, and their covariance is also nonnegative. □

We can now complete the second step proving that the construction on the previous proposition produces PQD variables.

Theorem 1.9 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \ldots, n$, X_i and Y_i are PQD. Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be such that, for each $i = 1, \ldots, n$, when considered as functions of the ith coordinate alone, they are both nondecreasing or both nonincreasing, and let $X = f(X_1, \ldots, X_n)$ and $Y = g(Y_1, \ldots, Y_n)$. Then X and Y are PQD.*

Proof For all $x, y \in \mathbb{R}$, let $X^* = \mathbb{I}_{(-\infty,x]}(X)$ and $Y^* = \mathbb{I}_{(-\infty,y]}(Y)$. The transformations used to construct X^* and Y^* have the same monotonicity properties as f and g, and so, by the previous proposition,

$$\mathrm{Cov}\bigl(X^*, Y^*\bigr) = \mathbf{P}(X \le x, Y \le y) - \mathbf{P}(X \le x)\mathbf{P}(Y \le y) \ge 0,$$

that is, X and Y are PQD. □

A useful and obvious extension of this result is the following.

Corollary 1.10 *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \ldots, n$, X_i and Y_i are PQD. Let U and V be independent and independent of $(X_1, Y_1), \ldots, (X_n, Y_n)$. Let $f, g : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be such that, for each $i = 2, \ldots, n+1$, when considered as functions of the ith coordinate alone, they are both nondecreasing or both nonincreasing, and let $X = f(U, X_1, \ldots, X_n)$ and $Y = g(V, Y_1, \ldots, Y_n)$. Then X and Y are PQD.*

Proof Just repeat the arguments of the proof of Theorem 1.9 taking into account the independence of U and V and of these with the remaining variables. □

This characterizations provide an easy way to construct examples of PQD random variables: all one has to do is apply monotone transformations with the same monotonicity direction. As an application of the previous results, we may prove the nonnegativity of Kendall's τ.

Corollary 1.11 *Let (X_1, Y_1) and (X_2, Y_2) be independent with the distribution of (X, Y). If X and Y are PQD, then the Kendall's τ is nonnegative.*

Proof Kendall's τ is defined as $\mathrm{Cov}(U, V)$, where $U = \mathrm{sgn}(X_2 - X_1)$ and $V = \mathrm{sgn}(Y_2 - Y_1)$, where $\mathrm{sgn}(x)$ represents the sign of x. Then, it is enough to verify that U and V are PQD. But this is a direct consequence of Theorem 1.9. □

Table 1.1 PQD but not associated random variables

x_1	x_2	$\mathbf{P}(X_1 = x_1, X_2 = x_2)$	x_1	x_2	$\mathbf{P}(X_1 = x_1, X_2 = x_2)$
0	0	$\frac{3}{14}$	1	0	$\frac{2}{14}$
0	1	$\frac{1}{14}$	1	1	$\frac{2}{14}$
0	2	$\frac{1}{14}$	1	3	$\frac{3}{14}$
0	3	$\frac{2}{14}$			

The drawback with positive quadrant dependence is due to its bivariate nature. In fact, it is quite clear from the definition that this dependence regards two given random variables and does not allow any manipulation concerning sequences of random variables, unless done pairwise. But then, this would quickly drive us into difficulties whenever more than two variables were involved, rendering proofs on results about partial sums, for example, rather difficult to handle and requiring extra assumptions to take care of higher-order joint distributions. There are several ways to extend positive quadrant dependence, a few of which have received some interest in the literature, and some of these will be referred in Sect. 1.5. The extension that proved to be most successful, because it allows for a sufficiently rich theoretical body and has a wide scope of applicability, has been what is known as (positive) association, introduced by Esary, Proschan and Walkup [36].

Definition 1.12 The random variables $X_1, \ldots, X_n$ are *associated* if, given two coordinatewise nondecreasing functions $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$,

$$\mathrm{Cov}\big(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)\big) \geq 0 \tag{1.3}$$

whenever the covariance exists.

A sequence of random variables X_n, $n \in \mathbb{N}$, is *associated* if, for every $n \in \mathbb{N}$, the family of variables $X_1, \ldots, X_n$ is associated.

Naturally, we may replace nondecreasing functions by nonincreasing functions in the definition above.

Remark 1.13 It is obvious that if X and Y are associated, they are also PQD. The converse is generally not true as illustrated by the following example from Joag-Dev [53].

Example 1.14 Consider discrete variables X_1 and X_2 with joint distribution characterized by Table 1.1. It is a simple matter of routine to verify that these random variables are indeed PQD. In fact, taking into account (1.1), for this joint distribution it is enough to verify the nonnegativity of $H(x_1, x_2) = \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2) - \mathbf{P}(X_1 \leq x_1)\mathbf{P}(X_2 \leq x_2)$ for $(x_1, x_2) = (0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 3)$. These values are given in Table 1.2.

Define now $f(x_1, x_2) = \mathbb{I}_{(0,+\infty)}(x_1)$ and $g(x_1, x_2) = \mathbb{I}_{(1,+\infty)}(x_2)$. These functions are obviously coordinatewise nondecreasing, and

Table 1.2 Verification that the distribution in Table 1.1 is PQD

x_1	x_2	$H(x_1, x_2)$	x_1	x_2	$H(x_1, x_2)$
0	0	$\frac{1}{28}$	1	0	0
0	1	0	1	1	0
0	2	$\frac{1}{28}$	1	3	0
0	3	0			

$$\begin{aligned}&\mathrm{Cov}\big(f(X_1, X_2), g(X_1, X_2)\big)\\&\quad = \mathbf{P}(X_1 > 0, X_2 > 1) - \mathbf{P}(X_1 > 0)\mathbf{P}(X_2 > 1) = -\frac{1}{28},\end{aligned}$$

so X_1 and X_2 are not associated.

The following property is an immediate consequence of the definition of association.

Theorem 1.15 *Let $X_1, \ldots, X_n$ be associated random variables and consider coordinatewise nondecreasing functions $f_1, \ldots, f_k : \mathbb{R}^n \longrightarrow \mathbb{R}$. Then the random variables $Y_1 = f_1(X_1, \ldots, X_n), \ldots, Y_k = f_k(X_1, \ldots, X_n)$ are associated.*

Remark 1.16 In the statement above we can require all the functions to be nonincreasing for the result to still hold.

The properties proved above for PQD variables that depend only on the nonnegativeness of the covariances are immediately valid for associated variables. We state them here without proof, as this would be just repeating the already used arguments.

Theorem 1.17 *Let $X_1, \ldots, X_n$ be associated variables. These random variables are independent if and only if* $\mathrm{Cov}(X_i, X_j) = 0$, $i, j = 1, \ldots, n$, $i \neq j$.

For an extension of this independence characterization, see Corollary 2.2. The following result extends Theorem 1.9 and is proved in exactly the same way.

Theorem 1.18 *Let $X_1, \ldots, X_n$ be independent of the variables $Y_1, \ldots, Y_m$. Assume that $X_1, \ldots, X_n$ are associated variables and also that $Y_1, \ldots, Y_m$ are associated. Then the variables $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ are associated.*

It is now simple to generate new families of associated variables starting from a given set of associated random variables by applying monotone transformations with the same monotonicity direction. So, to illustrate such a procedure, repeating the constructions described in Example 1.3, we obtain associated variables if the initial ones are already associated. It would be interesting to be able to start the construction from independent variables, but this requires some more results.

Example 1.19 If X_n, $n \in \mathbb{N}$, are associated random variables, then the sequence of partial sums $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, is associated. This is an immediate consequence of Theorem 1.15.

Example 1.20 Given the random variables $X_1, \ldots, X_n$, define the ordered statistics $X_{k:n} =$ the kth smallest among $X_1, \ldots, X_n$. These order statistics are nondecreasing transformations of the $X_1, \ldots, X_n$, so if these are associated, the same holds for $X_{1:n}, \ldots, X_{n:n}$.

Example 1.21 One model that will be used to illustrate the usefulness of some assumptions later starts with a sequence of random variables Y_n, $n \in \mathbb{N}$, and $m \in \mathbb{N}$ fixed and defines $X_n = \max(Y_n, Y_{n+1}, \ldots, Y_{n+m})$. If the Y_n's are associated, so are the X_n's. We will show later that we can relax the assumption on the Y_n's.

1.3 Characterizations and Constructive Properties

In this section we will study a few alternative characterizations of association relaxing the family of functions for which we should verify the sign of the covariances.

Theorem 1.22 *The random variables $X_1, \ldots, X_n$ are associated if and only if, for every coordinatewise nondecreasing functions $\gamma_1, \gamma_2 : \mathbb{R}^n \longrightarrow \{0, 1\}$,*

$$\operatorname{Cov}\bigl(\gamma_1(X_1, \ldots, X_n), \gamma_2(X_1, \ldots, X_n)\bigr) \geq 0.$$

Proof If the variables are associated, the conclusion is immediate from the definition of association. To prove the other implication, let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be coordinatewise nondecreasing functions. From Hoeffding's formula (1.2) it follows that

$$\begin{aligned} &\operatorname{Cov}\bigl(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)\bigr) \\ &\quad = \int_{\mathbb{R}^n} \operatorname{Cov}(\mathbb{I}_{\{f(X_1,\ldots,X_n)>s\}}, \mathbb{I}_{\{g(X_1,\ldots,X_n)>t\}})\, ds\, dt. \end{aligned} \tag{1.4}$$

Now, as f is coordinatewise nondecreasing, the same is true for the function $\gamma_s(x_1, \ldots, x_n) = \mathbb{I}_{\{f(x_1,\ldots,x_n)>s\}}$ and likewise about $\gamma_t(x_1, \ldots, x_n) = \mathbb{I}_{\{g(x_1,\ldots,x_n)>t\}}$. Thus, the integrand in (1.4) is nonnegative for every $s, t \in \mathbb{R}$, so it follows that

$$\operatorname{Cov}\bigl(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)\bigr) \geq 0,$$

as required. □

Remark 1.23 Notice that a nondecreasing $\{0, 1\}$-valued function γ defined on $\mathbb{R}$ is of the form $\mathbb{I}_{[a,+\infty)}(x)$ or $\mathbb{I}_{(a,+\infty)}(x)$, for some $a \in \mathbb{R}$. If γ is defined on $\mathbb{R}^n$ for some $n \geq 2$ and is of the form $\mathbb{I}_{B_1 \times \cdots \times B_n}(x_1, \ldots, x_n)$ where each B_i is either $[a_i, +\infty)$ or $(a_i, +\infty)$, $a_i \in \mathbb{R}$, then γ is coordinatewise nondecreasing. However, for the case $n \geq 2$, there are coordinatewise nondecreasing $\{0, 1\}$-valued functions that are not of this form. For an example, take $B = \{(x_1, \ldots, x_n) : x_1 + \cdots + x_n \geq 0\}$ and $\gamma(x_1, \ldots, x_n) = \mathbb{I}_B(x_1, \ldots, x_n)$.

We may now prove that random variables are associated with themselves.

Theorem 1.24 *Every random variable X is associated with itself.*

Proof Taking into account the previous characterization of association, it suffices to verify that $\mathrm{Cov}(\gamma_1(X), \gamma_2(X)) \geq 0$ for all $\{0, 1\}$-valued nondecreasing functions γ_1 and γ_2. But, taking into account Remark 1.23, such functions are necessarily of the form $\gamma_1(x) = \mathbb{I}_{[a_1,+\infty)}(x)$ or $\gamma_1(x) = \mathbb{I}_{(a_1,+\infty)}(x)$, and the same representation holds for γ_2. Thus, we will have either $\gamma_1(x) \leq \gamma_2(x)$ for every $x \in \mathbb{R}$ or $\gamma_1(x) \geq \gamma_2(x)$ for every $x \in \mathbb{R}$. Assume, with loss of generality, that $\gamma_1(x) \leq \gamma_2(x)$. Then, $\gamma_1(x)\gamma_2(x) = \gamma_1^2(x) = \gamma_1(x)$ and

$$\begin{aligned}
&\mathrm{Cov}\big(\gamma_1(X), \gamma_2(X)\big) \\
&\quad = \mathrm{E}\big(\gamma_1(X)\gamma_2(X)\big) - \mathrm{E}\gamma_1(X)\mathrm{E}\gamma_2(X) \\
&\quad = \mathrm{E}\gamma_1(X) - \mathrm{E}\gamma_1(X)\mathrm{E}\gamma_2(X) = \mathrm{E}\gamma_1(X)\big(1 - \mathrm{E}\gamma_2(X)\big) \geq 0.
\end{aligned}$$

□

This result, together with Theorem 1.18, implies that independent variables are associated.

Corollary 1.25 *If X_n, $n \in \mathbb{N}$, are independent random variables, then they are associated.*

Proof According to Theorem 1.24, X_1 is associated with itself, the same holds for X_2, and these variables are independent. Thus, from Theorem 1.18 it follows that X_1 and X_2 are associated. Now X_3 is associated with itself and independent from (X_1, X_2), so Theorem 1.18 again implies that X_1, X_2 and X_3 are associated. Applying successively this argument, it follows that, for every $n \in \mathbb{N}$, the variables $X_1, \ldots, X_n$ are associated, so the corollary is proved. □

We can now extend the construction in Examples 1.19, 1.20 and 1.21 starting from independent random variables.

Example 1.26 If X_n, $n \in \mathbb{N}$, are independent random variables, the sequence of partial sums $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, is associated.

Example 1.27 Given the independent random variables $X_1, \ldots, X_n$, define the order statistics $X_{k:n} =$ the kth smallest among $X_1, \ldots, X_n$. These order statistics are associated random variables.

Example 1.28 Let Y_n, $n \in \mathbb{N}$, be independent random variables, $m \in \mathbb{N}$ be fixed, and define $X_n = \max(Y_n, Y_{n+1}, \ldots, Y_{n+m})$. Then the variables X_n, $n \in \mathbb{N}$, are associated.

Some more simple examples can be added to our list.

Example 1.29 Consider a moving average model $X_n = a_0\varepsilon_n + \cdots + a_q\varepsilon_{n-q}$, where ε_n are independent random variables, and $a_0, \ldots, a_q$ have the same sign. Then the variables X_n, $n \in \mathbb{N}$, are associated.

Example 1.30 Consider now an auto-regressive model $X_n = c_1 X_{n-1} + \cdots + c_q X_{n-q}$, $n \geq 1$, where $c_1, \ldots, c_q > 0$, and $X_{1-q}, \ldots, X_0$ are initial independent random variables. Then the variables X_n, $n \in \mathbb{N}$, are associated.

The family of functions for which we have to check (1.3) to prove the association of a given collection of variables may be reduced in different directions than those considered in Theorem 1.22. In this theorem we reduced the family of functions by imposing some restriction on the possible values for these functions. No regularity or smoothness was assumed on these functions. It is convenient to restrict the need of verifying (1.3) to suitable families of test functions that share some smoothness property. A first result on this direction is proved next, requiring the test functions to be bounded and continuous. We need a preparatory result to prove the announced characterization.

Lemma 1.31 *Let* $X_1, \ldots, X_n$ *be random variables and assume that, given any coordinatewise nondecreasing, continuous and bounded functions* $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$, *we have* $\mathrm{Cov}(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \geq 0$. *Then, for every coordinatewise nondecreasing and right continuous functions* $\gamma_1, \gamma_2 : \mathbb{R}^n \longrightarrow \{0, 1\}$ *it holds that* $\mathrm{Cov}(\gamma_1(X_1, \ldots, X_n), \gamma_2(X_1, \ldots, X_n)) \geq 0$.

Proof Let $\gamma_1 : \mathbb{R}^n \longrightarrow \{0, 1\}$ coordinatewise nondecreasing and right continuous. We will first show that γ_1 is the limit of a sequence f_k of coordinatewise nondecreasing, continuous and bounded functions. For this, define $A = \gamma_1^{-1}(\{1\})$. We start by proving that A is closed. Given a sequence $x_m \in A$, $m \in \mathbb{N}$, that is convergent to some z, construct a new sequence y_m as follows: for each $i = 1, \ldots, n$, choose the ith coordinate $y_{i,m}$ satisfying $z_i = \frac{1}{2}(x_{i,m} + y_{i,m})$ if $x_{i,m} < z_i$ and $y_{i,m} = x_{i,m}$ otherwise. Then, for every $m \in \mathbb{N}$, $x_{i,m} \leq y_{i,m}$, $z_i \leq y_{i,m}$, $i = 1, \ldots, n$, and $\|y_m - z\|_2 = \|x_m - z\|_2$. As γ_1 is nondecreasing, it follows that, for each $m \in \mathbb{N}$, $1 = \gamma_1(x_m) \leq \gamma_1(y_m)$, and thus $y_m \in A$. Moreover, the sequence y_m is lexicographically nonincreasing (that is, every coordinate of y_{m+1} is less or equal to the corresponding one of y_m), thus the right continuity of γ_1 implies that $\gamma_1(z) = \lim_{m\to+\infty} \gamma_1(y_m) = 1$, that is, $z \in A$, which proves that A is closed.

Define now $f_k(x) = \max(1 - kd(x, A), 0)$, where $d(x, A)$ represents the Euclidean distance of x to the set A. Each function f_k is continuous and obviously verifies that $f_k(x) \in [0, 1]$ for every $x \in \mathbb{R}$. As we have proved that A is closed, it is clear that $f_k \searrow \mathbb{I}_A$.

Next, we prove that each f_k is coordinatewise nondecreasing or, equivalently, that $d(x, A)$ is a nonincreasing function of $x \in \mathbb{R}^n$. Choose $\varepsilon > 0$ and $z_\varepsilon \in A$ such that $\|x - z_\varepsilon\|_2 < d(x, A) + \varepsilon$. Given y lexicographically larger or equal than x, define $t_\varepsilon = z_\varepsilon + (y - x)$. Then $z_\varepsilon \leq t_\varepsilon$, $t_\varepsilon \in A$, and $\|x - z_\varepsilon\|_2 = \|x - z_\varepsilon\|_2$. As $\varepsilon > 0$ is arbitrarily chosen, it follows that $d(y, A) \leq d(x, A)$, thus f_k is nondecreasing.

Finally, defining g_k analogously with respect to γ_2, it follows now by dominated convergence that

$$\begin{aligned}&\mathrm{Cov}\big(\gamma_1(X_1,\ldots,X_n),\gamma_2(X_1,\ldots,X_n)\big)\\&\quad=\lim_{k\to+\infty}\mathrm{Cov}\big(f_k(X_1,\ldots,X_n),g_k(X_1,\ldots,X_n)\big),\end{aligned}$$

thus $\mathrm{Cov}(\gamma_1(X_1,\ldots,X_n),\gamma_2(X_1,\ldots,X_n))\geq 0$. □

Theorem 1.32 *The random variables $X_1,\ldots,X_n$ are associated if and only if $\mathrm{Cov}(f(X_1,\ldots,X_n),g(X_1,\ldots,X_n))\geq 0$ for all coordinatewise nondecreasing, continuous and bounded functions $f,g:\mathbb{R}^n\longrightarrow\mathbb{R}$.*

Proof Let $\gamma_1,\gamma_2:\mathbb{R}^n\longrightarrow\{0,1\}$ be coordinatewise nondecreasing, $A_1=\gamma_1^{-1}(\{1\})$ and $A_2=\gamma_2^{-1}(\{1\})$. Given $\varepsilon>0$, let $C\subset A_1$ be a compact such that

$$\mathbf{P}\big((X_1,\ldots,X_n)\in A_1\big)\leq\mathbf{P}\big((X_1,\ldots,X_n)\in C\big)+\varepsilon \tag{1.5}$$

and

$$C_1=C+[0,+\infty)^n=\big\{c+t,c\in C,t=(t_1,\ldots,t_n),t_i\geq 0,i=1,\ldots,n\big\}\subset A_1,$$

as γ_1 is coordinatewise nondecreasing. Let $x_k=c_k+t_k$, $k\in\mathbb{N}$, be a sequence in C_1, convergent to some z. As C is compact, there exists a subsequence c_{k_ℓ}, $\ell\in\mathbb{N}$, convergent to some $s\in C$. Then, the corresponding subsequence t_{k_ℓ} is convergent to $z-s$, so all the coordinates of $z-s$ are nonnegative, and $z=s+(z-s)\in C_1$, that is, C_1 is closed. It is obvious, from the construction of the set C_1, that $\mathbb{I}_{C_1}$ is coordinatewise nondecreasing, right continuous and $\mathbb{I}_{C_1}\leq\gamma_1$. We can repeat the construction to obtain a closed set $C_2\subset A_2$ such that $\mathbb{I}_{C_2}$ is coordinatewise nondecreasing, right continuous and $\mathbb{I}_{C_2}\leq\gamma_2$. Now, taking into account Lemma 1.31, we have that $\mathrm{Cov}(\mathbb{I}_{C_1}(X_1,\ldots,X_n),\mathbb{I}_{C_2}(X_1,\ldots,X_n))\geq 0$. On the other hand, given the construction made,

$$\mathrm{E}\big(\gamma_1(X_1,\ldots,X_n)\gamma_2(X_1,\ldots,X_n)\big)\geq\mathrm{E}\big(\mathbb{I}_{C_1}(X_1,\ldots,X_n)\mathbb{I}_{C_2}(X_1,\ldots,X_n)\big).$$

From (1.5) it follows that $\mathrm{E}\gamma_1(X_1,\ldots,X_n)\leq\mathrm{E}\mathbb{I}_{C_1}(X_1,\ldots,X_n)+\varepsilon$ and analogously $\mathrm{E}\gamma_2(X_1,\ldots,X_n)\leq\mathrm{E}\mathbb{I}_{C_2}(X_1,\ldots,X_n)+\varepsilon$. So, finally we have

$$\begin{aligned}&\mathrm{Cov}\big(\gamma_1(X_1,\ldots,X_n),\gamma_2(X_1,\ldots,X_n)\big)\\&\quad\geq\mathrm{E}\big(\mathbb{I}_{C_1}(X_1,\ldots,X_n)\mathbb{I}_{C_2}(X_1,\ldots,X_n)\big)\\&\qquad-\big(\mathrm{E}\mathbb{I}_{C_1}(X_1,\ldots,X_n)+\varepsilon\big)\big(\mathrm{E}\mathbb{I}_{C_1}(X_1,\ldots,X_n)+\varepsilon\big)\\&\quad\geq\mathrm{Cov}\big(\mathbb{I}_{C_1}(X_1,\ldots,X_n),\mathbb{I}_{C_2}(X_1,\ldots,X_n)\big)-2\varepsilon-\varepsilon^2.\end{aligned}$$

As $\varepsilon>0$ was arbitrarily chosen, if follows that $\mathrm{Cov}(\gamma_1(X_1,\ldots,X_n),\gamma_2(X_1,\ldots,X_n))\geq 0$, so, taking into account Theorem 1.22, we have that the variables $X_1,\ldots,X_n$ are associated. □

The characterization of association depending only on continuous functions allows us to obtain the preservation of association through convergence in distribution. Let us introduce some more notation: we will denote by $\stackrel{d}{\longrightarrow}$ convergence in distribution.

Theorem 1.33 *For each $k \in \mathbb{N}$, let $X_{1,k}, \ldots, X_{n,k}$ be associated random variables and assume that, as $k \longrightarrow +\infty$, $(X_{1,k}, \ldots, X_{n,k}) \xrightarrow{d} (X_1, \ldots, X_n)$. Then the random variables $X_1, \ldots, X_n$ are associated.*

Proof Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be coordinatewise nondecreasing, continuous and bounded functions. Then, for each $k \in \mathbb{N}$, $\mathrm{Cov}(f(X_{1,k}, \ldots, X_{n,k}), g(X_{1,k}, \ldots, X_{n,k})) \geq 0$. As the f, g and fg are all bounded and continuous functions, it follows that

$$\begin{aligned}
&\mathrm{Cov}\big(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)\big) \\
&\quad = \lim_{k\to+\infty} \mathrm{E}\big(f(X_{1,k}, \ldots, X_{n,k}) g(X_{1,k}, \ldots, X_{n,k})\big) \\
&\qquad - \mathrm{E}f(X_{1,k}, \ldots, X_{n,k})\mathrm{E}g(X_{1,k}, \ldots, X_{n,k}) \\
&\quad = \lim_{k\to+\infty} \mathrm{Cov}\big(f(X_{1,k}, \ldots, X_{n,k}), g(X_{1,k}, \ldots, X_{n,k})\big) \geq 0.
\end{aligned}$$

Thus, taking into account Theorem 1.32, the variables $X_1, \ldots, X_n$ are associated. □

We can still reduce somewhat the family of test functions for the definition of association.

Theorem 1.34 *The random variables $X_1, \ldots, X_n$ are associated if and only if* $\mathrm{Cov}(f(X_1, \ldots, X_n), g(X_1, \ldots, X_n)) \geq 0$, *for all coordinatewise nondecreasing, bounded functions $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ with bounded first partial derivatives.*

Proof Taking into account Theorem 1.32, it is enough to prove that, under the assumptions given, the inequality $\mathrm{Cov}(h_1(X_1, \ldots, X_n), h_2(X_1, \ldots, X_n)) \geq 0$ holds for all coordinatewise nondecreasing, continuous and bounded functions h_1 and h_2. This will follow if we prove that each such function h_i is the limit of a sequence of coordinatewise nondecreasing, bounded with first partial derivatives functions, by dominated convergence. Denote, for all $k \in \mathbb{N}$ and $x \in \mathbb{R}^n$, $\chi_k(x) = (\frac{k}{2\pi})^{n/2} \exp(-\frac{k\|x\|_2}{2})$ and define

$$h_{i,k}(x) = \int_{\mathbb{R}^n} h_i(x-y)\chi_k(y)\, dy_1 \cdots dy_n.$$

Then, each $h_i = \lim_{k\to+\infty} h_{i,k}$, so the result follows. □

It is now possible to give a complete characterization of association for Gaussian families of random variables. This is a nice result due to Pitt [82] giving a full description using only the covariances of the random variables.

Theorem 1.35 *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a Gaussian random vector. Then, the variables $X_1, \ldots, X_n$ are associated if and only if* $\mathrm{Cov}(X_i, X_j) \geq 0$, *$i, j = 1, \ldots, n$, $i \neq j$.*

Proof If the variables are associated, the covariances are obviously nonnegative, as follows immediately from the definition, so we need only to prove the other implication. Of course, without loss of generality, we may assume that $\mathbf{X}$ is centred. Assume for the moment that the covariance matrix $\Sigma = [\sigma_{ij}]$ of $(X_1, \ldots, X_n)$ is invertible, so that there exists a density $p_{\mathbf{X}}$. Consider $\mathbf{Z}$ a random vector with the same distribution as $\mathbf{X}$ but independent from $\mathbf{X}$. For each $\alpha \in [0, 1]$, define $\mathbf{Y}(\alpha) = \alpha\mathbf{X} + (1 - \alpha^2)^{1/2}\mathbf{Z}$, so $\mathbf{Y}(\alpha)$ is Gaussian centred with covariance matrix Σ, and $\operatorname{Cov}(X_j, Y_j(\alpha)) = \alpha\sigma_{ij}$, $i, j = 1, \ldots, n$. Further, define, for each $\alpha \in [0, 1]$, $F(\alpha) = \mathrm{E}(f(\mathbf{X})g(\mathbf{Y}(\alpha)))$, where f and g are coordinatewise nondecreasing, bounded with bounded first partial derivatives. Thus, F is continuous, and $F(1) - F(0) = \mathrm{E}(f(\mathbf{X})g(\mathbf{X})) - \mathrm{E}(f(\mathbf{X})g(\mathbf{Z})) = \operatorname{Cov}(f(\mathbf{X}), g(\mathbf{X}))$, due to the independence of $\mathbf{Z}$ and $\mathbf{X}$. Taking into account Theorem 1.34, it is enough to prove that F is differentiable and has nonnegative derivative in $(0, 1)$.

In order to find a more suitable representation for F, consider the conditional density

$$\begin{aligned}
p(\alpha, x, y) &= p_{\mathbf{Y}(\alpha)|\mathbf{X}}(y|x) \\
&= \frac{p_{\mathbf{Y}(\alpha),\mathbf{X}}(y, x)}{p_{\mathbf{X}}(x)} = \frac{\partial^n}{\partial y_1 \cdots \partial y_n}\mathbf{P}\big(Y_1(\alpha) \le y_1, \ldots, Y_n(\alpha) \le y_n | \mathbf{X} = x\big) \\
&= \big(1 - \alpha^2\big)^{-n/2} p_{\mathbf{X}}\big(\big(1 - \alpha^2\big)^{-1/2}(\alpha x - y)\big).
\end{aligned}$$

Then

$$\begin{aligned}
F(\alpha) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x)g(y)p_{\mathbf{Y}(\alpha),\mathbf{X}}(y, x)\, dy\, dx \\
&= \int_{\mathbb{R}^n} f(x)p_{\mathbf{X}}(x) \int_{\mathbb{R}^n} g(y)p(\alpha, x, y)\, dy\, dx \\
&= \int_{\mathbb{R}^n} p_{\mathbf{X}}(x) f(x)h(\alpha, x)\, dx,
\end{aligned}$$

where $h(\alpha, x) = \int_{\mathbb{R}^n} g(y)p(\alpha, x, y)\, dy$. This function may be rewritten as a convolution. Indeed, denoting $p_\alpha(x) = (1 - \alpha^2)^{-n/2} p_{\mathbf{X}}((1 - \alpha^2)^{-1/2}x)$, we have

$$h(\alpha, x) = \int_{\mathbb{R}^n} g(y)p_\alpha(\alpha x - y)\, dy = \int_{\mathbb{R}^n} g(\alpha x - y)p_\alpha(y)\, dy.$$

As g is coordinatewise nondecreasing and $\alpha \in (0, 1)$, it follows that h has bounded, continuous and nonnegative partial derivatives $\frac{\partial h}{\partial x_i}$, $i = 1, \ldots, n$. Moreover, as the conditional density p decreases exponentially at infinity, we can differentiate h with respect to α inside the integral, that is,

$$\frac{\partial h}{\partial \alpha}(\alpha, x) = \int_{\mathbb{R}^n} g(y)\frac{\partial p}{\partial \alpha}(\alpha, x, y)\, dy.$$

The representation (C.5) for the derivative of p of course implies a similar expression for the corresponding derivative of h. Thus, differentiating again under the integral sign, we find that

$$
\begin{aligned}
F'(\alpha) &= \int_{\mathbb{R}^n} p_{\mathbf{X}}(x) f(x) \frac{\partial h}{\partial \alpha}(\alpha, x)\, dx \\
&= -\frac{1}{\alpha} \int_{\mathbb{R}^n} f(x) p_{\mathbf{X}}(x) \left(\sum_{i,j=1}^{n} \sigma_{ij} \frac{\partial^2 h}{\partial x_i\, \partial x_j}(\alpha, x) - \sum_{i=1}^{n} x_j \frac{\partial h}{\partial x_i}(\alpha, x) \right) dx.
\end{aligned}
$$

Integrating by parts on the first sum, we find, taking into account that f is bounded and h has bounded derivatives,

$$
\begin{aligned}
\alpha F'(\alpha) = & -\int_{\mathbb{R}^n} \sum_{i,j=1}^{n} \sigma_{ij} \frac{\partial}{\partial x_i} \big(f(x) p_{\mathbf{X}}(x) \big)\, dx \\
& + \int_{\mathbb{R}^n} f(x) p_{\mathbf{X}}(x) \sum_{i=1}^{n} x_j \frac{\partial h}{\partial x_j}(\alpha, x)\, dx \\
= & -\int_{\mathbb{R}^n} \sum_{i,j=1}^{n} \sigma_{ij} \left(\frac{\partial p_{\mathbf{X}}}{\partial x_i}(x) f(x) + \frac{\partial f}{\partial x_i}(x) p_{\mathbf{X}}(x) \right) \frac{\partial h}{\partial x_j}(\alpha, x)\, dx \\
& + \int_{\mathbb{R}^n} f(x) p_{\mathbf{X}}(x) \sum_{i=1}^{n} x_j \frac{\partial h}{\partial x_i}(\alpha, x)\, dx. \qquad (1.6)
\end{aligned}
$$

If $\Sigma^{-1} = [s_{ij}]$, we can write, taking into account the symmetry of Σ (thus Σ^{-1} is also symmetric),

$$
\begin{aligned}
&\sum_{i,j=1}^{n} \sigma_{ij} \frac{\partial p_{\mathbf{X}}}{\partial x_i}(x) \frac{\partial h}{\partial x_j}(\alpha, x) \\
&\quad = \frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij} p_{\mathbf{X}}(x) \frac{\partial}{\partial x_i} \left(\sum_{k,\ell=1}^{n} s_{k\ell} x_k x_\ell \right) \frac{\partial h}{\partial x_j}(\alpha, x) \\
&\quad = \sum_{i,j=1}^{n} \sigma_{ij} p_{\mathbf{X}}(x) \left(\sum_{k=1}^{n} s_{ki} x_k \right) \frac{\partial h}{\partial x_j}(\alpha, x) \\
&\quad = \sum_{j=1}^{n} \frac{\partial h}{\partial x_j}(\alpha, x) \sum_{k=1}^{n} x_k \sum_{i=1}^{n} s_{ki} \sigma_{ij} = \sum_{j=1}^{n} \frac{\partial h}{\partial x_j}(\alpha, x) x_j.
\end{aligned}
$$

Thus, in the final expression on the right of (1.6), the integration on the first summation cancels with the second integral, leaving us with

$$
F'(\alpha) = \frac{1}{\alpha} \int_{\mathbb{R}^n} p_{\mathbf{X}}(x) \sum_{i,j=1}^{n} \sigma_{ij} \frac{\partial f}{\partial x_i}(x) \frac{\partial h}{\partial x_j}(\alpha, x)\, dx.
$$

As f is assumed nondecreasing and we have proved that h has nonnegative partial derivatives, it follows from the assumption $\sigma_{ij} \geq 0$ that $F'(\alpha) \geq 0$. Consequently, $F(1) - F(0) = \mathrm{Cov}(f(\mathbf{X}), g(\mathbf{X})) \geq 0$.

It remains to prove the result when Σ is not invertible. In this case, take $\mathbf{Z}$ to be a random Gaussian random vector centred with covariance matrix I_n, the identity matrix, and independent of $\mathbf{X}$. Then the random vector $\mathbf{X}_k = \mathbf{X} + \frac{1}{k}\mathbf{Z}$ has covariance matrix with all entries nonnegative, that is, the coordinates of $\mathbf{X}_k$ are associated. Now, as $k \longrightarrow +\infty$, $\mathbf{X}_k \stackrel{d}{\longrightarrow} \mathbf{X}$; thus, taking into account Theorem 1.33, we get that the coordinate variables $X_1, \ldots, X_n$ of $\mathbf{X}$ are associated. $\square$

1.4 Abstract Spaces

The notion of association only depends on using nondecreasing functions; thus, it is natural trying to extend this dependence structure to random variables taking values in more general spaces, as long as these have an order relation defined. The main contribution to this extension comes from Lindqvist [60].

Let us start by setting some notation and basic definitions. We will restrain ourselves to the more important facts concerning order relations without going into details and proofs on this direction, unless this is really relevant for the scope of this text. For a general treatment of topological issues with order relations, we refer the reader to Nachbin [67]. For the sequel of this section, recall that a polish space is a separable, complete and metrizable topological space.

Definition 1.36 We call S a *partially ordered polish* space, or *POP* space, a complete and separable metric where there is defined a partial order relation $\leq$ such that the set $G = \{(x, y) \in S \times S : x \leq y\}$ is closed in the product space $S \times S$.

There is an obvious way to build product structures based on POP spaces: given POP spaces $S_1, S_2, \ldots$, the product $S_1 \times S_2 \times \cdots$ is a POP space with respect to the partial ordering

$$(x_1, x_2, \ldots) \leq (y_1, y_2, \ldots) \quad \Leftrightarrow \quad x_i \leq y_i, \quad i = 1, 2, \ldots.$$

Notice that, to simplify the notation, we have used the same symbol $\leq$ for the order relation in each POP space. When $S = \mathbb{R}$, this is what is done to define the usual partial order relation in the Euclidean space $\mathbb{R}^n$. We will refer to this as the *usual order* in $\mathbb{R}^n$.

Definition 1.37 Let S be a POP space, and $A \subset S$. The set A is said to be *totally ordered* if for all $x, y \in A$, we have either $x \leq y$ or $y \leq x$.

Now we must define what is understood by nondecreasing sets and functions.

Definition 1.38 Let S be a POP space. A set $A \subset S$ is *nondecreasing* if whenever $x \in A$ and $x \leq y$, it holds that $y \in A$. A set $A \subset S$ is *nonincreasing* if whenever $x \in A$ and $y \leq x$, it holds that $y \in A$.

Of course, not all sets are nondecreasing. Simple examples of sets in $S = \mathbb{R}^2$ that are not nondecreasing with respect to the usual order are $\{(0, 1), (1, 0)\}$ or $\{0\} \times \mathbb{R}$. We will need to identify conveniently induced nondecreasing sets.

Definition 1.39 Let $A \subset S$, a POP space. Define $\mathrm{Inc}(A)$ as the smallest nondecreasing set that includes A and $\mathrm{Dec}(A)$ as the largest nonincreasing set included in A.

An explicit description is $\mathrm{Inc}(A) = \{y \in S : \text{there exists } x \in A \text{ such that } x \leq y\}$. For example, in $\mathbb{R}^2$ and with respect to the usual order, $\mathrm{Inc}(\{(0,1),(1,0)\}) = [0,+\infty) \times [1,+\infty) \cup [1,+\infty) \times [0,+\infty)$.

We can now extend association to S-valued random elements.

Definition 1.40 Let S be a POP space, and X an S-valued random variable. We say that X is *associated* if, for all nondecreasing sets A_1 and A_2,

$$\mathbf{P}(X \in A_1 \cap A_2) \geq \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2).$$

Remark 1.41 Notice that we are adopting a different language with this definition. In fact, looking to Definition 1.12 of association of a family of random variables $X_1, \ldots, X_n$, we are now adopting the expression $\mathbf{X} = (X_1, \ldots, X_n)$ is associated. Thus, the definition above, when $S = \mathbb{R}^n$, means that the n coordinate variables of the random vector are associated.

Remark 1.42 A set A is nondecreasing if and only if its complement A^c is nonincreasing. Thus, reasoning as is (1.1), one can, in the definition above, replace nondecreasing sets by nonincreasing sets.

Remark 1.43 For the case $S = R$, Definition 1.40 is equivalent to X being PQD with itself, which always holds as follows from Theorem 1.24 and Remark 1.13.

Example 1.44 There do exist random vectors that are not associated in the sense just introduced. Take, for instance, $S = \mathbb{R}^2$ with its usual order relation, and consider $\mathbf{X}$ with distribution $\mathbf{P}(\mathbf{X} = (0,1)) = \mathbf{P}(\mathbf{X} = (1,0)) = \frac{1}{2}$. If we take $A_1 = \mathrm{Inc}\{(0,1)\} = \{(x_1,x_2) : x_1 \geq 0, x_2 \geq 1\}$ and $A_2 = \mathrm{Inc}\{(1,0)\} = \{(x_1,x_2) : x_1 \geq 1, x_2 \geq 0\}$, then $\mathbf{P}(\mathbf{X} \in A_1) \times \mathbf{P}(\mathbf{X} \in A_2) = \frac{1}{4}$. On the other hand, $A_1 \cap A_2 = [1,+\infty)^2$, and thus $\mathbf{P}(\mathbf{X} \in A_1 \cap A_2) = 0$. That is, this random vector is not associated. Notice that the relevant feature explored in this case is the fact that the distribution of $\mathbf{X}$ is concentrated in two points that are not comparable using the order relation defined in $\mathbb{R}^2$.

We will verify soon that the definition just introduced really coincides with Definition 1.12 when treating real random variables. For this, we need to prove alternative characterizations of association in POP spaces.

Definition 1.45 Let S be a POP space. A nondecreasing set $A \subset S$ is *compact generated* if $A = \mathrm{Inc}(K)$ for some compact set K. A nonincreasing set $A \subset S$ is *compact generated* if $A = \mathrm{Dec}(K)$ for some compact set K.

It is proved in Nachbin [67] (see page 44) that nondecreasing or nonincreasing compact generated sets are closed.

We still need to clarify what is a monotone function between two POP spaces, although this notion should be by now clear.

Definition 1.46 Let S_1 and S_2 be two POP spaces. A function $f : S_1 \longrightarrow S_2$ is *nondecreasing* if $x \leq y$ (in S_1) implies $f(x) \leq f(y)$ (in S_2). A function $f : S_1 \longrightarrow S_2$ is *nonincreasing* if $x \leq y$ (in S_1) implies $f(x) \geq f(y)$ (in S_2).

We can now prove some equivalent characterizations of association.

Theorem 1.47 *Let S be a POP space, and X an S-valued random variable. The following statements are equivalent*:

(a) *X is associated*;
(b) *for all nondecreasing functions $f, g : S \longrightarrow \mathbb{R}$, $\mathrm{E}(f(X)g(X)) \geq \mathrm{E}f(X)\mathrm{E}g(X)$*;
(c) *for all nondecreasing closed sets A_1 and A_2,*

$$\mathbf{P}(X \in A_1 \cap A_2) \geq \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2);$$

(d) *for all nondecreasing compact generated sets A_1 and A_2,*

$$\mathbf{P}(X \in A_1 \cap A_2) \geq \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2).$$

Proof (a) $\Rightarrow$ (b): Use Hoeffding's formula (1.2) to write

$$\begin{aligned}&\mathrm{E}\big(f(X)g(X)\big) - \mathrm{E}f(X)\mathrm{E}g(X)\\ &\quad= \int_{\mathbb{R}^2} \mathbf{P}\big(f(X) > u, g(X) > v\big) - \mathbf{P}\big(f(X) > u\big)\mathbf{P}\big(g(X) > v\big)\, du\, dv.\end{aligned}$$

Now, as f and g are nondecreasing, the sets $A_1(u) = \{x \in S : f(x) > u\}$ and $A_2(v) = \{x \in S : g(x) > v\}$ are, for every $u, v \in \mathbb{R}$, nondecreasing, so it follows that

$$\begin{aligned}&\mathrm{E}\big(f(X)g(X)\big) - \mathrm{E}f(X)\mathrm{E}g(X)\\ &\quad= \int_{\mathbb{R}^2} \mathbf{P}\big(X \in A_1(u) \cap A_2(v)\big) - \mathbf{P}\big(X \in A_1(u)\big)\mathbf{P}\big(X \in A_2(v)\big)\, du\, dv \geq 0.\end{aligned}$$

(b) $\Rightarrow$ (a): Given nondecreasing sets A_1 and A_2, take $f = \mathbb{I}_{A_1}$ and $g = \mathbb{I}_{A_2}$. These functions are obviously nondecreasing, so the implication follows.

(a) $\Rightarrow$ (c) and (c) $\Rightarrow$ (d): These are obvious.

(d) $\Rightarrow$ (a): Let A_1 and A_2 be nondecreasing sets and fix $\varepsilon > 0$. It is possible to choose compact sets $K_1 \subset A_1$ and $K_2 \subset A_2$ such that $\mathbf{P}(A_1 \setminus K_1) < \varepsilon$ and $\mathbf{P}(A_2 \setminus K_2) < \varepsilon$. Consider now $H_1 = \mathrm{Inc}(K_1)$ and $H_2 = \mathrm{Inc}(K_2)$. Then, obviously $K_1 \subset H_1 \subset A_1$, $\mathbf{P}(A_1 \setminus H_1) < \varepsilon$, $K_2 \subset H_2 \subset A_2$ and $\mathbf{P}(A_2 \setminus H_2) < \varepsilon$. Finally,

$$\begin{aligned}&\mathbf{P}(X \in A_1 \cap A_2) - \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2)\\ &\quad\geq \mathbf{P}(X \in H_1 \cap H_2) - \big(\mathbf{P}(X \in H_1) + \varepsilon\big)\big(\mathbf{P}(X \in H_2) + \varepsilon\big)\\ &\quad\geq \mathbf{P}(X \in H_1 \cap H_2) - \mathbf{P}(X \in H_1)\mathbf{P}(X \in H_2) - 2\varepsilon - \varepsilon^2.\end{aligned}$$

As $\varepsilon > 0$ is arbitrarily chosen, it follows that

$$\begin{aligned}&\mathbf{P}(X \in A_1 \cap A_2) - \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2)\\ &\quad\geq \mathbf{P}(X \in H_1 \cap H_2) - \mathbf{P}(X \in H_1)\mathbf{P}(X \in H_2) \geq 0.\end{aligned}$$

$\square$

The equivalence between (a) and (b) above shows that association in abstract ordered spaces, given by Definition 1.40, coincides with the Definition 1.12 by choosing $S=\mathbb{R}^n$ with its usual partial order.

For real random variables, we have proved in Theorem 1.32 that nondecreasing, bounded and continuous functions are enough to characterize the association of a family of random variables. It is possible to extend this characterization, but it depends on some extra ordering structure of the POP space.

Definition 1.48 A POP space S is *normally ordered* if given disjoint sets A_1, closed and nonincreasing, and A_2, closed and nondecreasing, there exists a nondecreasing and continuous function $f:S\longrightarrow[0,1]$ such that $f(x)=0$ if $x\in A_1$ and $f(x)=1$ if $x\in A_2$.

The above is not the definition of a normally ordered space adopted in Nachbin [67], but it is proved there that this is an equivalent characterization. The above definition is more convenient for our purposes. Notice that it is easily verified that, for every $n\in\mathbb{N}$, $\mathbb{R}^n$ is normally ordered (with respect to the usual partial ordering).

Theorem 1.49 *Let S be a normally ordered POP space. Then an S-valued random variable is associated if and only if for all nondecreasing, bounded and continuous functions $f,g:S\longrightarrow\mathbb{R}$, $\mathrm{E}(f(X)g(X))\geq \mathrm{E}f(X)\mathrm{E}g(X)$.*

Proof It is clear from Theorem 1.47(b) that if X is associated, the inequality stated holds for every nondecreasing, bounded and continuous functions $f,g:S\longrightarrow\mathbb{R}$. To prove this result, it is now enough to check that if $\mathrm{E}(f(X)g(X))\geq \mathrm{E}f(X)\mathrm{E}g(X)$ is verified for all nondecreasing, bounded and continuous functions $f,g:S\longrightarrow\mathbb{R}$, then we have Theorem 1.47(d). Let A_1 and A_2 be nondecreasing compact generated sets and choose $\varepsilon>0$. As S is complete and separable, there exist compact sets $K_1\subset A_1^c$ and $K_2\subset A_2^c$ such that $\mathbf{P}(A_1^c\setminus K_1)<\varepsilon$ and $\mathbf{P}(A_2^c\setminus K_2)<\varepsilon$. Put $G_1=\mathrm{Dec}(K_1)$ and $G_2=\mathrm{Dec}(K_2)$. Then, it is obvious that $G_1\subset A_1^c$, $\mathbf{P}(A_1^c\setminus G_1)<\varepsilon$, $G_2\subset A_2^c$ and $\mathbf{P}(A_2^c\setminus G_2)<\varepsilon$. Now, since S is normally ordered, there exist nondecreasing and continuous functions $f_1,f_2:S\longrightarrow[0,1]$ such that $f_1(x)=0$ if $x\in G_1$, $f_1(x)=1$ if $x\in A_1$, $f_2(x)=0$ if $x\in G_2$, $f_2(x)=1$ if $x\in A_2$. Moreover, it is clear that

$$\begin{aligned}
\mathbf{P}(X\in A_1)&\leq\int_S f_1(x)\mathbf{P}_X(dx)\leq\mathbf{P}\big(G_1^c\big)<\mathbf{P}(A_1)+\varepsilon,\\
\mathbf{P}(X\in A_2)&\leq\int_S f_2(x)\mathbf{P}_X(dx)\leq\mathbf{P}\big(G_2^c\big)<\mathbf{P}(A_2)+\varepsilon,\\
\mathbf{P}(X\in A_1\cap A_2)&\leq\int_S f_1(x)f_2(x)\mathbf{P}_X(dx)\leq\mathbf{P}\big(G_1^c\cap G_2^c\big)<\mathbf{P}(A_1\cap A_2)+2\varepsilon.
\end{aligned}$$

So, finally, we obtain

$$\begin{aligned}
&\mathbf{P}(X\in A_1\cap A_2)-\mathbf{P}(X\in A_1)\mathbf{P}(X\in A_2)\\
&\quad\geq \mathrm{E}\big(f_1(X)f_2(X)\big)-2\varepsilon-\mathrm{E}f_1(X)\mathrm{E}f_2(X)\geq-2\varepsilon,
\end{aligned}$$

so, as $\varepsilon > 0$ is arbitrary, the result follows. □

The basic properties that we have proved before for associated families of random variables may be extended to this enlarged framework. We start with the extended version of Theorem 1.15.

Theorem 1.50 *Let S_1 and S_2 be POP spaces, $f : S_1 \longrightarrow S_2$ a nondecreasing measurable function, and X an S_1-valued associated random variable. Then $Y = f(X)$ is an S_2-valued associated random variable.*

Proof Let A_1 and A_2 be nondecreasing subsets of S_2. Then, obviously $f^{-1}(A_1)$ and $f^{-1}(A_2)$ are nondecreasing subsets of S_1. Thus,

$$\begin{aligned}&\mathbf{P}(Y \in A_1 \cap A_2)\\ &\quad= \mathbf{P}\big(X \in f^{-1}(A_1) \cap f^{-1}(A_2)\big)\\ &\quad\geq \mathbf{P}\big(X \in f^{-1}(A_1)\big)\mathbf{P}\big(X \in f^{-1}(A_2)\big) = \mathbf{P}(Y \in A_1)\mathbf{P}(Y \in A_2).\end{aligned}$$ □

The following is a generalized version of Theorem 1.18.

Theorem 1.51 *Let S_1 and S_2 be POP spaces. Let X_1 be an S_1-valued associated random variable, and X_2 be an S_2-valued associated random variable. If X_1 and X_2 are independent, then (X_1, X_2) is an $S_1 \times S_2$-valued associated random variable.*

Proof Let $f, g : S_1 \times S_2 \longrightarrow \mathbb{R}$ be nondecreasing functions. It is obvious that, for each $x_1 \in S_1$, $f(x_1, \cdot), g(x_1, \cdot) : S_2 \longrightarrow \mathbb{R}$ are nondecreasing functions. Thus,

$$\mathrm{E}\big(f(x_1, X_2)g(x_1, X_2)\big) \geq \mathrm{E}\big(f(x_1, X_2)\big) \geq \mathrm{E}\big(g(x_1, X_2)\big).$$

Then, taking into account the independence between X_1 and X_2, we have

$$\begin{aligned}&\mathrm{E}\big(f(X_1, X_2)g(X_1, X_2)\big)\\ &\quad= \mathrm{E}\big(\mathrm{E}\big(f(X_1, X_2)g(X_1, X_2)\big)|X_1\big)\\ &\quad\geq \mathrm{E}\big(\mathrm{E}\big(f(X_1, X_2)\big)\mathrm{E}\big(g(X_1, X_2)\big)|X_1\big) = \mathrm{E}\big(f(X_1, X_2)\big)\mathrm{E}\big(g(X_1, X_2)\big),\end{aligned}$$

so the association of (X_1, X_2) follows from Theorem 1.47(b). □

There is an interesting and relevant relation between association and the properties of the order structure of the POP space. It allows us to find, as a corollary, a generalized version of Theorem 1.24.

Theorem 1.52 *Let S be a POP space, and X be an S-valued random variable. If there exists a totally ordered set $A \subset S$ such that $\mathbf{P}(X \in A) = 1$, then X is associated.*

Proof Let A_1 and A_2 be nondecreasing sets in S. Assume that it is possible to choose $x \in A \cap A_1 \cap A_2^c$ and $y \in A \cap A_1^c \cap A_2$. Then, as A is totally ordered, we have either $x \le y$ or $y \le x$. Assume, without loss of generality, that $x \le y$. We have then that $x \in A_1$, which is a nondecreasing set, so it follows that $y \in A_1$, but this is incompatible with $y \in A \cap A_1^c \cap A_2$, so this choice of x and y is not possible. Thus, we proved that either $A \cap A_1 \cap A_2^c = \varnothing$ or $A \cap A_1^c \cap A_2 = \varnothing$.

Now assume that $A \cap A_1 \cap A_2^c = \varnothing$. Then $\mathbf{P}(X \in A_1 \cap A_2^c) = \mathbf{P}(X \in A \cap A_1 \cap A_2^c) = 0$, so $\mathbf{P}(X \in A_1) = \mathbf{P}(X \in A_1 \cap A_2)$, and hence $\mathbf{P}(X \in A_1 \cap A_2) \ge \mathbf{P}(X \in A_1)\mathbf{P}(X \in A_2)$, that is, X is associated. The other case is proved symmetrically. □

Corollary 1.53 *Let S be a POP space. Then every S-valued random variable X is associated if and only if S is totally ordered.*

Proof If S is totally ordered, it follows immediately from Theorem 1.52 that every S-valued variable is associated. Assume now that S is not totally ordered. Thus there exist $x, y \in S$ such that they are not comparable, that is, neither $x \le y$ or $y \le x$ holds. Define the S-valued random variable X with distribution $\mathbf{P}(X = x) = \mathbf{P}(X = y) = \frac{1}{2}$. Then $\mathbf{P}(X \in \mathrm{Inc}(\{x\}) \cap \mathrm{Inc}(\{y\})) = 0$ and $\mathbf{P}(X \in \mathrm{Inc}(\{x\}))\mathbf{P}(X \in \mathrm{Inc}(\{y\})) = \frac{1}{4}$. Thus, X is not associated. □

Remark 1.54 Notice that the final part of the proof above reproduces the construction made in Remark 1.44 to give an example of a nonassociated $\mathbb{R}^2$-valued variable.

These results show that the order structure of the underlying space is crucial to characterize association. One consequence is that if we change the representation of a family of random variables, using a different base space, this could mean that we might loose the association property.

Example 1.55 Consider a real random variable X and define the random point mass δ_X. This is a random variable with values in $\mathcal{N}$, the space of measures on $\mathbb{R}$ such that, for every Borel set $A \subset \mathbb{R}$, $\mu(A) \in \{0, 1, 2, \ldots\}$. Each $\mu \in \mathcal{N}$ is representable in the form $\mu = \sum_n \delta_{x_n}$, where $x_n \in \mathbb{R}$, not necessarily distinct. Denote the support of μ by $\mathrm{supp}(\mu) = \{x_1, x_2, \ldots\}$. Then we may define $\mu_1 \le \mu_2$ if and only if $\mathrm{supp}(\mu_1) \subset \mathrm{supp}(\mu_2)$. This clearly defines a partial order relation in $\mathcal{N}$. It is also clear that this order is not total: it is not possible to order elements of $\mathcal{N}$ whose supports do not satisfy an inclusion relation. Thus δ_X is not associated with itself, although, as a real random variable, X is associated with itself.

To conclude this section, we prove the extended version of Theorem 1.33.

Theorem 1.56 *Let S be a normally ordered POP space, and X_n, $n \in \mathbb{N}$, associated S-valued random variables. If there exists an S-valued random variable X such that $X_n \xrightarrow{d} X$, then X is associated.*

Proof Taking into account Theorem 1.49, it is enough to consider nondecreasing, bounded and continuous functions $f, g : S \longrightarrow \mathbb{R}$ and verify that $\mathrm{E}(f(X)g(X)) \ge \mathrm{E}f(X)\mathrm{E}g(X)$. As each X_n is associated and the functions are nondecreasing, we have that $\mathrm{E}(f(X_n)g(X_n)) \geq \mathrm{E}f(X_n)\mathrm{E}g(X_n)$. The convergence in distribution implies that $\mathrm{E}(f(X_n)g(X_n)) \longrightarrow \mathrm{E}(f(X)g(X))$, $\mathrm{E}f(X_n) \longrightarrow \mathrm{E}f(X)$ and $\mathrm{E}g(X_n) \longrightarrow \mathrm{E}g(X)$, because every function is bounded and continuous, so the conclusion follows immediately. □

1.5 Some Other Weak Dependence Notions

In this section we return to the more elementary framework of treating families of real random variables. There are several variants of notions of positive dependence trying to weaken association while extending the positive quadrant dependence to larger families of variables, that is, liberating this dependence from treating just two random variables. We will introduce some of these notions and briefly establish some basic relations between them.

1.5.1 Some Other Positive Dependence Notions

The most natural extensions of positive quadrant dependence, described on what follows, were introduced by Joag-Dev [53]. Let us introduce some notation before defining the new concepts of dependence. Given random variables $X_1, X_2, \ldots$, a set $A \subset \mathbb{N}$ and $x_1, \ldots, x_{|A|} \in \mathbb{R}$, where $|A|$ represents the cardinality of A, denote $\mathbb{I}_A(x_1, \ldots, x_{|A|}) = \prod_{i\in A} \mathbb{I}_{(x_i,+\infty)}(X_i)$ and $\mathbb{J}_A(x_1, \ldots, x_{|A|}) = \prod_{i\in A} \mathbb{I}_{(-\infty,x_i]}(X_i) = \prod_{i\in A}(1 - \mathbb{I}_{(x_i,+\infty)}(X_i))$. We will write just $\mathbb{I}_A$ or $\mathbb{J}_A$ when confusion does not arise.

Definition 1.57 A family of random variables $X_1, \ldots, X_n$ is *strongly positive orthant dependent* (*SPOD*) if, given any disjoint $A, B \subset \{1, \ldots, n\}$ and real x_j's and y_k's,

$$\mathrm{Cov}\big(\mathbb{I}_A(x_1, \ldots, x_{|A|}), \mathbb{I}_B(y_1, \ldots, y_{|B|})\big) \geq 0,$$
$$\mathrm{Cov}\big(\mathbb{J}_A(x_1, \ldots, x_{|A|}), \mathbb{J}_B(y_1, \ldots, y_{|B|})\big) \geq 0,$$
$$\mathrm{Cov}\big(\mathbb{I}_A(x_1, \ldots, x_{|A|}), \mathbb{J}_B(y_1, \ldots, y_{|B|})\big) \leq 0.$$

A sequence of random variables X_n, $n \in \mathbb{N}$, is *strongly positive orthant dependent* (*SPOD*) if for any $n \in \mathbb{N}$, the random variables $X_1, \ldots, X_n$ are strongly positive orthant dependent.

Definition 1.58 A family of random variables $X_1, \ldots, X_n$ is *linearly positive quadrant dependent* (*LPQD*) if, given any disjoint $A, B \subset \{1, \ldots, n\}$ and positive λ_j's, the random variables $\sum_{i\in A} \lambda_i X_i$ and $\sum_{j\in B} \lambda_j X_j$ are pairwise quadrant dependent.

The following are obvious consequences of the definitions above.

Proposition 1.59 *Let $X_1, \ldots, X_n$ be associated random variables. Then $X_1, \ldots, X_n$ are SPOD.*

Table 1.3 SPOD but not LPQD random variables

x_1	x_2	x_3	$p(x_1,x_2,x_3)$	x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	$\frac{3}{14}$	1	0	0	$\frac{2}{14}$
0	1	0	$\frac{1}{14}$	1	0	1	$\frac{1}{14}$
0	2	0	$\frac{1}{14}$	1	1	0	$\frac{1}{14}$
0	2	1	$\frac{2}{14}$	1	2	1	$\frac{3}{14}$

Table 1.4 Pairwise verifications concerning SPOD ($\varepsilon \in (0,1)$)

$(j,k)=(1,2)$

x	y	$C_{1,2}(x-\varepsilon,y-\varepsilon)$	x	y	$C_{1,2}(x-\varepsilon,y-\varepsilon)$
0	0	0	1	0	0
0	1	0	1	1	0
0	2	1	2	0	0

$(j,k)=(1,3)$

x	y	$C_{1,3}(x-\varepsilon,y-\varepsilon)$	x	y	$C_{1,3}(x-\varepsilon,y-\varepsilon)$
0	0	0	1	0	0
0	1	0	1	1	$\frac{3}{28}$

$(j,k)=(2,3)$

x	y	$C_{2,3}(x-\varepsilon,y-\varepsilon)$	x	y	$C_{2,3}(x-\varepsilon,y-\varepsilon)$
0	0	0	1	1	$\frac{11}{98}$
0	1	0	2	0	0
1	0	0	2	1	$\frac{17}{98}$

Proposition 1.60 *Let $X_1,\dots,X_n$ be associated random variables. Then $X_1,\dots,X_n$ are LPQD.*

The examples below show that neither SPOD nor LPQD implies the other. We start by an example of SPOD random variables that are not LPQD.

Example 1.61 Consider discrete variables X_1, X_2 and X_3 with joint distribution characterized by $p(x_1,x_2,x_3)=\mathbf{P}(X_1=x_1,X_2=x_2,X_3=x_3)$, given in Table 1.3. To verify that these variables are indeed SPOD, we need to compute all the covariances mentioned in Definition 1.57. Obviously, it is enough to consider the case where $A,B\neq\varnothing$. If $A=\{j\}$ and $B=\{k\}$, then $\mathrm{Cov}(\mathbb{J}_A,\mathbb{J}_B)=\mathrm{Cov}(\mathbb{I}_A,\mathbb{I}_B)$ and $\mathrm{Cov}(\mathbb{I}_A,\mathbb{J}_B)=-\mathrm{Cov}(\mathbb{I}_A,\mathbb{I}_B)$, so it is enough to compute $C_{j,k}(x,y)=\mathrm{Cov}(\mathbb{I}_A,\mathbb{I}_B)=\mathrm{Cov}(\mathbb{I}_{(x,+\infty)}(X_j),\mathbb{I}_{(y,+\infty)}(X_k))$ for every possible values of $j,k=1,2,3$, $j\neq k$, x and y and check that these covariance are nonnegative. As the variables are discrete, we compute, in Table 1.4, these covariances for every point located just a little to the left and below the position of each of the possible values of

Table 1.5 Triple verifications concerning SPOD ($\varepsilon \in (0,1)$)

$(j,k,\ell) = (1,2,3), (2,1,3)$

x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$	x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$
0	0	0	0	1	0	0	0
0	0	1	0	1	0	1	$\frac{3}{28}$
0	1	0	0	1	1	0	0
0	1	1	$\frac{11}{98}$	1	1	1	$\frac{9}{98}$
0	2	0	0	1	2	0	0
0	2	1	$\frac{17}{98}$	1	2	1	$\frac{6}{49}$

$(j,k,\ell) = (1,3,2), (3,1,2)$

x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$	x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$
0	0	0	0	1	0	0	0
0	0	1	0	1	0	1	0
0	0	2	0	1	0	2	$\frac{3}{28}$
0	1	1	0	1	1	1	0
0	1	0	$\frac{11}{98}$	1	1	1	$\frac{5}{98}$
0	1	2	$\frac{17}{98}$	1	1	2	$\frac{15}{49}$

$(j,k,\ell) = (3,2,1), (3,2,1)$

x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$	x	y	z	$C_{(j,k),\ell}(x-\varepsilon, y-\varepsilon, z-\varepsilon)$
0	0	0	0	1	1	0	0
0	0	1	0	1	1	1	$\frac{1}{28}$
0	1	0	0	2	0	0	0
0	1	1	$\frac{3}{28}$	2	0	1	0
1	0	0	0	2	1	0	0
1	0	1	0	2	1	1	$\frac{1}{28}$

the pair (X_j, X_k). We also have to consider the case where $A = \{j,k\}$ and $B = \{\ell\}$, with $\ell \neq j,k$. In this case,

$$\begin{aligned} C^*_{j,k,\ell}(x,y,z) &= \operatorname{Cov}(\mathbb{J}_A, \mathbb{J}_B) \\ &= \operatorname{Cov}\big(\mathbb{I}_{(x,+\infty)}(X_j), \mathbb{I}_{(z,+\infty)}(X_\ell)\big) \\ &\quad + \operatorname{Cov}\big(\mathbb{I}_{(y,+\infty)}(X_k), \mathbb{I}_{(z,+\infty)}(X_\ell)\big) \\ &\quad + \operatorname{Cov}\big(\mathbb{I}_{(x,+\infty)}(X_j)\mathbb{I}_{(y,+\infty)}(X_k), \mathbb{I}_{(z,+\infty)}(X_\ell)\big). \end{aligned}$$

So, we need to compute the final covariance term

$$C_{(j,k),\ell}(x,y,z) = \operatorname{Cov}\big(\mathbb{I}_{(x,+\infty)}(X_j)\mathbb{I}_{(y,+\infty)}(X_k), \mathbb{I}_{(z,+\infty)}(X_\ell)\big).$$

which are given in Table 1.5. As all these covariances are nonnegative, it follows that all the $C^*_{j,k,\ell}(x,y,z) = \operatorname{Cov}(\mathbb{J}_A, \mathbb{J}_B)$ are nonnegative.

Table 1.6 LPQD but not SPOD random variables

x_1	x_2	x_3	$p(x_1,x_2,x_3)$	x_1	x_2	x_3	$p(x_1,x_2,x_3)$
1	1	1	$\frac{4}{17}$	2	3	1	$\frac{1}{17}$
1	1	2	$\frac{1}{17}$	3	1	2	$\frac{1}{17}$
1	2	2	$\frac{1}{17}$	3	2	1	$\frac{1}{17}$
1	3	2	$\frac{1}{17}$	3	3	1	$\frac{1}{17}$
2	1	2	$\frac{1}{17}$	3	3	2	$\frac{4}{17}$
2	2	1	$\frac{1}{17}$				

Finally, we still have to verify what happens with

$$\begin{aligned} D_{j,k,\ell}(x,y,z) = \mathrm{Cov}(\mathbb{I}_A, \mathbb{J}_B) \\ &= -\mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(X_j), \mathbb{I}_{(z,+\infty)}(X_\ell)\big) \\ &\quad - \mathrm{Cov}\big(\mathbb{I}_{(y,+\infty)}(X_k), \mathbb{I}_{(z,+\infty)}(X_\ell)\big) \\ &\quad + \mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(X_j)\mathbb{I}_{(y,+\infty)}(X_k), \mathbb{I}_{(z,+\infty)}(X_\ell)\big). \end{aligned}$$

As all values on the right have been computed, it is easy to verify that the only nonzero values are

$$\begin{aligned} D_{1,2,3}(1,1,1) &= D_{2,1,3}(1,1,1) = -\frac{13}{98}, \\ D_{1,2,3}(1,2,1) &= D_{2,1,3}(2,1,1) = -\frac{31}{196}, \\ D_{1,3,2}(1,1,1) &= D_{3,1,2}(1,1,1) = -\frac{3}{49}, \\ D_{1,3,2}(1,1,2) &= D_{3,1,2}(1,1,2) = -\frac{1}{49}, \\ D_{2,3,1}(1,1,1) &= D_{3,2,1}(1,1,1) = -\frac{1}{14}, \\ D_{2,3,1}(2,1,1) &= D_{3,2,1}(1,2,1) = -\frac{1}{14}, \end{aligned}$$

all conveniently nonpositive, so the random variables X_1, X_2 and X_3 are SPOD.

On the other hand,

$$\mathbf{P}(X_1 > 0, X_2 + X_3 > 1) - \mathbf{P}(X_1 > 0)\mathbf{P}(X_2 + X_3 > 1) = -\frac{1}{28},$$

so X_1, X_2 and X_3 are not LPQD.

Now we show an example of random variables that are LPQD but not SPOD.

Example 1.62 Consider again three random variables X_1, X_2 and X_3 with joint distribution $p(x_1,x_2,x_3) = \mathbf{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3)$ described in Table 1.6. It is easily verified that these random variables are not SPOD. In fact,

$$\mathbf{P}(X_1 > 1, X_2 > 1, X_3 > 1) = \frac{4}{17} \approx 0.2353$$
$$< \mathbf{P}(X_1 > 1, X_2 > 1)\mathbf{P}(X_3 > 1) = \frac{8 \times 9}{17^2} \approx 0.2491.$$

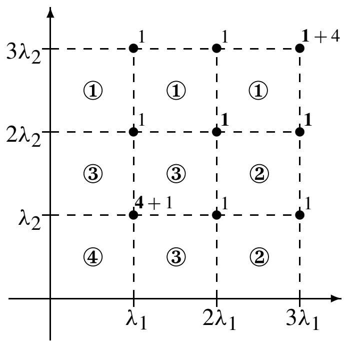
The verification that the random variables are not LPQD is somewhat lengthy, so we will just describe how it goes in a few cases, leaving the rest for the reader. First notice that verification that the variables are LPQD is equivalent to verifying that

$$\mathbf{P}(\lambda_1 X_j + \lambda_2 X_k > c_1 | X_\ell > c_2) \geq \mathbf{P}(\lambda_1 X_j + \lambda_2 X_k > c_1) \tag{1.7}$$

for all $\lambda_1, \lambda_2 > 0$, $c_1, c_2 \in \mathbb{R}$ and any permutation (j, k, ℓ) of $(1, 2, 3)$, when $\mathbf{P}(X_\ell > c_2) > 0$. The case where this later probability is null makes the right-hand side in the inequality characterizing LPQD also null, so the inequality is trivially verified. Let us consider $(j, k, \ell) = (1, 2, 3)$, the other cases being treated analogously. As X_3 only takes the values 1 and 2, it is enough to verify the case where $c_2 \in [1, 2)$. In such a case, $\mathbf{P}(X_3 > c_2) = \frac{9}{17}$. Now rewriting

$$\mathbf{P}(\lambda_1 X_j + \lambda_2 X_k > c_1 | X_\ell > c_2) = \frac{\mathbf{P}(\lambda_1 X_j + \lambda_2 X_k > c_1, X_\ell > c_2)}{\mathbf{P}(X_\ell > c_2)},$$

the events that contribute to the numerator will all be among the events contributing to the right-hand side of (1.7). But, for this right-hand side, there will be some more events, corresponding to the cases where $X_3 = 1$. The description of the distribution of (X_j, X_k) is now as indicated in the picture below, where the symbol $\bullet$ identifies the possible positions of (x_1, x_2), and the numbers next to each $\bullet$ are equal to $17\mathbf{P}(X_j = x_1, X_k = x_2)$. The numbers in **bold** identify how many cases contribute to the probability corresponding to $X_3 = 1$, so those that should be removed from the computation when calculating the probability conditional on $X_3 > c_2$ (recall that $c_2 \in [1, 2)$):



Assume now that the right-hand side of (1.7) is of the form $\frac{a}{17}$. Then, the value of the left-hand side is of the form $\frac{a-1}{9}$, $\frac{a-2}{9}$, $\frac{a-3}{9}$ or $\frac{a-7}{9}$, depending on the value of c_1, corresponding to the positions identified by the circled numbers. When c_1 defines a region corresponding to those marked by ①, the conditional probability is

$\frac{a-1}{9}$, as there is only one point satisfying $X_3 > c_2 \in [1, 2)$ above this region. Thus, we should verify that $\frac{a-1}{9} > \frac{a}{17}$ or, equivalently, $a > \frac{17}{8}$. The possible values for a for this choice are $a = 5, 6, 7$, all larger that $\frac{17}{8}$. For the region marked by ②, we should verify that $\frac{a-2}{9} > \frac{a}{17}$, as this region include two points satisfying $X_3 \leq 1.5$. This is equivalent to $a > \frac{34}{8} = 4.25$, and the possible values for a are 6 or 7. Repeating the arguments, on region ③ we should verify that $\frac{a-3}{9} > \frac{a}{17}$ or, equivalently, that $a > \frac{51}{8} = 6.375$, and the possible values for a are 8 or 10. Finally on region ④ we should verify $\frac{a-7}{9} > \frac{a}{17}$, which is equivalent to $a > \frac{119}{8} = 14.857$, and, in this region, $a = 17$. Thus, all the conditions for the LPQD-ness of the random variables have been verified for this permutation of indexes and conditioning. To complete the verification, we must do likewise for the remaining permutations and conditioning possibilities, but this is only a matter of routine repetition of the approach just described.

The following characterizations of independence, analogous to Corollary 1.6 and Theorem 1.17, are straightforward, reproducing the arguments for the proof of Corollary 1.6.

Theorem 1.63 *Let $X_1, \ldots, X_n$ be SPOD or LPQD random variables. These random variables are independent if and only if* $\mathrm{Cov}(X_i, X_j) = 0$ *for all $i \neq j$.*

1.5.2 Positive Dependencies and Stochastic Ordering

The notions of positive dependence discussed below, and a few others, have been used to study stochastic order relations between random variables. We give here a brief account of some results in this direction. The main purpose is to introduce some notions that imply the association but are easier to verify. A more complete picture of basic results can be found in Barlow and Proschan [5].

Definition 1.64 Let random variables X and Y have a joint density function or joint probability function, in case of a discrete distribution, f. The function f is *totally positive of order 2* (*TP2*) if, for all $x_1 < x_2$ and $y_1 < y_2$,

$$\begin{vmatrix} f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{vmatrix} \geq 0.$$

We say that X and Y are *TP2* if their joint density function or probability function is totally positive of order 2.

Theorem 1.65 *Let X and Y be TP2 random variables. Then* $\mathbf{P}(Y > y|X = x)$ *is, for every $y \in \mathbb{R}$, a nondecreasing function of x.*

Proof Assume that the joint distribution is absolutely continuous. It follows from Definition 1.64 that, for all $x_1 < x_2$,

$$\begin{vmatrix} \int_{-\infty}^{y} f(x_1,t)\,dt & \int_{y}^{+\infty} f(x_1,t)\,dt \\ \int_{-\infty}^{y} f(x_2,t)\,dt & \int_{y}^{+\infty} f(x_2,t)\,dt \end{vmatrix} \geq 0.$$

Replacing the first column by the sum of the two columns, we find that

$$\int_{-\infty}^{+\infty} f(x_1,t)\,dt \times \int_{y}^{+\infty} f(x_2,t)\,dt \geq \int_{-\infty}^{+\infty} f(x_2,t)\,dt \times \int_{y}^{+\infty} f(x_1,t)\,dt$$

or, equivalently

$$\mathbf{P}(Y > y|X = x_2) \geq \mathbf{P}(Y > y|X = x_1).$$

The discrete case is proved with an obvious modification of the argument above. □

It is now easily seen that the conclusion of this result implies the association of two random variables.

Theorem 1.66 *Let X and Y be such that $\mathbf{P}(Y > y|X = x)$ is, for every $y \in \mathbb{R}$, a nondecreasing function of x. Then X and Y are associated.*

Proof Define, for each $x \in \mathbb{R}$, $F_x(y) = \mathbf{P}(Y \leq y|X = x)$ and $g(u,x) = \inf\{y : u \leq F_x(y)\}$, the generalized inverse of $F_x(\cdot)$, for each fixed x. It follows from the nondecreaseness of $\mathbf{P}(Y > y|X = x)$ that g is nondecreasing in both arguments. Moreover, if U is uniform on $[0, 1]$, then $g(U, x)$ has distribution function F_x. Thus, if the random variable U is chosen independent from X, the distribution of (X, Y) coincides with the distribution of $(X, g(U, X))$. Taking into account Theorem 1.15, it follows that the random variables X and $g(U, X)$ are associated, and thus, as association only depends on the joint distributions, X and Y are associated. □

It is possible to prove a more general version of the preceding results, dealing with an arbitrary number of variables. But first, it is helpful to have a formalization of the nondecreasingness property that has been referred.

Definition 1.67 The random variables $X_1, \ldots, X_n$ are *stochastically nondecreasing* if, for each $j = 2, \ldots, n$, $\mathbf{P}(X_j > x_j|X_1 = x_1, \ldots, X_{j-1} = x_{j-1})$ is nondecreasing in $x_1, \ldots, x_{j-1}$.

The final goal is to prove the association of a family of random variables, using the TP2 condition. As when considering only two random variables, the notion of stochastic increasingness is an intermediate step. For this more general framework, we still need to check that the TP2 property is inherited by subfamilies.

Lemma 1.68 *Let $X_1, \dots, X_n$ have a joint density function or probability function (if the distribution is discrete) f that is TP2 in each pair of arguments. Then $X_1, \dots, X_{n-1}$ have a joint density function or probability function g that is TP2 in each pair of arguments.*

Proof Let $j, k \in \{1, \dots, n-1\}$, $j \neq k$, $x_i \in \mathbb{R}$, $i = 1, \dots, n$, but $i \neq j, k$, be fixed and denote, for simplicity, $f^*(x, y, z) = f(x_1, \dots, x_{j-1}, x, x_j, \dots, x_{k-1}, y, x_k, \dots, x_{n-1}, z)$ and $g^*(x, y) = g(x_1, \dots, x_{j-1}, x, x_j, \dots, x_{k-1}, y, x_k, \dots, x_{n-1})$. Obviously,

$$g(x_1, \dots, x_{n-1}) = \int f(x_1, \dots, x_n)\, dx_n$$

and

$$g^*(x, y) = \int f^*(x, y, z)\, dz.$$

We want then to prove that g^* is TP2. For this, let $x_1 < x_2$ and $y_1 < y_2$. Then, we need to prove that g^* verifies the condition of Definition 1.64, that is, $g^*(x_1, y_1) g^*(x_2, y_2) - g^*(x_1, y_2) g^*(x_2, y_1) \geq 0$:

$$\begin{aligned} & g^*(x_1, y_1) g^*(x_2, y_2) - g^*(x_1, y_2) g^*(x_2, y_1) \\ & \quad = \int_{\mathbb{R}^2} \frac{f^*(x_2, y_2, z_2)}{f^*(x_2, y_1, z_2)} f^*(x_1, y_1, z_1) f^*(x_2, y_1, z_2)\, dz_1\, dz_2 \\ & \qquad - \int_{\mathbb{R}^2} \frac{f^*(x_1, y_2, z_2)}{f^*(x_1, y_1, z_2)} f^*(x_2, y_1, z_1) f^*(x_1, y_1, z_2)\, dz_1\, dz_2. \end{aligned} \tag{1.8}$$

Each of these integrals is separated into two by integrating on the sets $\{z_1 < z_2\}$ and $\{z_1 > z_2\}$. In the integrals over $\{z_1 > z_2\}$, we make the change of variables $(z_1, z_2) = (u_2, u_1)$. Then, for the first integral on the right in (1.8), we find

$$\begin{aligned} & \int_{\{z_1 > z_2\}} \frac{f^*(x_2, y_2, z_2)}{f^*(x_2, y_1, z_2)} f^*(x_1, y_1, z_1) f^*(x_2, y_1, z_2)\, dz_1\, dz_2 \\ & \quad = \int_{\{u_1 < u_2\}} \frac{f^*(x_2, y_2, u_1)}{f^*(x_2, y_1, u_1)} f^*(x_1, y_1, u_2) f^*(x_2, y_1, u_1)\, dz_1\, dz_2, \end{aligned}$$

and analogously for the second integral in (1.8). We find then an expression with four terms. Putting together the first and fourth on one side and the second and third on the other side, we find

$$
\begin{aligned}
&g^*(x_1,y_1)g^*(x_2,y_2) - g^*(x_1,y_2)g^*(x_2,y_1)\\
&\quad= \int_{\{z_1<z_2\}} \left(\frac{f^*(x_2,y_2,z_2)}{f^*(x_2,y_1,z_2)} - \frac{f^*(x_1,y_2,z_1)}{f^*(x_1,y_1,z_1)} \right)\\
&\qquad \times f^*(x_1,y_1,z_1) f^*(x_2,y_1,z_2)\, dz_1\, dz_2\\
&\qquad + \int_{\{z_1<z_2\}} \left(\frac{f^*(x_2,y_2,z_2)}{f^*(x_2,y_1,z_2)} - \frac{f^*(x_1,y_2,z_2)}{f^*(x_1,y_1,z_2)} \right)\\
&\qquad \times f^*(x_1,y_1,z_2) f^*(x_2,y_1,z_1)\, dz_1\, dz_2\\
&\quad= \int_{\{z_1<z_2\}} \left(\frac{f^*(x_2,y_2,z_2)}{f^*(x_2,y_1,z_2)} - \frac{f^*(x_1,y_2,z_1)}{f^*(x_1,y_1,z_1)} \right)\\
&\qquad \times \big(f^*(x_1,y_1,z_1) f^*(x_2,y_1,z_2) - f^*(x_2,y_1,z_1) f^*(x_1,y_1,z_2) \big)\, dz_1\, dz_2\\
&\qquad + \int_{\{z_1<z_2\}} \left(\frac{f^*(x_2,y_2,z_2)}{f^*(x_2,y_1,z_2)} - \frac{f^*(x_1,y_2,z_2)}{f^*(x_1,y_1,z_2)} \right.\\
&\qquad + \left. \frac{f^*(x_2,y_2,z_1)}{f^*(x_2,y_1,z_1)} - \frac{f^*(x_1,y_2,z_1)}{f^*(x_1,y_1,z_1)} \right)\\
&\qquad \times f^*(x_1,y_1,z_2) f^*(x_2,y_1,z_1)\, dz_1\, dz_2.
\end{aligned}
$$

Now, in the second integral the two differences inside the large parentheses are nonnegative because f^* is TP2 in each pair of arguments (z_2 is fixed on the first difference, while z_1 is fixed on the second). On what concerns the first integral, we have, again because f^* is TP2 in each pair of arguments,

$$
\frac{f^*(x_2,y_2,z_2)}{f^*(x_2,y_1,z_2)} \geq \frac{f^*(x_2,y_2,z_1)}{f^*(x_2,y_1,z_1)} \geq \frac{f^*(x_1,y_2,z_1)}{f^*(x_1,y_1,z_1)},
$$

so the difference is also nonnegative. As f^* is obviously nonnegative, we finally have that $g^*(x_1,y_1)g^*(x_2,y_2) - g^*(x_1,y_2)g^*(x_2,y_1) \geq 0$, as required. □

We can now extend Theorem 1.65.

Theorem 1.69 *Let $X_1,\ldots,X_n$ have a joint density function or probability function f that is TP2 in each pair of arguments. Then $X_1,\ldots,X_n$ are stochastically nondecreasing.*

Proof Denote by f_j the distribution of the vector $(X_1,\ldots,X_j)$, $j=1,\ldots,n$. It follows from the previous lemma that each f_j is TP2. In particular, f_2 is totally positive of order 2, that is, X_1 and X_2 are TP2, so, it follows from Theorem 1.65 that $\mathbf{P}(X_2 > x_2 | X_1 = x_1)$ is nondecreasing in x_1. For the case $j=3$, the function $f_3(x_1,x_2,x_3)$ is, for each fixed x_1, TP2 in x_2 and x_3, so, again according to Theorem 1.65, $\mathbf{P}(X_3 > x_3 | X_1 = x_1, X_2 = x_2)$ is nondecreasing in x_2. Analogously, for each fixed x_2, we conclude that $\mathbf{P}(X_3 > x_3 | X_1 = x_1, X_2 = x_2)$ is nondecreasing in x_1. So, it follows that $\mathbf{P}(X_3 > x_3 | X_1 = x_1, X_2 = x_2)$ is nondecreasing in x_1 and x_2. We can now recurse this argument to conclude that, for each $j=2,\ldots,n$, $\mathbf{P}(X_j > x_j | X_1 = x_1, \ldots, X_{j-1} = x_{-1})$ is nondecreasing in $x_1,\ldots,x_{j-1}$. □

Finally, we prove the association of stochastic nondecreasing variables.

Theorem 1.70 *Let $X_1, \ldots, X_n$ be stochastic nondecreasing random variables. Then $X_1, \ldots, X_n$ are associated.*

Proof All the arguments used in the proof of Theorem 1.66 are applicable in this multivariate framework, so this result follows. □

1.5.3 Some Negative Dependence Notions

Negative dependence is naturally introduced by reversing the inequalities in the definitions of positive dependencies above. There are, however, a few cares to be taken when dealing with association to remove some choices where obviously the covariances could not become negative. The negative counterpart of PQD was studied in Lehmann [58] proving the analogous of the PQD results referred before. The appearance of a negative association happened in Joag-Dev and Proschan [54] and Joag-Dev [53]. Some other negative dependence notions appeared when trying to adapt some of the positive dependencies mentioned above to their negative counterparts as in Shaked [92] or Block, Savits and Shaked [17]. We will not include this material here, as much of it is derived with arguments very similar to those used in the previous subsection. These negative counterparts have attracted some attention, especially on what concerns inequalities about partial sums. In some sense, the negative association will make moment inequalities easier to be obtained, as it implies that covariances are negative, so the covariances between partial sums are smaller than in the independent case. This general impression about negative dependencies was, probably, at the origin of less attention in earlier years, when compared to the interest in positive association, but there are, of course, quite a few specific features that finally caught the attention. We will give a brief account of negative dependence and a few basic properties in this subsection. These notions will not be developed in further chapters of this text.

Definition 1.71 Two random variables X and Y are said to be *negatively quadrant dependent* (*NQD*) if, for all $x, y \in \mathbb{R}$,

$$H(x, y) = \mathbf{P}(X > x, Y > y) - \mathbf{P}(X > x)\mathbf{P}(Y > y) \leq 0.$$

The following is obvious but useful for adapting the proofs and results from the PQD results.

Proposition 1.72 *X and Y are NQD if and only if X and $-Y$ are PQD.*

The notion relies on Hoeffding's formula (1.2) for its basic properties. Thus, Corollaries 1.5 and 1.6 have immediate counterparts.

Proposition 1.73 *Let X and Y be NQD random variables. Then*

(a) $\mathrm{Cov}(X, Y) \le 0$;
(b) *X and Y are independent if and only if* $\mathrm{Cov}(X, Y) = 0$.

The adaptation of Proposition 1.8, Theorem 1.9 and Corollary 1.10 requires a little more care because of the direction of the monotonicity of the transformations involved, as is obvious from Proposition 1.72. We will include here the statements without proof, as these are obvious taking into account the previous comment.

Proposition 1.74 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \dots, n$, X_i and Y_i are NQD. Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be such that, for each $i = 1, \dots, n$, when considered as functions of the ith coordinate alone, one is nondecreasing, and the other is nonincreasing, and let $X = f(X_1, \dots, X_n)$ and $Y = g(Y_1, \dots, Y_n)$. Then* $\mathrm{Cov}(X, Y) \le 0$.

Theorem 1.75 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \dots, n$, X_i and Y_i are PQD. Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be such that, for each $i = 1, \dots, n$, when considered as functions of the ith coordinate alone, one is nondecreasing, and the other is nonincreasing, and let $X = f(X_1, \dots, X_n)$ and $Y = g(Y_1, \dots, Y_n)$. Then X and Y are NQD.*

Corollary 1.76 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent pairs of random variables such that, for each $i = 1, \dots, n$, X_i and Y_i are PQD. Let U and V be independent and independent from $(X_1, Y_1), \dots, (X_n, Y_n)$. Let $f, g : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be such that, for each $i = 2, \dots, n+1$, when considered as functions of ith coordinate alone, one is nondecreasing, and the other is nonincreasing, and let $X = f(U, X_1, \dots, X_n)$ and $Y = g(V, Y_1, \dots, Y_n)$. Then X and Y are NQD.*

For the definition of negative association, one should take care on the choice of the arguments that are passed to the nondecreasing functions in order to avoid turning the definition useless.

Definition 1.77 The variables $X_1, \dots, X_n$ are *negatively associated* if, for all disjoint $A, B \subset \{1, \dots, n\}$ and coordinatewise nondecreasing functions $f : \mathbb{R}^{|A|} \longrightarrow \mathbb{R}$, $g : \mathbb{R}^{|B|} \longrightarrow \mathbb{R}$,

$$\mathrm{Cov}\big(f(X_i, i \in A), g(X_j, j \in B)\big) \le 0 \tag{1.9}$$

whenever the covariance exists.

A sequence of random variables X_n, $n \in \mathbb{N}$, is *negatively associated* if, for every $n \in \mathbb{N}$, the family of variables $X_1, \dots, X_n$ is negatively associated.

The following are immediate conversions to negatively dependent variables of Theorems 1.15, 1.17 and 1.18, with proofs that just rephrase the arguments used before.

Theorem 1.78 *Let $X_1, \dots, X_n$ be negatively associated random variables and consider coordinatewise nondecreasing functions $f_1, \dots, f_k : \mathbb{R}^n \longrightarrow \mathbb{R}$. Then the random variables $Y_1 = f_1(X_1, \dots, X_n), \dots, Y_k = f_k(X_1, \dots, X_n)$ are negatively associated.*

Theorem 1.79 *Let $X_1, \dots, X_n$ be negatively associated variables. These random variables are independent if and only if* $\mathrm{Cov}(X_i, X_j) = 0$, $i, j = 1, \dots, n$, $i \neq j$.

Theorem 1.80 *Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ be independent and disjoint families of the variables. Assume that $X_1, \dots, X_n$ are negatively associated variables and also that $Y_1, \dots, Y_m$ are negatively associated. Then the variables $X_1, \dots, X_n$, $Y_1, \dots, Y_m$ are negatively associated.*

To finalize this subsection, we show an example of families of random variables that are negatively associated.

Example 1.81 Let $x_1, \dots, x_n$ be distinct and fixed real numbers, and $X_1, \dots, X_n$ be random variables such that $\mathbf{X} = (X_1, \dots, X_n)$ has distribution verifying $\mathbf{P}(\mathbf{X} = (x_{\sigma(1)}, \dots, x_{\sigma(n)}) = \frac{1}{n!}$ for every permutation σ of the set $\{1, \dots, n\}$. If $n = 2$, it is easily verified that X_1 and X_2 are negatively associated. In fact, is this case, given nondecreasing functions f and g,

$$\mathrm{Cov}\big(f(X_1), g(X_2)\big) = \frac{1}{4}\big(f(x_1) - f(x_2)\big)\big(g(x_2) - g(x_1)\big) \leq 0.$$

The general case can now be proved by induction, so assume that this is true for a family of $n-1$ variables. We may assume, without loss of generality, that f and g take the same value when we permute their arguments. Denote $x_* = x_1 \wedge \cdots \wedge x_n$. Let $\mathbb{L}$ the random variable identifying the index of $\mathbf{X}$ that takes the value x_*, and $A, B \subset \{1, \dots, n\}$ be disjoint. Now we have

$$\begin{aligned}
&\mathrm{Cov}\big(f(X_i, i \in A), g(X_j, j \in B)\big) \\
&\quad = \mathrm{E}\big(\mathrm{Cov}\big(f(X_i, i \in A), g(X_j, j \in B) | \mathbb{L}\big)\big) \\
&\qquad + \mathrm{Cov}\big(\mathrm{E}\big(f(X_i, i \in A) | \mathbb{L}\big) \mathrm{E}\big(g(X_j, j \in B) | \mathbb{L}\big)\big).
\end{aligned}$$

By the induction hypothesis, $\mathrm{Cov}(f(X_i, i \in A), g(X_j, j \in B) | \mathbb{L}) \leq 0$, thus the first term on the right above is less than or equal to zero. As we have assumed that f does not change by permuting its arguments, $\mathrm{E}(f(X_i, i \in A) | \mathbb{L})$ takes only two different values, depending on whether $\mathbb{L} \in A$ or not. In the first case, where $\mathbb{L} \in A$, as f is nondecreasing, the value of the conditional expectation will be smaller than the value corresponding to the second case. Of course, the same behaviour is observed for the conditional expectation $\mathrm{E}(g(X_j, j \in B) | \mathbb{L})$. That is, the final term on the right is the covariance between a nondecreasing function of $\mathbb{L}$ and a nonincreasing function of $\mathbb{L}$, so it is less than or equal to zero.

Chapter 2
Inequalities

Abstract This chapter sets the basic tools to prove the asymptotic results that are to come in the following chapters. The first three sections are concerned with different types of inequalities on joint distributions of associated random variables and moments of sums. It is interesting that, although association is defined with a somewhat vague requirement, it is possible to recover versions for moment inequalities which are quite close to the independent case, thus paving the way to find asymptotic results that are also similar to the ones found in the independence framework. One of the key issues with association is the ability to control joint distributions from the marginal distributions using the covariance structure of the random variables. This is explored mainly in Sects. 2.5 and 2.6. The inequalities proved in these sections will provide the means to use the coupling technique, common to prove convergence results. Sect. 2.5 shows that at least the convergence in distribution is concerned with the covariance structure that completely describes the behaviour of associated variables. This chapter is a fundamental one for the remaining text.

2.1 Introduction

As usual, inequalities play an important role in the development of a theory, as much of the proving efforts are spent obtaining good estimates of suitable quantities. The first result on inequalities for associated random variables appeared in Lebowitz [55], controlling covariances of blocks of variables motivated by the need to control some Hamiltonians appearing in the Ising spin models with ferromagnetic interactions. These are essentially covariance inequalities on transformations defined with indicator functions. The natural development considering more general transformations appeared with the contributions by Newman [69, 70], where the main goal was, however, the extension of inequalities on characteristic functions to transformed associated random variables. These inequalities on characteristic functions were motivated by the study of central limit problems using the classical approach: decomposing sums into sums of blocks and trying to treat these as if they were independent. This led the work in Newman [69, 70] to one of the main tools in the association literature about convergence in distribution, Theorem 2.37 and its extension to transformed variables, Theorem 2.40. Moreover, inequality (2.26) showed the importance of covariances on the characterization of the dependence structure

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8_2,

of associated random variables, leading naturally to conditions on the decay rate of the covariances when dealing with central limit problems, invariance principles or statistical estimation issues. The meaningfulness of the coefficient $u(n)$, introduced by Cox and Grimmett [25] (see Definition 2.13), is well justified by this characteristic functions inequality. As a sort of a side effect of the previous, it is worth mentioning the control of covariances of indicator functions by the covariances of the original associated variables, with a first version appearing in Yu [109] and later extended in Cai and Roussas [24]. This inequality, as expressed in Corollary 2.36, has shown to be of significant importance in the analysis of invariance principles and also when studying the behaviour of statistical estimators. The interest on extending covariance inequalities was also developed into another direction, controlling the covariance of not necessarily monotone transformations of the variables, with upper bounds depending on the derivatives of these transformations. A general methodology for approaching the control of such covariances really follows from a few concepts introduced in Newman [69, 70], although explicit results only appeared somewhat later, Bulinski [22]. The control of moments for partial sums was first treated in Birkel [13], where the usual $n^{r/2}$ bounds were proved under suitable decay rates on the pairwise covariances, expressed through a convenient coefficient, to be introduced below in Definition 2.13, and were later extended by Shao and Yu [94]. These tools were used by Masry [65] to prove almost optimal convergence rates for density estimators based on associated samples.

Finally, exponential inequalities have been an important tool for studying convergence rates, especially for laws of large numbers or large deviations. The first such inequality was proved by Prakasa Rao [83], who obtained an upper bound that is too weak to characterize convergence results. A more useful exponential inequality for bounded associated variables appeared in Ioannides and Roussas [48] and was used by the authors to prove the first convergence rates for Strong Laws of Large Numbers. This inequality was also used in statistical estimation to characterize convergence rates when approximating distribution functions in Azevedo and Oliveira [3] or Henriques and Oliveira [41] or for density estimators [42]. The approach was based on a block decomposition of partial sums, as for the proofs of Central Limit Theorems. This was later extended, dropping the boundedness of the variables by using truncation, by Oliveira [75]. Curiously, what seems to be the weak point of this approach was the treatment of the coupling independent variables used in the proof. This aspect was further improved by Sung [98] and Xing, Yang and Liu [106], always with the motivation of improving the convergence rates for the laws of large numbers in mind, who were able to obtain exponential inequalities that lead to convergence rates in the law of large numbers arbitrarily close to the rate for independent variables.

2.2 Block and Tail Inequalities

In this section we prove two simple inequalities whose proofs use essentially the same arguments. The first one concerns joint distributions and pairwise joint distributions, separating the variables, and was proved by Lebowitz [55]. This inequality can be thought of as a first version of the control of the distance between joint distributions and the product of marginal distributions that would be found in the case of independent random variables. This idea will be extended into a finer and more clear statement in Sect. 2.5, where this distance is truly controlled in terms of the covariances of the associated variables. The second inequality concerns distribution functions and tail probabilities, proving what one could expect to find, given the definition of association of random variables, that is, that joint probabilities tend to increase with respect to the product of the corresponding marginal ones. This is a direct consequence of the fact that covariances are nonnegative, which is one of the sources of difficulties when treating associated variables: covariances of moments of sums tend to increase with respect to what could be found for independent variables, so the inequalities seem to somehow go into a wrong direction.

Given $A, B \subset \{1, \dots, n\}$ and $x_1, \dots, x_n \in \mathbb{R}$, define

$$\begin{aligned} H_{A,B}(x_1, \dots, x_n) &= \mathbf{P}(X_i > x_i, i \in A \cup B) \\ &\quad - \mathbf{P}(X_j > x_j, j \in A)\mathbf{P}(X_k > x_k, k \in B), \\ H_{j,k}(x_1, \dots, x_n) &= H_{\{j\},\{k\}}(x_1, \dots, x_n). \end{aligned}$$

We will write, for simplicity, just $H_{A,B}$ or $H_{j,k}$, unless confusion arises. Notice that $H_{j,k} = \mathrm{Cov}(\mathbb{I}_{(x_j,+\infty)}(X_j), \mathbb{I}_{(x_k,+\infty)}(X_k)) \geq 0$, as the functions $\mathbb{I}_{(x,+\infty)}(u)$ are, for each fixed x, nondecreasing in u. Moreover, as already remarked before (see page 3),

$$\begin{aligned} H_{1,2}(x_1, x_2) &= \mathbf{P}(X_1 > x_1, X_2 > x_2) - \mathbf{P}(X_1 > x_1)\mathbf{P}(X_2 > x_2) \\ &= \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2) - \mathbf{P}(X_1 \leq x_1)\mathbf{P}(X_2 \leq x_2) \\ &= \mathrm{Cov}\big(\mathbb{I}_{(-\infty,x_1]}(X_1), \mathbb{I}_{(-\infty,x_2]}(X_2)\big). \end{aligned}$$

Theorem 2.1 (Lebowitz inequality) *Let $X_1, \dots, X_n$ be associated random variables, and $A, B \subset \{1, \dots, n\}$. Then $0 \leq H_{A,B} \leq \sum_{j\in A,k\in B} H_{j,k}$.*

Proof Recall the definition of $\mathbb{I}_A = \prod_{i\in A} \mathbb{I}_{(x_i,+\infty)}(X_i)$ (see page 22) and put $S_A = \sum_{i\in A} \mathbb{I}_{(x_i,+\infty)}(X_i)$, and analogously for $\mathbb{I}_B$ and S_B. Then, obviously,

$$H_{A,B} = \mathrm{Cov}(\mathbb{I}_A, \mathbb{I}_B), \qquad \sum_{j\in A,k\in B} H_{j,k} = \mathrm{Cov}(S_A, S_B)$$

and

$$\mathrm{Cov}(S_A, S_B) = \mathrm{Cov}(S_A - \mathbb{I}_A, S_B) + \mathrm{Cov}(\mathbb{I}_A, S_B - \mathbb{I}_B) + \mathrm{Cov}(\mathbb{I}_A, \mathbb{I}_B).$$

All the $\mathbb{I}_A$, $\mathbb{I}_B$, S_A and S_B are nondecreasing transformations of the $X_1, \dots, X_n$, thus are associated, according to Theorem 1.15. It follows that $H_{A,B} = \mathrm{Cov}(\mathbb{I}_A, \mathbb{I}_B) \geq 0$. Let us now fix $j \in A$. Then

$$S_A - \mathbb{I}_A = \sum_{\substack{\ell\in A \\ \ell\neq j}} \mathbb{I}_\ell + \mathbb{I}_j\Bigg(1 - \prod_{\substack{\ell\in A \\ \ell\neq j}} \mathbb{I}_\ell\Bigg).$$

The first term on the right does not depend on X_j, while the second is the product of a nondecreasing function of X_j, $\mathbb{I}_j$, by a nonnegative factor that does not depend on X_j. Thus, it follows that $S_A - \mathbb{I}_A$ is a nondecreasing function of X_j. Repeating this argument for each choice of $j \in A$ and each $k \in B$, it follows that $S_A - \mathbb{I}_A$ and $S_B - \mathbb{I}_B$ are nondecreasing in each variable they depend on. Thus, due to the association, $\mathrm{Cov}(S_A - \mathbb{I}_A, S_B) \geq 0$ and $\mathrm{Cov}(\mathbb{I}_A, S_B - \mathbb{I}_B) \geq 0$, so the theorem follows. □

This result implies immediately a very simple characterization of independence between associated variables completely described in terms of the covariances, generalizing Theorem 1.17.

Corollary 2.2 *Let $X_1, \ldots, X_n$ be associated random variables, and $A, B \subset \{1, \ldots, n\}$. Then the random variables X_i, $i \in A$, are jointly independent of X_j, $j \in B$, if and only if* $\mathrm{Cov}(X_i, X_j) = 0$ *for every $i \in A$ and $j \in B$.*

Theorem 2.1 above gives an upper bound for a term that may be thought of as a joint covariance using pairwise covariances. Next we prove a similar result but concerning directly the joint distributions. In this case, association implies a lower bound for the joint distributions.

Theorem 2.3 *Let $X_1, \ldots, X_n$ be associated random variables, and $x_1, \ldots, x_n \in \mathbb{R}$. For every $A \subset \{1, \ldots, n\}$,*

(a) $\mathbf{P}(X_i > x_i, i \in A) \geq \prod_{i \in A} \mathbf{P}(X_i > x_i)$.
(b) $\mathbf{P}(X_i \leq x_i, i \in A) \geq \prod_{i \in A} \mathbf{P}(X_i \leq x_i)$.

Proof (a) may be rewritten as $\mathrm{E}(\prod_{i \in A} \mathbb{I}_{(x_i,+\infty)}(X_i)) \geq \prod_{i \in A} \mathrm{E}(\mathbb{I}_{(x_i,+\infty)}(X_i))$. By permutating the random variables, which does not affect association, we may assume, without loss of generality, that $A = \{1, \ldots, k\}$ for some $k \leq n$. Then,

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbb{I}_{(x_k,+\infty)}(X_k), \prod_{i=1}^{k-1} \mathbb{I}_{(x_i,+\infty)}(X_i)\right) \\
&\quad = \mathrm{E}\left(\mathbb{I}_{(x_k,+\infty)}(X_k) \prod_{i=1}^{k-1} \mathbb{I}_{(x_i,+\infty)}(X_i)\right) - \mathrm{E}\big(\mathbb{I}_{(x_k,+\infty)}(X_k)\big)\mathrm{E}\left(\prod_{i=1}^{k-1} \mathbb{I}_{(x_i,+\infty)}(X_i)\right) \\
&\quad \geq 0.
\end{aligned}$$

Iterating now this argument, (a) follows. As for (b), apply the same argument to the decreasing transformations $\mathbb{I}_{(-\infty,x_i]}(X_i) = 1 - \mathbb{I}_{(x_i,+\infty)}(X_i)$. □

A useful bound for $H_{j,k}$, in terms of the covariances of the original random variables will be proved later in Corollary 2.36.

The inequality in Theorem 2.1 concerns a special kind of nondecreasing transformations. In fact, the same is still true for the inequalities in Theorem 2.3. However, it is possible to go beyond nondecreasing and even nonmonotone transformations of associated variables if these functions are dominated by nondecreasing ones, in a convenient sense as introduced by Newman [69].

Definition 2.4 Let $f, g : \mathbb{R}^n \longrightarrow \mathbb{C}$, where $n \in \mathbb{N}$. We write $f \preceq g$ if $g - \mathrm{Re}(e^{i\alpha} f)$ is coordinatewise nondecreasing for every $\alpha \in \mathbb{R}$.

Remark 2.5 Notice that as $g = \frac{1}{2}[(g - \mathrm{Re}(f)) + (g - \mathrm{Re}(-f))]$, if $f \preceq g$, then g is real-valued and coordinatewise nondecreasing.

Remark 2.6 If f is a real-valued function, then it is obvious, by choosing $\alpha = \pi$ or $\alpha = 0$, that $f \preceq g$ if and only if both $g + f$ and $g - f$ are nondecreasing.

We first state a result allowing to deal with characteristic functions through the relation "$\preceq$".

Proposition 2.7 *Let f and g be functions defined on $\mathbb{R}^n$. Assume f is real-valued and $f \preceq g$. Let ρ be a complex-valued function defined on $\mathbb{R}$ such that, for every $u, v \in \mathbb{R}$, $|\rho(u) - \rho(v)| \leq |u - v|$. Then $\rho \circ f \preceq g$.*

Proof We need to prove that $g - \mathrm{Re}(e^{i\alpha} \rho \circ f)$ is nondecreasing. Let $t, s \in \mathbb{R}^n$ be such that $t \geq s$ (in the coordinatewise sense, that is, $t_j \geq s_j$, $j = 1, \ldots, n$). Then, taking into account the assumption on ρ, we have

$$\begin{aligned}
&\left|\mathrm{Re}\big(e^{i\alpha}\rho\big(f(t)\big)\big) - \mathrm{Re}\big(e^{i\alpha}\rho\big(f(s)\big)\big)\right| \\
&\quad \leq \left|e^{i\alpha}\rho\big(f(t)\big) - e^{i\alpha}\rho\big(f(s)\big)\right| = \left|\rho\big(f(t)\big) - \rho\big(f(s)\big)\right| \leq \left|f(t) - f(s)\right|.
\end{aligned}$$

Now, as $f \preceq g$, both $g + f$ and $g - f$ are nondecreasing. So, if $f(t) - f(s) > 0$, use the first to find $|f(t) - f(s)| = f(t) - f(s) \leq g(t) - g(s)$, and in case $f(t) - f(s) < 0$, use the later to find $|f(t) - f(s)| = f(s) - f(f) \leq g(t) - g(s)$. That is, we have in either case $|f(t) - f(s)| \leq g(t) - g(s)$. Finally, as g is nondecreasing,

$$\begin{aligned}
&\left|\big(g(t) + \mathrm{Re}\big(e^{i\alpha}\rho\big(f(t)\big)\big)\big) - \big(g(s) + \mathrm{Re}\big(e^{i\alpha}\rho\big(f(s)\big)\big)\big)\right| \\
&\quad \geq g(t) - g(s) - \left|\mathrm{Re}\big(e^{i\alpha}\rho\big(f(t)\big)\big) - \mathrm{Re}\big(e^{i\alpha}\rho\big(f(s)\big)\big)\right| \geq 0. \qquad \square
\end{aligned}$$

Notice that in the previous result we can choose $\rho(u) = e^{iu}$.

Lemma 2.8 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Let f_1, f_2, g_1, g_2 be functions defined on $\mathbb{R}^n$ for some $n \in \mathbb{N}$, such that $f_1 \preceq g_1$ and $f_2 \preceq g_2$. Then*

$$\begin{aligned}
&\left|\mathrm{Cov}\big(f_1(X_1, \ldots, X_n), f_2(X_1, \ldots, X_n)\big)\right| \\
&\quad \leq 2\left|\mathrm{Cov}\big(g_1(X_1, \ldots, X_n), g_2(X_1, \ldots, X_n)\big)\right|. \qquad (2.1)
\end{aligned}$$

Proof Assume that f_1, f_2, g_1, g_2 are real-valued functions. Then, it is enough to prove

$$\begin{aligned}
&\left|\mathrm{Cov}\big(g_1(X_1, \ldots, X_n), g_2(X_1, \ldots, X_n)\big)\right| - \mathrm{Cov}\big(f_1(X_1, \ldots, X_n), h(X_1, \ldots, X_n)\big) \\
&\quad \geq 0
\end{aligned}$$

both for $h = f_2$ and $h = -f_2$. Now, according to Remark 2.6, $g_1 + f_1$, $g_1 - f_1$, $g_2 + f_2$ and $g_2 - f_2$ are all nondecreasing functions. Thus, for both considered choices of h, g_2+h and g_2-h are nondecreasing. Notice further that, according to Remark 2.5, g_1 and g_2 are nondecreasing functions. Thus, taking into account the association of the random variables, we have

$$\begin{aligned}
&\left|\mathrm{Cov}\big(g_1(X_1,\dots,X_n),g_2(X_1,\dots,X_n)\big)\right| \\
&\qquad -\mathrm{Cov}\big(f_1(X_1,\dots,X_n),h(X_1,\dots,X_n)\big) \\
&\quad =\mathrm{Cov}\big(g_1(X_1,\dots,X_n),g_2(X_1,\dots,X_n)\big) \\
&\qquad -\mathrm{Cov}\big(f_1(X_1,\dots,X_n),h(X_1,\dots,X_n)\big) \\
&\quad =\frac{1}{2}\big[\mathrm{Cov}\big(g_1(X_1,\dots,X_n)+f_1(X_1,\dots,X_n), \\
&\qquad\quad g_2(X_1,\dots,X_n)-h(X_1,\dots,X_n)\big) \\
&\qquad +\mathrm{Cov}\big(g_1(X_1,\dots,X_n)-f_1(X_1,\dots,X_n), \\
&\qquad\quad g_2(X_1,\dots,X_n)+h(X_1,\dots,X_n)\big)\big] \\
&\quad \geq 0,
\end{aligned}$$

again due to the association of the random variables for the final step. If f_1 and f_2 are complex-valued functions, separate them into the real and imaginary parts and apply twice the previous upper bound. □

Remark 2.9 Notice that, if one of the functions f_1 or f_2 is real-valued, we may drop the coefficient 2 in (2.1). This inequality, for real functions, has appeared in Birkel [14], while the extension to complex-valued functions was considered in Newman [69].

These inequalities give us a means to prove an extension of Theorem 2.1 considering smooth but not necessarily monotone transformations of the random variables.

Theorem 2.10 (Bulinsky inequality) *Let X_n, $n\in\mathbb{N}$, be associated random variables. Assume that $A,B\subset\mathbb{N}$ are two finite sets and that f_1 and f_2 are real-valued functions defined on $\mathbb{R}^{|A|}$ and $\mathbb{R}^{|B|}$, respectively, partially differentiable with bounded first-order partial derivatives. Then*

$$\left|\mathrm{Cov}\big(f_1(X_i,i\in A),f_2(X_j,j\in B)\big)\right|\leq\sum_{i\in A,j\in B}\left\|\frac{\partial f_1}{\partial t_i}\right\|_\infty\left\|\frac{\partial f_2}{\partial t_j}\right\|_\infty\mathrm{Cov}(X_i,X_j). \tag{2.2}$$

Proof Define the following functions:

$$g_1(s_1,\dots,s_{|A|})=\sum_{i\in A}\left\|\frac{\partial f}{\partial t_i}\right\|_\infty s_i \quad\text{and}\quad g_2(s_1,\dots,s_{|B|})=\sum_{j\in B}\left\|\frac{\partial g}{\partial t_j}\right\|_\infty s_j.$$

Then g_1-f_1, g_1+f_1, g_2-f_2 and g_2+f_2 are nondecreasing functions, that is, $f_1\preceq g_1$ and $f_2\preceq g_2$. Then, applying Lemma 2.8 and taking into account Remark 2.9, the theorem follows immediately. □

A useful consequence of Theorem 2.10 gives an upper bound when considering transformed associated random variables.

Corollary 2.11 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume that $A, B \subset \mathbb{N}$ are two finite sets and that h is a bounded real-valued function defined on $\mathbb{R}$ with bounded first-order derivative. Then*

$$\left|\mathrm{Cov}\left(\prod_{i\in A} h(X_i), \prod_{j\in B} h(X_j)\right)\right| \leq \|h\|_\infty^{a+b-1} \|h'\|_\infty^2 \sum_{i\in A, j\in B} \mathrm{Cov}(X_i, X_j).$$

Proof Apply Theorem 2.10 to $f_1(s_1, \ldots, s_{|A|}) = h(s_1) \cdots h(s_{|A|})$ and $f_2(s_1, \ldots, s_{|B|}) = h(s_1) \cdots h(s_{|B|})$. □

Remark 2.12 Notice that the inequalities proved in Theorem 2.10 and Corollary 2.11 makes the control of dependence through the pairwise covariances.

Following the above remark, we close the section introducing an essential coefficient, firstly used by Cox and Grimmett [25], for the control of the dependence for associated variables.

Definition 2.13 Let X_n, $n \in \mathbb{N}$, be a sequence of random variables. Denote

$$u(n) = \sup_{k\in\mathbb{N}} \sum_{j:|j-k|\geq n} \mathrm{Cov}(X_j, X_k).$$

Remark 2.14 Notice that if we assume the random variables to be stationary, then

$$u(n) = \mathrm{E}X_1 + 2 \sum_{j=n+1}^{\infty} \mathrm{Cov}(X_1, X_j).$$

One can recognize this expression as the asymptotic variance in central limit theorems for dependent variables if we choose $n = 0$.

2.3 Moment Inequalities

Moment inequalities play a central role in proving asymptotic results for sums of random variables. As is well known, for independent random variables, the growth of $\mathrm{E}S_n^r$ is controlled by $n^{r/2}$. We find in this section sufficient conditions for this growth rate to hold for associated variables. The first such result was proved by Birkel [13] and later extended by Shao and Yu [94], obtaining the control for sums of transformations of associated variables, as described in Theorem 2.18 below. The route for the proof of both these versions is much alike, although the technicalities are somewhat different: find an appropriate control of covariances of powers of sums and use an induction argument to proceed. So, let us go into the first step for proving the main moment inequality, the control of covariances between sums of transformed variables.

Lemma 2.15 *Let $2 < p < r \le +\infty$, and f be an absolutely continuous function such that $\sup_{x\in\mathbb{R}} |f'(x)| \le C_0$. Assume that the random variables X_n, $n \in \mathbb{N}$, are associated with $\|f(X_n)\|_r < \infty$, $n \in \mathbb{N}$. Let A and B be two finite subsets of $\mathbb{N}$. Then*

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\left|\sum_{i\in A} f(X_i)\right|, \left|\sum_{j\in B} f(X_j)\right|^{p-1}\right)\right| \\
&\quad \le p\left(\mathrm{E}\left|\sum_{j\in B} f(X_j)\right|^p\right)^{(r-1)(p-2)/r_p} \left(\sum_{i\in A}\left\|f(X_j)\right\|_r\right)^{r(p-2)/r_p} \\
&\quad \times \left(C_0^2 \sum_{i\in A, j\in B} \mathrm{Cov}(X_i, X_j)\right)^{(r-p)/r_p}, \qquad (2.3)
\end{aligned}$$

where $r_p = r(p-1) - p$.

Proof Let $C_1 > 0$ be fixed and denote as usual by $|A|$ the cardinality of a set A. Define $g(t_1, \ldots, t_{|A|}) = |\sum_{i=1}^{|A|} f(t_i)|$ and

$$h(t_1, \ldots, t_{|B|}) = \left|\sum_{j=1}^{|B|} f(t_j)\right|^{p-1} \mathbb{I}_{\{|\sum_j f(t_j)| \le C_1\}} + C_1^{p-1}\mathbb{I}_{\{|\sum_j f(t_j)| > C_1\}}.$$

It is easily verified that

$$\left\|\frac{\partial g}{\partial t_i}\right\|_\infty \le C_1 \quad \text{and} \quad \left\|\frac{\partial h}{\partial t_j}\right\|_\infty \le (p-1)C_1^{p-2}C_0$$

(in fact, h is not differentiable at every point, but, as f has a bounded derivative, one may arbitrarily approximate h by a differentiable function and then take limits), and thus, from Theorem 2.10 it follows that

$$\left|\mathrm{Cov}\big(g(X_i, i \in A), h(X_j, j \in B)\big)\right| \le (p-1)C_1^{p-2}C_0^2 \sum_{i\in A, j\in B} \mathrm{Cov}(X_i, X_j).$$

To complete the proof, we now find an upper bound for

$$\begin{aligned}
&\left|\mathrm{Cov}\left(g(X_i, i \in A), \left|\sum_{j\in B} f(X_j)\right|^{p-1} - h(X_j, j \in B)\right)\right| \\
&\quad = \left|\mathrm{Cov}\left(g(X_i, i \in A), \left(\left|\sum_{j\in B} f(X_j)\right|^{p-1} - C_1^{p-1}\right)\mathbb{I}_{\{|\sum_{j\in B} f(X_j)| > C_1\}}\right)\right|.
\end{aligned}$$

By rewriting this expression in terms of mathematical expectations, this covariance is obviously less than or equal to

$$\begin{aligned}
&\max\left[\sum_{i\in A} \mathrm{E}\left(\left|f(X_i)\right|\left|\sum_{j\in B} f(X_j)\right|^{p-1} \mathbb{I}_{\{|\sum_{j\in B} f(X_j)| > C_1\}}\right),\right. \\
&\quad \left.\sum_{i\in A} \mathrm{E}\left|f(X_i)\right|\mathrm{E}\left(\left|\sum_{j\in B} f(X_j)\right|^{p-1} \mathbb{I}_{\{|\sum_{j\in B} f(X_j)| > C_1\}}\right)\right].
\end{aligned}$$

Applying the Hölder inequality to both terms, the above is still less than or equal to

$$\sum_{i\in A}\left\|f(X_i)\right\|_r\left(\mathrm{E}\left(\left|\sum_{j\in B}f(X_j)\right|^{(p-1)r/(r-1)}\mathbb{I}_{\{|\sum_{j\in B}f(X_j)|>C_1\}}\right)\right)^{(r-1)/r}.$$

Finally, using again the Hölder inequality followed by the Markov inequality, we find

$$\left|\mathrm{Cov}\left(g(X_i,i\in A),\left|\sum_{j\in B}f(X_j)\right|^{p-1}-h(X_j,j\in B)\right)\right|$$
$$\leq C_1^{-(r-p)/r}\sum_{i\in A}\left\|f(X_i)\right\|_r\left(\mathrm{E}\left|\sum_{j\in B}f(X_j)\right|^{p}\right)^{(r-1)/r}.$$

The lemma now follows by choosing

$$C_1=\left(\frac{\sum_{i\in A}\|f(X_i)\|_r\mathrm{E}(|\sum_{j\in B}f(X_j)|^p)^{(r-1)/r}}{C_0^2\sum_{i\in A,j\in B}\mathrm{Cov}(X_i,X_j)}\right)^{r/r_p}.$$ □

Remark 2.16 Notice that, in the proof of this inequality, the association of the random variables is only used through Bulinsky's inequality.

The previous lemma is essential for the proof of the main inequality in this section, as done in Shao and Yu [94]. We still need a technical lemma to achieve this extension.

Lemma 2.17 *Let* $\alpha,\beta\in(0,1)$ *and* $x,a,b,c\geq 0$. *If* $x\leq a+bx^{1-\alpha}+cx^{1-\beta}$, *then* $x\leq 2a+(4b)^{1/\alpha}+(4c)^{1/\beta}$.

Proof As

$$s^\theta t^{1-\theta}\leq s+t,\quad s,t\geq 0,\theta\in[0,1],\tag{2.4}$$

it follows that $bx^{1-\alpha}=((4b)^{(1-\alpha)/\alpha}b)^\alpha(\frac{x}{4})^{1-\alpha}\leq 4^{(1-\alpha)/\alpha}b^{1/\alpha}+\frac{x}{4}$. Analogously, $cx^{1-\beta}\leq 4^{(1-\beta)/\beta}b^{1/\beta}+\frac{x}{4}$. Using these bounds on the assumption, the lemma follows readily. □

We may now, following Shao and Yu [94], prove a moment bound for partial sums requiring a suitable decay rate on the covariance structure, expressed using the coefficient $u(n)$ introduced in Definition 2.13.

Theorem 2.18 *Let* $2<p<r\leq+\infty$, *and* f *be an absolutely continuous function such that* $\sup_{x\in\mathbb{R}}|f'(x)|\leq C_0$. *Assume that the random variables* X_n, $n\in\mathbb{N}$, *are associated*, $\mathrm{E}f(X_n)=0$, $\|f(X_n)\|_r<\infty$, $n\in\mathbb{N}$, *and* $u(n)\leq C_1n^{-\theta}$ *for some* $C_1>0$ *and* $\theta>0$. *Then, for each* $\varepsilon>0$, *there exists* K, *depending on* ε, r, p *and* θ, *such that*

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n} f(X_i)\right|^p \\
&\quad \le K\left[n^{1+\varepsilon}\max_{i\le n}\mathrm{E}\big(\big|f(X_i)\big|^p\big)+\left(n\max_{i\le n}\sum_{j=1}^{n}\big|\mathrm{Cov}\big(f(X_i),f(X_j)\big)\big|\right)^{p/2}\right. \\
&\qquad + n^{(r(p-1)-p-\theta(r-p))/(r-2)\vee(1+\varepsilon)}\max_{i\le n}\big\|f(X_i)\big\|_r^{r(p-2)/(r-2)} \\
&\qquad \left.\times\big(C_0^2C_1\big)^{(r-p)/(r-2)}\right]. \qquad (2.5)
\end{aligned}$$

Proof We shall prove the theorem by induction on the number of terms in the summation on the left side of (2.5). Notice that (2.5) is obvious for $n = 1$. So, assume that the theorem is true for each $k < n$. Denote, for each $n \in \mathbb{N}$, $T_n = \sum_{i=1}^{n} f(X_i)$ and $r_p = r(p-1) - p$, as in Lemma 2.15. Let $a \in (0, \frac{1}{2})$ be fixed, $m = [na] + 1$, and denote $k_n = [\frac{n}{2m}] + 1$. Now, decompose T_n into the sum of several blocks of length m:

$$\xi_\ell = \sum_{j=2(\ell-1)m+1}^{n\wedge(2\ell-1)m} f(X_j) \quad \text{and} \quad \eta_\ell = \sum_{j=(2\ell-m)+1}^{n\wedge 2\ell m} f(X_j), \quad \ell = 1,\dots,k_n.$$

Further, define the sums of alternating blocks:

$$T_{1,n} = \sum_{\ell=1}^{k_n}\xi_\ell \quad \text{and} \quad T_{2,n} = \sum_{\ell=1}^{k_n}\eta_\ell.$$

It is obvious, using the binomial inequality, that $\mathrm{E}|T_n|^p \le 2^{p-1}(\mathrm{E}|T_{1,n}|^p + \mathrm{E}|T_{2,n}|^p)$. We will concentrate on finding an upper bound for $\mathrm{E}|T_{1,n}|^p$, as the other mathematical expectation is analogous. Using again the binomial inequality, we find

$$\begin{aligned}
\mathrm{E}|T_{1,n}|^p &= \mathrm{E}\left(\sum_{\ell=1}^{k_n}|\xi_\ell||\xi_\ell + T_{1,n} - \xi_\ell|^{p-1}\right) \\
&\le 2^{p-2}\sum_{\ell=1}^{k_n}\mathrm{E}\big(|\xi_\ell|\big(|\xi_\ell|^{p-1} + |T_{1,n}-\xi_\ell|^{p-1}\big)\big) \\
&\le 2^{p-2}\sum_{\ell=1}^{k_n}\mathrm{E}|\xi_\ell|^p + 2^{p-2}\sum_{\ell=1}^{k_n}\mathrm{E}\big(|\xi_\ell||T_{1,n}-\xi_\ell|^{p-1}\big). \qquad (2.6)
\end{aligned}$$

Denote the first summation above by A_1 and the second by A_2. The association of the random variables enables the control of A_2 as, taking into account Lemma 2.15, it follows that

$$
\begin{aligned}
A_2 \le & \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell| \mathrm{E}|T_{1,n} - \xi_\ell|^{p-1} \\
& + \sum_{\ell=1}^{k_n} p\big(\mathrm{E}|T_{1,n} - \xi_\ell|^p\big)^{(r-1)(p-2)/r_p} \Big(m \max_{j\le m} \big\| f(X_j)\big\|_r\Big)^{r(p-2)/r_p} \\
& \times \big(C_0^2 m u(m)\big)^{(r-p)/r_p}. \qquad (2.7)
\end{aligned}
$$

Using the binomial and Hölder inequalities, we get that the first term above is less than or equal to

$$
\begin{aligned}
& 2^{p-2} \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell| \mathrm{E}\big(|\xi_\ell|^{p-1} + |T_{1,n}|^{p-1}\big) \\
& \quad \le 2^{p-2} \Bigg[\sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell|^p + \sum_{\ell=1}^{k_n} \big(\mathrm{E}|T_{1,n}|^p\big)^{(p-1)/p} \mathrm{E}|\xi_\ell| \Bigg].
\end{aligned}
$$

Put $C_2 = \max_{j\le m} \| f(X_j)\|_r^{r(p-2)/r_p} (C_0^2 u(m))^{(r-p)/r_p}$. Then, using again the binomial inequality on $\mathrm{E}|T_{1,n} - \xi_\ell|^p$, we obtain

$$
\begin{aligned}
A_2 \le & \, 2^{p-2} A_1 + 2^{p-2}\big(\mathrm{E}|T_{1,n}|^p\big)^{(p-1)/p} \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell| \\
& + p 2^{p-1} C_2 m \Bigg(\sum_{\ell=1}^{k_n} \big(\mathrm{E}|\xi_\ell|^p\big)^{1-(r-2)/r_p} + k_n \big(\mathrm{E}|T_{1,n}|^p\big)^{1-(r-2)/r_p} \Bigg). \qquad (2.8)
\end{aligned}
$$

Using (2.4), we easily see that

$$
C_2 m \sum_{\ell=1}^{k_n} \big(\mathrm{E}|\xi_\ell|^p\big)^{1-(r-2)/r_p} \le \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell|^p + k_n (C_2 m)^{r_p/(r-2)},
$$

so, replacing in (2.8), we have

$$
\begin{aligned}
A_2 \le & \, 2^{p-2}(1+2p) A_1 + 2^{p-2}\big(\mathrm{E}|T_{1,n}|^p\big)^{(p-1)/p} \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell| \\
& + p 2^{p-1} (C_2 m k_n)^{r_p/(r-2)} + p 2^{p-1} C_2 m k_n \big(\mathrm{E}|T_{1,n}|^p\big)^{1-(r-2)/r_p}.
\end{aligned}
$$

Insert now this into (2.6) and use Lemma 2.17, to find that

$$
\begin{aligned}
\mathrm{E}|T_{1,n}|^p \le & \, 2^{p-2}\big(1 + 2^{p-2}(1+2p)\big) A_1 \\
& + p 2^{2p-3} (C_1 m k_n)^{r_p/(r-2)} \\
& + 2^{2(p-2)} \big(\mathrm{E}|T_{1,n}|^p\big)^{(p-1)/p} \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell| \\
& + p 2^{2p-3} C_2 m k_n \big(\mathrm{E}|T_{1,n}|^p\big)^{1-(r-2)/r_p}
\end{aligned}
$$

$$\leq 2^{p-1}\big(1+2^{p-2}(1+2p)\big)A_1 + 2^{2p(p-1)}\left(\sum_{\ell=1}^{k_n}\mathrm{E}|\xi_\ell|\right)^p + 3p2^{2(p-1)}(C_2mk_n)^{r_p/(r-2)}.$$

To simplify the writing of the expressions, denote

$$a_p = 2^{p-1}\big(1+2^{p-2}(1+2p)\big)$$

and

$$b_p = 3p2^{2(p-1)}\max_{j\leq m}\|f(X_j)\|_r^{r(p-2)/r_p}\big(C_0^2C_1\big)^{(r-p)/(r-2)}.$$

Then, recalling the assumption on $u(n)$, we have

$$\mathrm{E}|T_{1,n}|^p \leq a_pA_1 + 2^{2p(p-1)}\left(\sum_{\ell=1}^{k_n}\big(\mathrm{E}\xi_\ell^2\big)^{1/2}\right)^p + b_p(mk_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}.$$

Denote $v_n = \max_{i\leq n}\sum_{j=1}^n|\mathrm{Cov}(f(X_i),f(X_j))|$, so that $\mathrm{E}\xi_\ell^2 \leq mv_m \leq mv_n$. Hence,

$$\mathrm{E}|T_{1,n}|^p \leq a_pA_1 + 2^{2p(p-1)}k_n^p(mv_n)^{p/2} + b_p(mk_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}. \tag{2.9}$$

Of course, $T_{2,n}$ verifies an analogous inequality. Therefore,

$$\mathrm{E}|T_n|^p \leq 2^{p-1}a_p\left(\sum_{\ell=1}^{k_n}\mathrm{E}|\xi_\ell|^p + \sum_{\ell=1}^{k_n}\mathrm{E}|\eta_\ell|^p\right) + 2^{(2p+1)(p-1)+1}k_n^p(mv_n)^{p/2} + 2^pb_p(mk_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}.$$

We may now use the induction hypothesis to bound the summations $\sum_{\ell=1}^{k_n}\mathrm{E}|\xi_\ell|^p$ and $\sum_{\ell=1}^{k_n}\mathrm{E}|\eta_\ell|^p$, so it follows that

$$\begin{aligned}\mathrm{E}|T_n|^p \leq{}& 2^pa_pk_nK\Big(m^{1+\varepsilon}\max_{j\leq n}\mathrm{E}\big|f(X_j)\big|^p + (mv_n)^{p/2} \\ &+ 2b_pm^{(r_p+\theta(p-r))/(r-2)\vee(1+\varepsilon)}\Big) \\ &+ 2^{(2p+1)(p-1)+1}k_n^p(mv_n)^{p/2} + 2^pb_p(mk_n)^{r_p/(r-2)}m^{\theta(p-r)/(r-2)}.\end{aligned}$$

Choosing $a = (2^{p-1}a_p)^{-1/\varepsilon}$ and

$$K = \max\left(\frac{2^{(2p+1)(p-1)+1}a^{p(\varepsilon-1)/2}}{1-a^{(p/2-1)(1+\varepsilon)}}, \frac{2^pb_pa^{(1+\varepsilon)r_p/(r-2)}}{1-2b_pa^{(1+\varepsilon)(p(r-1)/(r-2)-2)}}\right),$$

we get inequality (2.5), so the proof of the theorem is concluded. □

The inequality just proved in Theorem 2.18 above plays an important role in the study of convergence in distribution of empirical processes, allowing the control of the moments of increments needed to prove the tightness of the empirical process (refer to Sect. 5.4). An extension to the multivariate case, with an application to density estimation may be found in Masry [65]. A version of this inequality for LPQD random variables has been proved in Louhichi [61].

If we assume that $u(n)$ decreases fast enough, we may be more explicit about the growth rate of the third term in (2.5).

Corollary 2.19 *Under the assumptions of Theorem 2.18, if* $\theta \geq \frac{r(p-2)}{2(r-p)}$, *then, for each* $\varepsilon > 0$, *there exists* K, *depending on* ε, r, p *and* θ, *such that*

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n} f(X_i)\right|^p \\
&\quad \leq K\left[n^{1+\varepsilon} \max_{i\leq n} \mathrm{E}\big(\big|f(X_i)\big|^p\big) + \left(n \max_{i\leq n} \sum_{j=1}^{n} \big|\mathrm{Cov}\big(f(X_i), f(X_j)\big)\big|\right)^{p/2}\right. \\
&\qquad \left. + n^{p/2} \max_{i\leq n} \big\|f(X_i)\big\|_r^{r(p-2)/(r-2)} \big(C_0^2 C_1\big)^{(r-p)/(r-2)}\right]. \qquad (2.10)
\end{aligned}$$

The following result is an immediate consequence of Corollary 2.19, choosing f as the identity function and $\varepsilon = \frac{p-2}{2}$, to obtain the $n^{r/2}$ growth rate for the rth moment of partial sums as for independent random variables.

Corollary 2.20 *Let* $2 < p < r \leq \infty$, *and* X_n, $n \in \mathbb{N}$, *be centred and associated random variables satisfying* $u(n) \leq C_1 n^{-\theta}$ *for some* $C_1 > 0$, *with* $\theta \geq \frac{r(p-2)}{2(r-p)}$, *and* $\|X_n\|_r < \infty$ *for* $n \geq 1$. *Then, there exists a constant* $K = K(p,r)$ *such that, for all* $n \geq 1$,

$$\begin{aligned}
\mathrm{E}\left|\sum_{i=1}^{n} X_i\right|^p \leq K n^{p/2}&\left[\max_{i\leq n} \mathrm{E}|X_i|^p + \left(\max_{i\leq n} \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)\right)^{p/2}\right. \\
&\left. + \max_{i\leq n} \|X_i\|_r^{r(p-2)/(r-2)} C_1^{(r-p)/(r-2)}\right]. \qquad (2.11)
\end{aligned}$$

This corollary is essentially a version of the result proved by Birkel [13] that we state next for convenience later when studying invariance principles (see Sect. 5.4).

Corollary 2.21 *Let* $2 \leq p < r < \infty$, *and* X_n, $n \in \mathbb{N}$, *be centred and associated random variables such that* $u(n) \leq C_1 n^{-\theta}$ *for some* $C_1 > 0$ *and* $\theta > 0$, *and* $\sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^{r+\eta} < \infty$ *for some* $\eta > 0$. *Then, writing* $S_n = X_1 + \cdots + X_n$, *there exists a constant* $K > 0$ *such that*

$$\sup_{m\in\mathbb{N}} \mathrm{E}\big(|S_{m+n} - S_m|^p\big) \leq K n^{p/2}. \qquad (2.12)$$

Proof Just remark that, in Corollary 2.20, the constant K in (2.11) does not depend on n and the expression inside the large square brackets is bounded above by $\sup_n \mathrm{E}|X_n|^p + (u(0))^{p/2} + \sup_n \|X_i\|_r^{r(p-2)/(r-2)} C^{(r-p)/(r-2)}$, which is also independent from n. Of course, in (2.11) one does not have to start the summations at $i = 1$. □

2.4 Maximal Inequalities

An usual, the way to prove functional central limit theorems is based on suitable maximal inequalities. In fact, such inequalities are at the base of most of the arguments needed to prove the tightness of the sequences of random functions in the most popular spaces of continuous or càdlàg functions as described, for example, in Billingsley [10]. There is a huge literature on such problems that is directly inspired on this approach. Thus, we are interested on extending such inequalities to associated random variables. An application of the following results to proving functional central limit theorems will be treated later on Sect. 5.3.

The first maximal inequality appeared in Newman and Wright [71], controlling the second-order moments of maxima. We will then discuss some extensions to higher-order moments and conclude this section considering the case where we only have moments of order strictly smaller than 2.

Throughout this section we will be referring to the maxima of partial sums, so we introduce the following notation: for each $n \in \mathbb{N}$, denote $M_n = \max(S_1, \dots, S_n)$ and $M_n^* = \max(|S_1|, \dots, |S_n|)$.

Theorem 2.22 *Let $X_n, n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Then $\mathrm{E}M_n^2 \leq \mathrm{Var}(S_n) = \mathrm{E}S_n^2$ for every $n \in \mathbb{N}$.*

Proof The inequality is trivial for $n = 1$. The proof will now be completed by an induction argument. Thus, let us assume that the result holds for the maxima of partial sums involving $n - 1$ variables and define, for each $n \in \mathbb{N}$,

$$\begin{aligned}
K_n &= \min(X_2 + \cdots + X_n, X_3 + \cdots + X_n, \dots, X_n, 0),\\
L_n &= \max(X_2, X_2 + X_3, \dots, X_2 + \cdots + X_n),\\
J_n &= \max(0, L_n).
\end{aligned}$$

Notice that all these variables depend only on $X_2, \dots, X_n$. It is clear that $K_n = X_2 + \cdots + X_n - J_n$, so, as $M_n = X_1 + J_n$,

$$\begin{aligned}
\mathrm{E}M_n^2 &= \mathrm{Var}(X_1) + 2\,\mathrm{Cov}(X_1, J_n) + \mathrm{E}J_n^2\\
&= \mathrm{Var}(X_1) + 2\,\mathrm{Cov}(X_1, X_2 + \cdots + X_n) - 2\,\mathrm{Cov}(X_1, K_n) + \mathrm{E}J_n^2.
\end{aligned}$$

The random variables K_n are increasing transformations of the original variables, so according to Theorem 1.15, they are associated and associated with X_1, and thus $\mathrm{Cov}(X_1, K_n) \geq 0$. Moreover, $J_n^2 \leq L_n^2$, so it follows that

$$\mathrm{E}M_n^2 \leq \mathrm{Var}(X_1) + 2\,\mathrm{Cov}(X_1, X_2 + \cdots + X_n) + \mathrm{E}L_n^2.$$

Finally, by the induction hypothesis, $\mathrm{E}L_n^2 \leq \mathrm{Var}(X_2 + \cdots + X_n)$, so the theorem follows. □

Given $j, n \in \mathbb{N}$, define

$$T_{j,n} = \begin{cases} j\text{th largest among } (S_1, \ldots, S_n) & \text{if } j \le n, \\ \min(S_1, \ldots, S_n) & \text{if } j > n. \end{cases}$$

Obviously, $T_{n,n} = \min(S_1, \ldots, S_n)$ and $T_{1,n} = \max(S_1, \ldots, S_n)$. The following gives a general inequality.

Lemma 2.23 *Let X_n, $n \in \mathbb{N}$, be associated random variables, and m a nondecreasing function such that $m(0) = 0$. Then, for all $j, n \in \mathbb{N}$,*

$$\mathrm{E}\int_0^{T_{j,n}} u m(du) \le \mathrm{E}\bigl(S_n m(T_{j,n})\bigr). \tag{2.13}$$

Thus, for every $c > 0$,

$$\lambda \mathbf{P}(T_{j,n} \ge c) \le \int_{\{T_{j,n} \ge c\}} S_n \, d\mathbf{P}. \tag{2.14}$$

Proof Write

$$m(T_{j,n})S_n = \sum_{k=0}^{n-1} S_{k+1}\bigl(m(T_{j,k+1}) - m(T_{j,k})\bigr) + \sum_{k=1}^{n-1}(S_{k+1} - S_k)m(T_{j,k}). \tag{2.15}$$

By the definition of $T_{j,n}$, if $k < j$, we have either $T_{j,k} = T_{j,k+1}$ or $S_{k+1} = Y_{j,k+1}$. Analogously, if $k \ge j$, we have either $T_{j,k} = T_{j,k+1}$ or $S_{k+1} \ge Y_{j,k+1}$. Thus, for every $k \ge 1$,

$$S_{k+1}\bigl(m(T_{j,k+1}) - m(T_{j,k})\bigr) \ge T_{j,k+1}\bigl(m(T_{j,k+1}) - m(T_{j,k})\bigr) \ge \int_{T_{j,k}}^{T_{j,k+1}} u m(du).$$

Now, take expectations, sum these terms, and recall that $T_{j,n} = S_n$ and that, due to the association of the random variables and that S_n are nondecreasing transformations, all the terms in the second summation of (2.15) are nonnegative, so (2.13) follows. Finally, to prove (2.14), choose $m(u) = \mathbb{I}_{[c,+\infty)}(u)$ and apply (2.13). □

Remark 2.24 The previous lemma was proved by Newman and Wright [72] for a somewhat more general framework. In fact, all that is used in the proof is just that $\mathrm{Cov}(X_{n+1}, m(S_1, \ldots, S_n)) \ge 0$. A sequence of random variables verifying this condition was called in Newman and Wright [72] a demimartingale.

We may now prove an extended version of the inequality in Theorem 2.22.

Theorem 2.25 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables. Then $\mathrm{E}T_{j,n}^2 \le \mathrm{E}S_n^2$.*

Proof Define the random variables $Z_1 = 0$ and $Z_k = \sum_{i=n-k+2}^{n} X_i$, $k = 2, 3, \dots,$ $n+1$, and, for $j \le n$, $Z_{j,n}$ the jth largest among $(Z_1, \dots, Z_n)$. Then, from Theorem 2.23 with $m(u) = u$ we have

$$\frac{1}{2}\mathrm{E}Z_{n-j+1,n}^2 \le \mathrm{E}(Z_n T_{n-j+1,n}) \le \mathrm{E}(Z_{n+1} T_{n-j+1,n}),$$

so $\mathrm{E}(Z_{n+1} - T_{n-j+1,n})^2 \le \mathrm{E}Z_{n+1}^2$, which is equivalent to the statement of this theorem. □

We may improve the upper bound for $\mathrm{E}M_n^2$ if we assume a more precise convergence decrease rate on the covariance of the random variables. The following result appeared much later in the literature (Yang, Su and Yu [108]) and was motivated by the search for convenient characterizations of the convergence rate in Strong Laws of Large Numbers.

Theorem 2.26 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables such that*

$$\sum_{i=1}^{\infty} u^{1/2}\big(2^i\big) < \infty. \tag{2.16}$$

Then, there exists a positive constant C such that

$$\mathrm{E}M_n^2 \le Cn\Big(\max_{k\le n} \mathrm{E}X_k^2 + 1\Big). \tag{2.17}$$

Proof For each $n \in \mathbb{N}$, define the sequence of random variables $Y_{i,n} = X_i \mathbb{I}_{[1,n]}(i)$, $i = 1, 2, \dots$. These random variables are obtained as nondecreasing transformations of the original ones and thus are associated. Consider, on the sequel, n fixed. Given $j, k \in \mathbb{N}$, define $S_j(k) = Y_{j+1,n} + \cdots + Y_{j+k,n}$ and, to deal with the nonstationarity, $s_k = \sup_{j\in\mathbb{N}} \|S_j(k)\|_2$, where $\|S_j(k)\|_2 = (\mathrm{E}(S_j(k)))^{1/2}$ is the L^2 norm of $S_j(k)$. Then, obviously,

$$\begin{aligned}
\big\|S_j(2k)\big\|_2 &\le \big\|S_j(k) + S_{j+k+[k^{1/3}]}(k)\big\|_2 + \big\|S_{j+k}\big(\big[k^{1/3}\big]\big)\big\|_2 + \big\|S_{j+2k}\big(\big[k^{1/3}\big]\big)\big\|_2 \\
&\le \big\|S_j(k) + S_{j+k+[k^{1/3}]}(k)\big\|_2 + 2k^{1/3} \sup_{i\in\mathbb{N}} \|Y_{i,n}\|_2 \\
&= \big\|S_j(k) + S_{j+k+[k^{1/3}]}(k)\big\|_2 + 2k^{1/3} \max_{i\le n} \|X_i\|_2.
\end{aligned}$$

Now we use the association of the random variables, implying that the covariances are nonnegative, so that

$$\begin{aligned}
\big\|S_j(k) + S_{j+k+[k^{1/3}]}(k)\big\|_2^2 &= \mathrm{E}\big(S_j(k) + S_{j+k+[k^{1/3}]}(k)\big)^2 \\
&\le 2s_k^2 + 2\sum_{i=j+1}^{j+k} \sum_{\ell=j+k+[k^{1/3}]+1}^{\infty} \mathrm{Cov}(Y_{i,n}, Y_{\ell,n}) \\
&\le 2s_k^2 + 2\sum_{i=j+1}^{j+k} u\big(\big[k^{1/3}\big]\big) = 2s_k^2 + 2ku\big(\big[k^{1/3}\big]\big).
\end{aligned}$$

Thus, inserting this in the previous majorization, it follows

$$s_{2k} \leq \sqrt{2} s_k + \sqrt{2ku\big(\big[k^{1/3}\big]\big)} + 2k^{1/3} s_1 .$$

We now use recursively the previous bound to find

$$\begin{aligned}
s_{2^r} &\leq 2^{r/2} s_1 + 2 s_1 \sum_{i=0}^{r-1} 2^{(r-1-i)/2+i/3} + 2^{r/2} \sum_{i=0}^{r-1} \sqrt{u\big(\big[2^{i/3}\big]\big)} \\
&\leq 2^{r/2} s_1 + 2^{(r+1)/2} s_1 \sum_{i=0}^{\infty} 2^{-i/6} + 2^{r/2} \sum_{i=0}^{\infty} \sum_{j=3i}^{3i+2} \sqrt{u\big(\big[2^{j/3}\big]\big)} \\
&\leq 14 \times 2^{r/2} s_1 + 3 \times 2^{r/2} \sum_{i=0}^{\infty} \sqrt{u\big(\big[2^{i}\big]\big)} \\
&\leq C 2^{r/2} (s_1 + 1),
\end{aligned}$$

where $C = \max(14, 3\sum_{i=0}^{\infty} \sqrt{u([2^i])})$. Assume now that $2^r \leq k < 2^{r+1}$. Then, due to the association,

$$\mathrm{E} S_k^2 \leq \mathrm{E}\big(S_0^2\big(2^{r+1}\big)\big) \leq s_{2^{2+1}}^2 \leq C 2^{(r+1)/2}\big(s_1^2 + 1\big) \leq 2C 2^{2/r}\big(s_1^2 + 1\big),$$

so, from Theorem 2.22 the result follows. □

Assumption (2.16) that has been used for the first time in Yang [107] is a rather mild one. In fact, (2.16) is verified if $u(n)$ is of order $(\log n)^{-2}(\log\log n)^{-3}$, much weaker than the typical hypothesis used in Sect. 2.3, where a polynomial decrease rate was often assumed.

The result above assumes the existence of second-order moments, but, being a statement that does not require stationarity, these moments may be unbounded. Of course, the case where the second-order moments are bounded is included in the framework of Theorem 2.26. As could be expected, it is possible to prove a version of the upper bound better adapted to this situation, which will be explored later when dealing with truncated variables, allowing to explore the behaviour of the truncating sequence.

Theorem 2.27 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables such that* $\sup_{n\in\mathbb{N}} \mathrm{E} X_n^2 < \infty$ *and*

$$K = \sup_{j\in\mathbb{N}} \sum_{k:k-j>1}^{\infty} \mathrm{Cov}^{1/2}(X_j, X_k) < \infty. \tag{2.18}$$

Then, for every $n \in \mathbb{N}$,

$$\mathrm{E} M_n^2 \leq 2n \sup_{n\in\mathbb{N}} \mathrm{E} X_n^2 + 4nK \Big(\sup_{n\in\mathbb{N}} \mathrm{E} X_n^2\Big)^{1/2} .$$

Proof First remark that, obviously, for every $j \le n$,

$$S_j^2 \le \max\Big(\Big(\min_{k\le n} S_k\Big)^2, \Big(\max_{k\le n} S_k\Big)^2\Big) \le \Big(\min_{k\le n} S_k\Big)^2 + \Big(\max_{k\le n} S_k\Big)^2,$$

so it follows that $M_n^2 \le (\min_{k\le n} S_k)^2 + (\max_{k\le n} S_k)^2 \le 2(\max_{k\le n} S_k)^2$. Thus, applying Theorem 2.25, we have that $\mathrm{E}(\max_{k\le n} S_k)^2 \le \mathrm{E}S_n^2$, so

$$\mathrm{E}\Big(\max_{k\le n} S_k^2\Big) \le 2\mathrm{E}S_n^2 = 2\sum_{j,k=1}^{n} \mathrm{Cov}(X_j, X_k).$$

Define $K_1 = \sup_{n\in\mathbb{N}} \mathrm{E}X_n^2$. Then, it is obvious that

$$\mathrm{Cov}(X_j, X_k) \le \big(\mathrm{E}X_j^2 \mathrm{E}X_k^2\big)^{1/2} \le K_1$$

and

$$\mathrm{Cov}(X_j, X_k) \le \big(K_1 \,\mathrm{Cov}(X_j, X_k)\big)^{1/2}.$$

Thus,

$$\sum_{j,k=1}^{n} \mathrm{Cov}(X_j, X_k) = \sum_{j=1}^{n} \mathrm{E}X_j^2 + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \mathrm{Cov}(X_j, X_k) \le K_1 n + 2K_1^{1/2} n K,$$

so the result follows. □

We now prove an extension to associated random variables of a well-known maximal inequality under independence.

Theorem 2.28 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Then, for all $\lambda > 0$ and $n \in \mathbb{N}$,*

$$\mathbf{P}\big(M_n^* \ge \lambda s_n\big) \le 2\mathbf{P}\big(|S_n| \ge (\lambda - \sqrt{2}) s_n\big). \tag{2.19}$$

Proof Define $M_n^+ = \max(0, S_1, \ldots, S_n)$. Given real numbers $x < y$, we have

$$\begin{aligned}
&\mathbf{P}\big(M_n^+ \ge y\big)\\
&\quad \le \mathbf{P}(S_n \ge x) + \mathbf{P}\big(M_{n-1}^+ \ge y, M_{n-1}^+ - S_n > y - x\big)\\
&\quad \le \mathbf{P}(S_n \ge x) + \mathbf{P}\big(S_{n-1}^+ \ge y\big)\mathbf{P}\big(M_{n-1}^+ - S_n > y - x\big)\\
&\quad \le \mathbf{P}(S_n \ge x) + \mathbf{P}\big(M_n^+ \ge y\big)\frac{\mathrm{E}(M_{n-1}^+ - S_n)^2}{(y-x)^2},
\end{aligned}$$

using Theorem 2.3, as M_{n-1}^+ and $S_n - M_{n-1}^+$ are associated. The mathematical expectation above may be rewritten as

$$\mathrm{E}\big(M_{n-1}^+ - S_n\big)^2 = \mathrm{E}\big[\max(X_n, X_n + X_{n-1}, \ldots, X_n + \cdots + X_1)^2\big] \le \mathrm{E}S_n^2,$$

taking into account Theorem 2.22. Now, if $(y-x)^2 \ge s_n^2 = \mathrm{E}S_n^2$, it follows then that

$$\mathbf{P}\big(M_n^+ \ge y\big) \le \frac{(y-x)^2}{(y-x)^2 - s_n^2}\mathbf{P}(S_n \ge x). \tag{2.20}$$

Repeating these arguments with $M_n^- = \max(0, -S_1, \ldots, -S_n)$ and adding to (2.20), we find, whenever $y - x \geq \sqrt{2}s_n$,

$$\mathbf{P}\big(M_n^* \geq y\big) \leq 2\mathbf{P}\big(|S_n| \geq x\big). \tag{2.21}$$

Finally, choosing $y = \lambda s_n$ and $x = (\lambda - \sqrt{2})s_n$, it follows that

$$\mathbf{P}\big(M_n^* \geq \lambda s_n\big) \leq 2\mathbf{P}\big(|S_n| \geq (\lambda - \sqrt{2})s_n\big).$$

□

Notice that (2.19) has exactly the same form as in the independent case (see Sect. 10 in Billingsley [10], for example).

The previous theorem implies, in a very simple way, a maximal inequality for higher-order moments, thus extending Theorem 2.22.

Corollary 2.29 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables with finite moments of order $p \geq 2$. Then $\mathrm{E}M_n^p \leq \sqrt{2}(\mathrm{E}|S_n|^p)^{1/p} + 2\mathrm{E}S_n^p$.*

Proof Recall that $M_n^+ = \max(0, S_1, \ldots, S_n)$. As M_n^+ is a nonnegative variable, by (2.19) and Hölder's inequality, it follows that, with $s_n^2 = \mathrm{E}S_n^2$,

$$\begin{aligned}
&\mathrm{E}\big(M_n^+\big)^p \\
&\quad = \int_0^{+\infty} \mathbf{P}\big(\big(M_n^+\big)^p > y\big)\,dy \leq \sqrt{2}s_n + 2\int_{\sqrt{2}s_n}^{+\infty} \mathbf{P}\big(|S_n|^p > y - \sqrt{2}s_n\big)\,dy \\
&\quad \leq \sqrt{2}s_n + 2\mathrm{E}S_n^p \leq \sqrt{2}\big(\mathrm{E}|S_n|^p\big)^{1/p} + 2\mathrm{E}S_n^p.
\end{aligned}$$

Obviously, the same applies to $M_n^- = \max(0, -S_1, \ldots, -S_n)$, so the result follows. □

If we only have moments of order smaller than 2, it is still possible to prove bounds for the corresponding moments of maxima. These will be useful later on, when analysing extensions of Strong Laws of Large Numbers. In this case we need a preparatory lemma providing control on the tail of M_n. In order to state the result, we need to introduce some extra notation. Recall the definition of $H_{j,k}$ (see page 3): $H_{j,k}(x, y) = \mathbf{P}(X_j > x, X_k > y) - \mathbf{P}(X_j > x)\mathbf{P}(X_k > y)$. Next, given $v > 0$, define $g_v(u) = \max(\min(u, v), -v)$ and

$$G_{j,k}(v) = \mathrm{Cov}\big(g_v(X_j), g_v(X_k)\big) = \int\int_{[-v,v]^2} H_{j,k}(x, y)\,dx\,dy.$$

It is obvious that $G_{j,k}(+\infty) = \mathrm{Cov}(X_j, X_k)$. Moreover, as each g_v is an increasing function bounded by v^2, it follows, assuming that the random variables X_n are associated, that $0 \leq G_v(x, y) \leq v^2$.

Lemma 2.30 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume that there exists a nonnegative random variable Y such that, for every $x > 0$, $\sup_{n\in\mathbb{N}} \mathbf{P}(|X_n| > x) \le \mathbf{P}(Y > x)$. Then, for every $x, m > 0$,*

$$\begin{aligned}\mathbf{P}(M_n > x) \le \frac{4n}{x^2}\mathrm{E}\big(Y^2\mathbb{I}_{\{Y\le m\}}\big) + \frac{4n}{x}\mathrm{E}(Y\mathbb{I}_{\{Y>m\}}) \\ + \frac{4nm^2}{x^2}\mathbf{P}(Y > m) + \frac{8}{x^2}\sum_{\substack{j,k=1\\ j\ne k}}^{n} G_{j,k}(m).\end{aligned}$$

Proof Let $m > 0$ be fixed and define, for each $j \in \mathbb{N}$,

$$\begin{aligned}X_{1,j} &= g_m(X_j), &\qquad S_{1,n} &= \sum_{j=1}^{n}(X_{1,j} - \mathrm{E}X_{1,j}),\\ M_{1,n} &= \max_{k\le n} S_{1,k}, &\qquad X_{2,j} &= X_j - X_{1,j}.\end{aligned}$$

As, obviously,

$$M_n \le M_{1,n} + \sum_{j=1}^{n}\big(|X_{2,j}| + \mathrm{E}|X_{2,j}|\big),$$

it follows from Markov's inequality that

$$\mathbf{P}(M_n > x) \le \mathbf{P}\left(M_{1,n} > \frac{x}{2}\right) + \frac{4}{x}\sum_{j=1}^{n}\mathrm{E}|X_{2,j}| \le \frac{4}{x^2}\mathrm{E}M_{1,n}^2 + \frac{4}{x}\sum_{j=1}^{n}\mathrm{E}|X_{2,j}|.$$

The random variables $X_{1,n}$ are associated, so we may apply Theorem 2.22 to bound the first term on the right above, to find

$$\begin{aligned}\mathbf{P}(M_n > x) &\le \frac{4}{x^2}\mathrm{E}S_{1,n}^2 + \frac{4}{x}\sum_{j=1}^{n}\mathrm{E}|X_{2,j}|\\ &\le \frac{4}{x^2}\sum_{j=1}^{n}\mathrm{E}X_{1,j}^2 + \frac{8}{x^2}\sum_{\substack{j,k=1\\ j\ne k}}^{n}\mathrm{Cov}(X_{1,j}, X_{1,k}) + \frac{4}{x}\sum_{j=1}^{n}\mathrm{E}|X_{2,j}|.\end{aligned}$$

Finally, notice that $\mathrm{E}X_{1,j}^2 = \int \mathbf{P}(X_{1,j}^2 > x)\,dx \le \mathrm{E}(Y^2\mathbb{I}_{\{Y\le m\}}) + m^2\mathbf{P}(Y > m)$ and, analogously, $\mathrm{E}|X_{2,j}| \le \mathrm{E}((|X_j| - m)\mathbb{I}_{\{|X_j|>m\}}) \le \mathrm{E}(Y\mathbb{I}_{\{Y>m\}})$. □

Now we may prove the announced maximal inequality.

Theorem 2.31 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume that there exists a nonnegative random variable Y such that, for some $p \in (1, 2)$, $\mathrm{E}Y^p < \infty$*

and, for every $x > 0$, $\sup_{n\in\mathbb{N}} \mathbf{P}(|X_n| > x) \le \mathbf{P}(Y > x)$. *Then, there exists a constant* $c_p > 0$, *depending only on* p, *such that*

$$\mathrm{E}\Big(\max_{k\le n} |S_k|^p\Big) \le c_p \left(n\mathrm{E}Y^p + \sum_{\substack{j,k=1\\ j\ne k}}^{n} \int_0^\infty x^{p-3} G_{j,k}(x)\, dx \right). \tag{2.22}$$

Proof Applying Lemma 2.30 with $m = x$ and taking into account that $p < 2$, we have

$$\begin{aligned}
\mathrm{E}M_n^p &= p \int_0^\infty x^{p-1} \mathbf{P}(M_n > x)\, dx \\
&\le p \int_0^\infty x^{p-1} \left(\frac{4n}{x^2} \mathrm{E}\big(Y^2 \mathbb{I}_{\{Y\le x\}}\big) + \frac{4n}{x} \mathrm{E}(Y \mathbb{I}_{\{Y>x\}}) + 4n\mathbf{P}(Y > x) \right) dx \\
&\quad + p \int_0^\infty x^{p-1} \frac{8}{x^2} \sum_{\substack{j,k=1\\ j\ne k}}^{n} G_{j,k}(x)\, dx \\
&\le 8n \int_0^\infty x^{p-3} \mathrm{E}\big(Y^2 \mathbb{I}_{\{Y\le x\}}\big) + x^{p-2} \mathrm{E}(Y \mathbb{I}_{\{Y>x\}}) + x^{p-1} \mathbf{P}(Y \ge x)\, dx \\
&\quad + 8 \int_0^\infty x^{p-3} \sum_{\substack{j,k=1\\ j\ne k}}^{n} G_{j,k}(x)\, dx \\
&\le 8n \left(\frac{1}{2-p} + \frac{1}{p-1} + \frac{1}{p} \right) \mathrm{E}Y^p + 8 \int_0^\infty x^{p-3} \sum_{\substack{j,k=1\\ j\ne k}}^{n} G_{j,k}(x)\, dx
\end{aligned}$$

and choose $c_p = 8 \max(\frac{1}{2-p} + \frac{1}{p-1} + \frac{1}{p}, 1)$. Finally, notice that one may replace the variables X_n by $-X_n$, keeping the association, so the previous inequality also applies, thus proving the theorem. □

2.5 Characteristic Functions

This section presents a few inequalities, first proved by Newman [68] (see also Newman [70] for a more complete presentation), controlling the distance between joint distributions and the product of marginal distributions, based on characteristic functions. As it will be shown, this distance is completely characterized by the covariances of the variables, and thus, when seeking for asymptotic results, it becomes natural to look for assumptions on the covariance structure. These inequalities will play a major role in proving central limit theorems for associated variables. An analogous inequality concerning moment generating functions is also proved. This later inequality, first used by Dewan and Prakasa Rao [30] in a different context, will be a useful tool for proving exponential inequalities of the next section and to derive convergence rates.

We start by setting some notation.

Definition 2.32 Given the random variable X, we denote its *characteristic function* by

$$\varphi_X(u) = \mathrm{E}e^{iuX}, \quad u \in \mathbb{R}.$$

Given random variables $X_1, \ldots, X_n$, we denote their *joint characteristic function* by

$$\varphi_{(X_1,\ldots,X_n)}(u_1, \ldots, u_n) = \mathrm{E}e^{i(u_1X_1+\cdots+u_nX_n)}, \quad u_1, \ldots, u_n \in \mathbb{R}.$$

Given a set $A \subset \{1, \ldots, n\}$, we denote

$$\varphi_A(u_1, \ldots, u_n) = \mathrm{E}e^{i\sum_{j\in A} u_jX_j}.$$

If $A = \{j\}$, we denote, for simplicity, φ_A by φ_j.

We first prove the inequality that describes the control of the distance between joint distributions and the product of marginal distributions for two random variables. The proof of the main result in this section, Theorem 2.37, is built on this version proceeding by induction.

Lemma 2.33 *Let X and Y be associated random variables. Then, for every $u, v \in \mathbb{R}$,*

$$\left|\mathrm{E}e^{iuX+ivY} - \mathrm{E}e^{iuX}\mathrm{E}e^{ivY}\right| \le |uv| \operatorname{Cov}(X, Y). \tag{2.23}$$

Proof The expression inside the absolute value may be rewritten as $\operatorname{Cov}(e^{iuX}, e^{ivY})$. Now, representing by $\mathbf{P}_X$, $\mathbf{P}_Y$ and $\mathbf{P}_{(X,Y)}$ the distributions of X, Y and (X, Y), respectively, and recalling that $H(s,t) = \mathbf{P}(X > s, Y > t) - \mathbf{P}(X > s)\mathbf{P}(Y > t)$ and integrating by parts (see Theorem C.4), we find

$$\begin{aligned}
\operatorname{Cov}\big(e^{iuX}, e^{ivY}\big) &= \int\!\!\int e^{ius+ivt}(\mathbf{P}_{(X,Y)} - \mathbf{P}_X \otimes \mathbf{P}_Y)(ds\,dt) \\
&= \int\!\!\int \frac{\partial^2}{\partial s\,\partial t} e^{ius+ivt} H(s,t)\,ds\,dt \\
&= \int\!\!\int iue^{ius}ive^{ivt} H(s,t)\,ds\,dt.
\end{aligned}$$

Noticing that due to the association of the variables, H is a nonnegative function, the lemma follows immediately from

$$\left|\operatorname{Cov}\big(e^{iuX}, e^{ivY}\big)\right| \le |uv| \int\!\!\int H(s,t)\,ds\,dt = |uv| \operatorname{Cov}(X, Y),$$

using Hoeffding's formula (1.2). □

Remark 2.34 Notice that the proof depends only on the fact that H has constant sign. Thus, the previous result holds for positively dependent variables or for negatively dependent variables taking, in this later case, the absolute value of the covariance on the left-hand side of (2.23). Of course, for general random variables, the upper bound would be

$$|uv| \int \int \bigl|H(s,t)\bigr|\, ds\, dt.$$

Before stating an extension of this lemma, it is useful to relate the covariances H with the covariances between the original random variables. This is a direct consequence of the two-dimensional version of the classical Berry–Esséen inequalities, given in Theorem A.2.

Lemma 2.35 *Let X and Y be associated random variables with absolutely continuous distributions. Assume that the marginal densities f_X and f_Y are bounded by M. Then, for every $T > 0$,*

$$H(x,y) = \operatorname{Cov}\bigl(\mathbb{I}_{(-\infty,x]}(X), \mathbb{I}_{(-\infty,y]}(Y)\bigr) \le M^* \left(T^2 \operatorname{Cov}(X,Y) + \frac{1}{T}\right), \tag{2.24}$$

where $M^* = \max(\frac{2}{\pi^2}, 45M)$.

Proof Using Corollary A.3 together with (2.23), the lemma follows immediately. □

Optimizing the choice of T on the previous result, we find the following important inequality.

Corollary 2.36 *Under the same assumptions as in Lemma 2.35, if* $\operatorname{Cov}(X,Y) > 0$, *we have that*

$$\operatorname{Cov}\bigl(\mathbb{I}_{(-\infty,x]}(X), \mathbb{I}_{(-\infty,y]}(Y)\bigr) \le \frac{1}{M^*} \operatorname{Cov}^{1/3}(X,Y). \tag{2.25}$$

Proof In (2.24), choose $T = (2\operatorname{Cov}(X,Y))^{-1/3}$, and the inequality follows. □

This inequality plays an important role while studying invariance principles and the asymptotics for the density and regression estimators that depend on transformations using indicator functions on the sequence of random variables. In fact, (2.25) enables the control of covariances between indicator functions using covariances between the original random variables. Thus, it gives a way to obtain sufficient conditions expressed in terms of the initial variables.

Now we extend Lemma 2.33 to any number of associated variables.

Theorem 2.37 (Newman inequality) *Let X_n, $n \in \mathbb{N}$, be associated variables. Then, for all $n \in \mathbb{N}$ and $u_1, \ldots, u_n \in \mathbb{R}$,*

$$\left|\varphi_{(X_1,\ldots,X_n)}(u_1,\ldots,u_n) - \prod_{j=1}^{n} \varphi_j(u_j)\right| \leq \frac{1}{2} \sum_{\substack{j,k=1 \\ j\neq k}}^{n} |u_j u_k| \operatorname{Cov}(X_j, X_k). \tag{2.26}$$

Proof When $n = 2$, inequality (2.26) reduces to (2.23), so we may proceed by induction on n to prove this theorem. Assume then that (2.26) holds whenever there are only $n-1$ variables involved. To prove the inequality for n variables, split the set $\{1, \ldots, n\}$ in the following way:

(a) if all the $u_1, \ldots, u_n$ have the same sign, take $A = \{1, \ldots, n-1\}$ and $B = \{n\}$;
(b) if not all the $u_1, \ldots, u_n$ have the same sign, take $A = \{j \in \{1, \ldots, n\} : u_j > 0\}$ and $B = \{1, \ldots, n\} \setminus A$.

Define now the variables $U = \sum_{j\in A} |u_j| X_j$ and $V = \sum_{j\in B} |u_j| X_j$. Notice that these variables are increasing transformations of the X_n's, so they are still associated. Moreover, we can write

$$\varphi_{(X_1,\ldots,X_n)}(u_1,\ldots,u_n) = \mathrm{E} e^{i(U-V)} = \varphi_{U-V}(1),$$
$$\varphi_A(u_j, j \in A) = \varphi_U(1) \quad \text{and} \quad \varphi_B(u_j, j \in B) = \varphi_V(-1).$$

Then,

$$\begin{aligned}
&\left|\varphi_{(X_1,\ldots,X_n)}(u_1,\ldots,u_n) - \prod_{j=1}^{n} \varphi_j(u_j)\right| \\
&\quad \leq \left|\varphi_{U-V}(1) - \varphi_U(1)\varphi_V(-1)\right| + \left|\varphi_U(1)\right| \left|\varphi_V(-1) - \prod_{j\in B} \varphi_j(u_j)\right| \\
&\qquad + \left|\prod_{j\in B} \varphi_j(u_j)\right| \left|\varphi_U(1) - \prod_{j\in A} \varphi_j(u_j)\right|.
\end{aligned} \tag{2.27}$$

Of course, characteristics functions have absolute values bounded by 1, so the second and third terms may be bounded using the induction hypothesis. It remains to bound the first term on (2.27): as U and V are associated, we refer to Lemma 2.33 and, with respect to the notation used in this lemma, choose, for the case (a), $u = v = 1$ if the u_j are positive or $u = v = -1$ if the u_j are negative, and for the case (b), $u = 1$ and $v = -1$. Applying now Lemma 2.33 and the induction hypothesis, we immediately get (2.27). □

Remark 2.38 A close look at the proof shows that the association assumption can be weakened. In fact, what is used throughout the proof is the fact that linear combinations with nonnegative coefficients (notice that we multiply by -1 the coefficients that are negative) of the random variables have nonnegative covariance. Thus, the previous result holds if the random variables are LPQD (see Definition 1.58).

Remark 2.39 Notice that, as far as what convergence in distribution is concerned, inequality (2.26) means that the covariance structure of the random variables completely determines the properties of the approximation of joint distributions to independence. This remark is at the heart of most of the results included in Chap. 4. It also justifies that, for random variables, it is natural to seek for assumptions on the behaviour of these covariances.

Next, we prove an extension of Theorem 2.37, allowing for applications going beyond associated random variables themselves, by considering suitable transformations of the variables. Of course, Theorem 2.37 still applies if we consider transformations of the initial variables that are either all increasing or all decreasing as, according to Theorem 1.15, such transformations keep the association. However, it is possible to prove a version of (2.26) for nonmonotone transformations of associated variables, using Lemma 2.8, if these functions are dominated by nondecreasing ones, as described by the relation "$\preceq$" introduced in Definition 2.4.

Theorem 2.40 *Let Y_n, $n \in \mathbb{N}$, be associated random variables. Assume that, for each $n \in \mathbb{N}$, f_n, g_n are real-valued functions such that $f_n \preceq g_n$, and denote $X_n = f_n(Y_1, Y_2, \ldots)$ and $X_n^* = g_n(Y_1, Y_2, \ldots)$. Then, for every $n \in \mathbb{N}$, given $A, B \subset \{1, \ldots, n\}$ and $u_1, \ldots, u_n \in \mathbb{R}$,*

$$\begin{aligned}
&\left|\varphi_{A\cup B}(u_1, \ldots, u_n) - \varphi_A(u_1, \ldots, u_n)\varphi_B(u_1, \ldots, u_n)\right| \\
&\quad \le 2 \sum_{j\in A, k\in B} |u_j u_k| \left|\mathrm{Cov}\big(X_j^*, X_k^*\big)\right| \qquad (2.28)
\end{aligned}$$

and

$$\left|\varphi_{(X_1,\ldots,X_n)}(u_1, \ldots, u_n) - \prod_{j=1}^n \varphi_j(u_j)\right| \le 2 \sum_{\substack{j,k=1 \\ j\ne k}}^n |u_j u_k| \,\mathrm{Cov}\big(X_j^*, X_k^*\big). \qquad (2.29)$$

Proof To prove (2.28), define $f_1(u_1, \ldots, u_n) = \exp(i \sum_{j\in A} u_j X_j)$ and $f_2(u_1, \ldots, u_n) = \exp(i \sum_{k\in B} u_k X_k)$. As $\sum_{j\in A} u_j f_j \preceq \sum_{j\in A} |u_j| g_j$, Proposition 2.7 applies, so (2.28) is an immediate consequence of Lemma 2.8.

To prove (2.29), argue by induction as in the proof of Theorem 2.37, decomposing the set $\{1, \ldots, n\}$ in exactly the same way and using decomposition (2.27). As previously, the second and third terms of this decomposition are controlled directly from the induction hypothesis. To control the first term, define $U = f_1(u_1, \ldots, u_n)$ and $V = f_s(u_1, \ldots, u_n)$, which, according to Proposition 2.7, verify $U \preceq \sum_{j\in A} |u_j| g_j$ and $V \preceq \sum_{k\in B} |u_k| g_k$, and apply again Lemma 2.8. □

A straightforward modification of the proof of Lemma 2.33 gives an upper bound in terms of moment generating functions. In fact, we may write

$$\begin{aligned}
\mathrm{Cov}\big(e^{\lambda X}, e^{\lambda Y}\big) &= \int\int e^{\lambda(s+t)} (\mathbf{P}_{(X,Y)} - \mathbf{P}_X \otimes \mathbf{P}_Y)(ds\,dt) \\
&= \int\int \lambda^2 e^{\lambda(s+t)} H(s,t)\,ds\,dt.
\end{aligned}$$

The following statement is now obvious.

Lemma 2.41 *Let X and Y be associated random variables such that $|X|, |Y| \leq C$ for some constant $C > 0$. Then, for every $\lambda \in \mathbb{R}$,*

$$\left|\mathrm{E}e^{\lambda(X+Y)} - \mathrm{E}e^{\lambda X}\mathrm{E}e^{\lambda Y}\right| \leq \lambda^2 e^{2\lambda C}\,\mathrm{Cov}(X, Y). \tag{2.30}$$

Taking into account that, for any $A \subset \{1, \ldots, n\}$, $\mathrm{E}e^{\lambda \sum_{j\in A} X_j} \leq e^{\lambda |A| C}$, where $|A|$ is the number of elements in A, the extension for n random variables is immediate, following the arguments of the proof of Theorem 2.37.

Theorem 2.42 *Let X_n, $n \in \mathbb{N}$, be associated variables such that $|X_n| \leq C$ for some constant $C > 0$, not depending n. Then, for all $n \in \mathbb{N}$ and $\lambda \in \mathbb{R}$,*

$$\left|\mathrm{E}e^{\lambda \sum_{j=1}^n X_j} - \prod_{j=1}^n \mathrm{E}e^{\lambda X_j}\right| \leq \frac{\lambda^2}{2} e^{n\lambda C} \sum_{\substack{i,j=1 \\ i\neq j}}^n \mathrm{Cov}(X_i, X_j). \tag{2.31}$$

2.6 Exponential Inequalities

One of the main tools used for characterizing convergence rates in strong laws has been convenient versions of the so-called Bernstein-type exponential inequalities. There exist several versions of such inequalities available in the literature for independent sequences of variables with assumptions of uniform boundedness or some, quite relaxed, control on their (centred or noncentred) moments. For associated random variables, a first exponential inequality was proved by Prakasa Rao [83], but it was too weak to be really useful. In fact, this inequality does not even recover the known results if one assumes the variables to be independent. A stronger inequality, effectively extending results from the independent case, was proved by Ioannides and Roussas [48]. The technique was based on decomposing S_n into the sum of convenient blocks in both cases. But, while Prakasa Rao [83] tried to control everything just using properties of the exponential function, Ioannides and Roussas [48] controlled the mathematical expectations by coupling the blocks by independent ones. The route of the proof consists then in controlling the distance between the original blocks and the coupling independent variables, achieved using an induction argument, and finding convenient bounds for the independent coupling terms. Their inequality was later extended in Oliveira [75], where the approximation to independence was controlled in a different way, based on Theorem 2.42, avoiding the induction argument and dropping some technical assumptions appearing in course of the proof proposed by Ioannides and Roussas [48]. One way to have some insight on how optimized the exponential inequality is was to use it to characterize convergence rates for Strong Laws of Large Numbers to see how close the optimal rates for independent variables remain. This will be explored later in Sect. 3.2. We should note at this point that the main term in such convergence rate characterizations is the one that controls the independent coupling terms. This led to some effort in improving the control of these independent like terms by Sun [98], Xing, Yang and Liu [106], Henriques and Oliveira [43] and Xing and Yang [105], but we defer this to Sect. 3.2. Nevertheless, in all these references, the way association was used to find the exponential inequality was essentially the same, and that is what we will be concentrating on this section.

Let us now introduce the notation to be used throughout this section. Let c_n, $n \in \mathbb{N}$, be a sequence of nonnegative real numbers such that $c_n \longrightarrow +\infty$ and, given the random variables X_n, $n \in \mathbb{N}$, define, for all $i, n \geq 1$,

$$\begin{aligned} &X_{1,i,n} = -c_n \mathbb{I}_{(-\infty,-c_n)}(X_i) + X_i \mathbb{I}_{[-c_n,c_n]}(X_i) + c_n \mathbb{I}_{(c_n,+\infty)}(X_i), \\ &X_{2,i,n} = (X_i - c_n)\mathbb{I}_{(c_n,+\infty)}(X_i), \qquad X_{3,i,n} = (X_i + c_n)\mathbb{I}_{(-\infty,-c_n)}(X_i). \end{aligned} \tag{2.32}$$

For each fixed $n \geq 1$, the variables $X_{1,1,n}, \ldots, X_{1,n,n}$ are uniformly bounded, and thus they may be treated using Theorem 2.42. Note that, for each fixed $n \geq 1$, all these variables are monotone transformations of the initial variables X_n. This implies that an association assumption is preserved by this construction.

The proof of an exponential inequality will use, besides the truncation introduced before, a convenient decomposition of the sums into blocks. This block decomposition is a means to an approximation to independence technique on the truncated variables. The tails will be treated directly using Laplace transforms.

Consider a sequence of natural numbers p_n such that, for each $n \geq 1$, $p_n < \frac{n}{2}$ and define r_n as the greatest integer less than or equal to $\frac{n}{2p_n}$. Define then, for $q = 1, 2, 3$ and $j = 1, \ldots, 2r_n$,

$$Y_{q,j,n} = \sum_{\ell=(j-1)p_n+1}^{jp_n} (X_{q,\ell,n} - \mathrm{E}X_{q,\ell,n}). \tag{2.33}$$

Finally, for all $q = 1, 2, 3$ and $n \geq 1$, define

$$\begin{aligned} Z_{q,n,od} &= \sum_{j=1}^{r_n} Y_{q,2j-1,n}, \qquad Z_{q,n,ev} = \sum_{j=1}^{r_n} Y_{q,2j,n}, \\ R_{q,n} &= \sum_{\ell=2r_n p_n+1}^{n} (X_{q,\ell,n} - \mathrm{E}X_{q,\ell,n}). \end{aligned} \tag{2.34}$$

The proof of the main result is now divided into the control of the bounded terms, corresponding to the index $q = 1$, and the control of the nonbounded terms that correspond to the indices $q = 2, 3$. The next result takes care of the approximation between the joint distribution of the blocks and what one would find if these blocks were independent.

Lemma 2.43 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables. Then, for every $\lambda > 0$,*

$$\left| \mathrm{E}e^{\lambda Z_{1,n,od}} - \prod_{j=1}^{r_n} \mathrm{E}e^{\lambda Y_{1,2j-1,n}} \right| \leq \frac{\lambda^2 n}{2} e^{\lambda n c_n} \sum_{j=p_n+2}^{(2r_n-1)p_n} \mathrm{Cov}(X_1, X_j), \tag{2.35}$$

and analogously for the term corresponding to $Z_{1,n,ev}$.

Proof Taking into account (2.34) and the fact that the variables defined in (2.32) are associated, we have, applying directly Theorem 2.42,

$$\left| \mathrm{E} e^{\lambda Z_{1,n,od}} - \prod_{j=1}^{r_n} \mathrm{E} e^{\lambda Y_{1,2j-1,n}} \right| \\ \leq \lambda^2 r_n p_n e^{2\lambda r_n p_n c_n} \sum_{1 \leq j < j' \leq r_n} \mathrm{Cov}(Y_{1,2j-1,n}, Y_{1,2j'-1,n}). \tag{2.36}$$

As $2r_n p_n \leq n$, we are left with the sum of the covariances to deal with. Using the stationarity of the variables, it follows that

$$\sum_{1 \leq j < j' \leq r_n} \mathrm{Cov}(Y_{1,2j-1,n}, Y_{1,2j'-1,n}) = \sum_{j=1}^{r_n - 1} (r_n - j) \, \mathrm{Cov}(Y_{1,1,n}, Y_{1,2j-1,n}).$$

A further invocation of the stationarity implies that

$$\begin{aligned} &\mathrm{Cov}(Y_{1,1,n}, Y_{1,2j-1,n}) \\ &\quad = \sum_{\ell=0}^{p_n - 1} (p_n - \ell) \, \mathrm{Cov}(X_{1,1,n}, X_{1,2jp_n+\ell+1,n}) \\ &\qquad + \sum_{\ell=1}^{p_n - 1} (p_n - \ell) \, \mathrm{Cov}(X_{1,\ell+1,n}, X_{1,2jp_n+1,n}) \\ &\quad \leq p_n \sum_{\ell=(2j-1)p_n+2}^{(2j+1)p_n} \mathrm{Cov}(X_{1,1,n}, X_{1,\ell,n}). \end{aligned} \tag{2.37}$$

We now analyse the covariances using the Hoeffding formula (1.2):

$$\begin{aligned} \mathrm{Cov}(X_{1,i,n}, X_{1,j,n}) = \int_{\mathbb{R}^2} &\mathbf{P}(X_{1,i,n} > u, X_{1,j,n} > v) \\ &- \mathbf{P}(X_{1,i,n} > u) \mathbf{P}(X_{1,j,n} > v) \, du \, dv. \end{aligned} \tag{2.38}$$

If we take into account the truncation made in (2.32), it follows that the integrand function vanishes outside the square $[-c_n, c_n]^2$. Moreover, for $u, v \in [-c_n, c_n]$, we may replace, in the integrand, the variables $X_{1,i,n}$ and $X_{1,j,n}$ by X_i and X_j, respectively, so that

$$\begin{aligned} &\mathrm{Cov}(X_{1,i,n}, X_{1,j,n}) \\ &\quad = \int_{[-c_n, c_n]^2} \mathbf{P}(X_i > u, X_j > v) - \mathbf{P}(X_i > u) \mathbf{P}(X_j > v) \, du \, dv \\ &\quad \leq \int_{\mathbb{R}^2} \mathbf{P}(X_i > u, X_j > v) - \mathbf{P}(X_i > u) \mathbf{P}(X_j > v) \, du \, dv = \mathrm{Cov}(X_i, X_j), \end{aligned}$$

due to the nonnegativity of the latter integrand function, as follows from the association of the original variables. Inserting this into (2.36) and (2.37), the lemma follows. $\square$

The next step is to find some convenient control on the variables that couple the blocks $Y_{1,2j-1,n}$. We first prove a small extension of the moment inequalities studied in Sect. 2.3, that is better suited for our present purposes.

Lemma 2.44 *Let $c > 0$ and $S_{1,n} = \sum_{i=1}^{n}(X_{1,i,n} - \mathrm{E}X_{1,i,n})$. Assume that the random variables X_n, $n \in \mathbb{N}$, are strictly stationary, associated and $u(0) < \infty$. Then $\mathrm{E}S_{1,n}^2 \leq 2nc_n^*$, where $c_n^* \geq c_n^2 + u(0)$.*

Proof Using the stationarity, we easily get that

$$\begin{aligned}\mathrm{E}S_{1,n}^2 &= n\,\mathrm{Var}(X_{1,1,n}) + 2\sum_{j=1}^{n-1}(n-j)\,\mathrm{Cov}(X_{1,1,n}, X_{1,j+1,n})\\ &\leq 2nc_n^2 + 2nu(0) \leq 2nc_n^*,\end{aligned}$$

since $\mathrm{Cov}(X_{1,1,n}, X_{1,j+1,n}) \leq \mathrm{Cov}(X_1, X_{j+1})$ due to the association of the random variables, as mentioned in the proof of the previous lemma. $\square$

Remark 2.45 Notice that we can assume that $c_n^* = 2c_n^2$, at least as $c_n \longrightarrow +\infty$, as is the case for our framework.

The previous inequality will help improving the control of the independent-like terms used in Ioannides and Roussas [48] and Oliveira [75]. An improvement based on the Hölder inequality appears in Sung [98], but the approach by Xing, Yan and Liu [106] that goes along the arguments to be used below produces a better upper bound.

Lemma 2.46 *Let $X_n, n \in \mathbb{N}$, be strictly stationary and associated random variables such that $u(0) < \infty$. If $0 < \lambda < \frac{1}{2c_n p_n}$, then*

$$\prod_{j=1}^{r_n} \mathrm{E}e^{\lambda Y_{1,2j-1,n}} \leq \exp\bigl(\lambda^2 nc_n^*\bigr),$$

and the same bound holds for $\prod_{j=1}^{r_n} \mathrm{E}e^{\lambda Y_{1,2j,n}}$.

Proof From the definition (2.33) it is obvious that $|Y_{1,2j-1,n}| \leq 2c_n p_n$. Using a Taylor expansion and Lemma 2.44, we get that, for each $j = 1, \ldots, r_n$,

$$\mathrm{E}e^{\lambda Y_{1,2j-1,n}} \leq 1 + \lambda^2 \mathrm{E}Y_{1,2j-1,n}^2 \sum_{k=2}^{\infty} \frac{(2c_n\lambda p_n)^{k-2}}{k!} \leq \exp\bigl(2\lambda^2 p_n c_n^*\bigr),$$

using the inequality $1 + x \leq e^x$ for ≥ 0 and taking into account the assumption on λ. Finally, recall that $2r_n p_n \leq n$. $\square$

We may now prove an exponential inequality for the sum of odd indexed or even indexed terms.

Lemma 2.47 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables. Assume that $p_n > c_n > u(0)$ and*

$$\frac{n}{c_n^4}\exp\left(\frac{n}{4c_n}\right)u(p_n) \le C_0 < \infty. \tag{2.39}$$

Then, for every $\varepsilon \in (0, \frac{c_n}{p_n})$,

$$\mathbf{P}\left(\frac{1}{n}|Z_{1,n,od}| > \varepsilon\right) \le (1+32C_0)\exp\left(-\frac{n\varepsilon^2}{8c_n^2}\right), \tag{2.40}$$

and analogously for $Z_{1,n,ev}$.

Proof Applying Markov's inequality and using Lemma 2.43, we find that, for every $\lambda > 0$ small enough,

$$\mathbf{P}\left(\frac{1}{n}|Z_{1,n,od}| > \varepsilon\right) \le \frac{\lambda^2 n}{2}\exp(\lambda n c_n - \lambda n\varepsilon)\sum_{j=p_n+2}^{(2r_n-1)p_n}\mathrm{Cov}(X_1, X_l) + \exp\big(\lambda^2 n c_n^* - \lambda n\varepsilon\big). \tag{2.41}$$

To optimize the exponent in the last term of the upper bound in (2.41), we choose $\lambda = \frac{\varepsilon}{2c_n^*}$, so that $\lambda^2 n c_n^* - \lambda n\varepsilon = -\frac{n\varepsilon^2}{4c_n^*}$. Notice that as $\varepsilon < \frac{c_n}{p_n}$, it follows that the requirement on λ of Lemma 2.46 is fulfilled. Replacing now λ in the first term of the upper bound and taking into account (2.39), we get that

$$\begin{aligned}
&\mathbf{P}\left(\frac{1}{n}|Z_{1,n,od}| > \varepsilon\right)\\
&\le \frac{\varepsilon^2 n}{8(c_n^*)^2}\exp\left(\frac{nc_n}{2c_n^*} - \frac{n\varepsilon^2}{2c_n^*}\right)\sum_{j=p_n+2}^{(2r_n-1)p_n}\mathrm{Cov}(X_1, X_l) + \exp\left(-\frac{n\varepsilon^2}{4c_n^*}\right)\\
&\le 32C_0\exp\left(-\frac{n\varepsilon^2}{2c_n^*}\right) + \exp\left(-\frac{n\varepsilon^2}{4c_n^*}\right)\\
&\le (1+32C_0)\exp\left(-\frac{n\varepsilon^2}{4c_n^*}\right)\\
&= (1+32C_0)\exp\left(-\frac{n\varepsilon^2}{8c_n^2}\right).
\end{aligned}$$

□

Remark 2.48 Note that assumption (2.39), which involves the covariance structure on the previous lemma, is much stronger than (2.16), used to control second-order moments of maxima.

To complete the treatment of the bounded terms, it remains to control the sum corresponding to the indices after $2r_n p_n$, that is, $R_{1,n}$.

Lemma 2.49 *Let X_n, $n \in \mathbb{N}$, be strictly stationary associated random variables and assume that*

$$\frac{n}{c_n p_n} \longrightarrow +\infty. \tag{2.42}$$

Then, for n large enough and every $\varepsilon > 0$, we have $\mathbf{P}(|R_{1,n}| > n\varepsilon) = 0$.

Proof Recall the definition of $R_{1,n} = \sum_{\ell=2r_n p_n+1}^{n}(X_{1,\ell,n} - \mathrm{E}X_{1,\ell,n})$. Taking into account the construction of r_n and p_n, we get that $|R_{1,n}| \le 2(n-2r_n p_n)c_n \le 4c_n p_n$. Now $\mathbf{P}(|R_{1,n}| > n\varepsilon) \le \mathbf{P}(\frac{4}{\varepsilon} > \frac{n}{c_n p_n})$ and, by (2.42), this is equal zero for n large enough. □

The variables $X_{2,i,n}$ and $X_{3,i,n}$ are associated but not bounded, even for fixed n. This means that Theorem 2.42 may not be applied to the sum of such terms. But, we may note that these variables depend only on the tails of the distributions of the original variables. So, by controlling the decrease rate of these tails we may prove an exponential inequality for sums of $X_{2,i,n}$ or $X_{3,i,n}$. A first upper bound using such an approach was obtained in Oliveira [75], where the association was not explored, and later improved by Xing, Yang and Liu [106], using explicitly the association of the random variables via the maximal inequality proved in Theorem 2.27.

Lemma 2.50 *Let X_n, $n \in \mathbb{N}$, be associated random variables such that (2.18) holds and there exist $M > 0$ and $\delta > 0$ such that*

$$\sup_{|t|\le\delta} \mathrm{E}e^{tX_1} \le M < +\infty. \tag{2.43}$$

Then, for $t \in (0, \delta]$ and $q = 2, 3$,

$$\mathbf{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(X_{q,i,n} - \mathrm{E}X_{q,i,n})\right| > n\varepsilon\right) \le \frac{2MCe^{-tc_n}}{nt^2\varepsilon^2} + \frac{(2M)^{1/2}Ce^{-tc_n/2}}{nt\varepsilon^2}, \tag{2.44}$$

where C is the constant introduced in Theorem 2.27.

Proof According to Theorem 2.27, using Markov's inequality, we have

$$\begin{aligned}
&\mathbf{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(X_{q,i,n} - \mathrm{E}X_{q,i,n})\right| > n\varepsilon\right) \\
&\quad\le \frac{1}{n^2\varepsilon^2}\mathrm{E}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(X_{q,i,n} - \mathrm{E}X_{q,i,n})\right|^2\right) \\
&\quad\le \frac{C}{n\varepsilon^2}\left(\sup_i \mathrm{E}X_{q,i,n}^2 + \left(\sup_i \mathrm{E}X_{q,i,n}^2\right)^{1/2}\right),
\end{aligned} \tag{2.45}$$

so we need to find upper bounds for $\mathrm{E}X_{q,i,n}^2$. Write $\overline{F}_i(x) = \mathbf{P}(X_i > x)$. Using again Markov's inequality, we get that, for $t \in (0, \delta)$, $\overline{F}_i(x) \le e^{-tx}\mathrm{E}e^{tX_i} \le Me^{-tx}$.

Writing the mathematical expectation as a Stieltjes integral and integrating by parts, we find

$$\mathrm{E}X_{2,i,n}^2 = -\int_{(c_n,+\infty)} (x-c_n)^2 \overline{F}_i(dx) = \int_{c_n}^{+\infty} 2(x-c_n)\overline{F}_i(x)\,dx \le 2M\frac{e^{-tc_n}}{t^2}.$$

Inserting this into (2.45), we get the result. □

Remark 2.51 Notice that in (2.44), if $c_n \longrightarrow +\infty$, the second term in (2.44) is the dominating one.

We can now collect all the previous upper bounds into the following result, stating an exponential inequality regardless of the boundedness of the variables.

Theorem 2.52 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables such that (2.18), (2.39), (2.42) and (2.43) hold. Then, for n large enough, $\varepsilon \in (0, \frac{c_n}{p_n})$ and $t \in (0, \delta]$,*

$$\mathbf{P}\left(\left|\sum_{i=1}^n X_i - \mathrm{E}X_i\right| > n\varepsilon\right) \le 2(1+32C_0)\exp\left(-\frac{n\varepsilon^2}{648c_n^2}\right) + \frac{36MCe^{tc_n}}{nt^2\varepsilon^2} + \frac{18(2M)^{1/2}Ce^{tc_n/2}}{nt\varepsilon^2}. \quad (2.46)$$

Proof Just write

$$\begin{aligned}
&\mathbf{P}\left(\left|\sum_{i=1}^n X_i - \mathrm{E}X_i\right| > n\varepsilon\right)\\
&\le \sum_{q=1}^3 \mathbf{P}\left(\left|\sum_{i=1}^n X_{q,i,n} - \mathrm{E}X_{q,i,n}\right| > \frac{n\varepsilon}{3}\right)\\
&\le \mathbf{P}\left(|Z_{1,n,od}| > \frac{n\varepsilon}{9}\right) + \mathbf{P}\left(|Z_{1,n,ev}| > \frac{n\varepsilon}{9}\right) + \mathbf{P}\left(|R_{1,n}| > \frac{n\varepsilon}{9}\right)\\
&\quad + \mathbf{P}\left(\left|\sum_{i=1}^n X_{2,i,n} - \mathrm{E}X_{2,i,n}\right| > \frac{n\varepsilon}{3}\right) + \mathbf{P}\left(\left|\sum_{i=1}^n X_{3,i,n} - \mathrm{E}X_{3,i,n}\right| > \frac{n\varepsilon}{3}\right)
\end{aligned}$$

and use (2.40) and (2.44) to conclude. □

These inequalities are not yet in an adequate form for characterizing convergence rates. This will be done in Sect. 3.2, essentially allowing ε to depend on n and identifying a convenient decrease rate such that the derived upper bound still defines a convergent series.

Chapter 3
Almost Sure Convergence

Abstract This chapter studies essentially Strong Laws of Large Numbers (SLLN) for associated variables and their applications to the characterization of asymptotics of statistical estimators under associated sampling. It is possible to prove SLLN under fairly general assumptions, but, in order to prove characterizations of convergence rates, a closer care on the control of the covariances, based on the inequalities studied in the previous chapter, is required. Sect. 3.2 handles this kind of results, proving almost optimal convergence rates, that is, convergence rates arbitrarily close to those for independent variables. There exist characterizations of convergence rates based on extensions of the Law of Iterated Logarithm to associated variables. Such results are deferred to Chap. 4, as their proofs require a few inequalities to be proved there. We include a section on large deviations, a not yet very explored theme under association. Here the assumptions on the decay rate of the covariances are much stronger, a behaviour as found for some other dependence structures. The approach and techniques used in this chapter are adapted in the final section to prove almost sure consistency results for nonparametric density and regression estimators based on associated samples.

3.1 Introduction

The first Strong Law of Large Numbers (SLLN) for associated random variables appears in Newman [70] under strict stationarity and a Cesàro convergence assumption on the covariance structure of the variables. The stationarity assumption was dropped in Birkel [15], who proved a version of the SLLN that can be interpreted as a generalized Kolmogorov's SLLN. This subject remained more or less quite until the contribution by Ioannides and Roussas [48] provided a means to prove convergence rates for the asymptotic behaviour of nonparametric estimators for the density or for the regression. In fact, these authors proved an exponential-type inequality, a particular case of Theorem 2.52, that enabled them to find rates assuming a convenient geometric decay rate for the covariances. The rates proved by Ioannides and Roussas [48] were rather slow, indicating that the exponential inequality used there was far from an optimal form, so the interest on optimizing these inequalities and, consequently, the rates derived from them was natural. The approach introduced by Ioannides and Roussas [48] was improved by Oliveira [75], Sung [98], Xing,

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8_3,

Yang and Liu [106] and Xing and Yang [105] so that the rates derived became arbitrarily close to the rates for independent variables. Another direction was pursued by Louhichi [62] and Louhichi and Soulier [64], who proved the Marcinkiewicz–Zygmund SLLN with less requirements on the existence of moments. The covariance structure plays a rather indirect role on these later type of results. More recent contributions were made by Yang, Su and Yu [108] using an approach based on maximal inequalities and a suitable moment control. Their results require a mild assumption on the existence of moments and a very weak condition on the decay rate for the covariance structure, which now plays again an important role. A suitable use of maximal inequalities enables still a small improvement.

As already remarked, these results on almost sure convergence naturally had an interest for the statistical literature based on associated samples. The main interest has been on density estimation. Roussas [85, 86] appears to be the first to prove some asymptotics for kernel density estimators both almost surely and in distribution, these to be treated here in a later chapter. These results were followed by the contributions by Cai and Roussas [23, 24], who also considered the estimation of distribution functions. The results obtained included some convergence rates characterization but assumed the sample to have pairwise joint distributions with densities that should be close to product densities, that is, assumptions similar to those appearing when considering strong mixing samples. A few interesting examples are left out by this absolute continuity assumption, as for instance, the model described in Example 1.28. Masry [65] proved various consistency results, including optimal almost sure convergence rates, for the kernel density estimator, using an approach based on moment inequalities and being less restrictive on the decay rate of the covariances, as it includes polynomially decreasing covariances. The pairwise joint distributions were still assumed to be absolutely continuous. The approach based on exponential inequalities requires geometrically decreasing covariances but, although first results were somewhat less encouraging, provides better convergence rates for the estimator. This approach was used in Oliveira [74] and Henriques and Oliveira [42] for the density estimator. The same method was also used in Azevedo and Oliveira [3] and Henriques and Oliveira [41] to deal with the estimation of distribution functions. A more general approach to estimation problems, trying to treat density and regression estimation, deals with the estimation of Radon–Nikodym derivatives in point process models. A few efforts dealing with associated samples were made by Ferrieux [38, 39] and Jacob and Oliveira [52]. As follows from Sect. 1.4 (see, for instance, Example 1.4), this is a different framework from considering associated samples of random variables, as the base space is changed in a way that the order structure is affected. Nevertheless, this abstract approach has consequences on results about estimation, enabling the treatment of pairwise distributions that are no longer absolutely continuous, as showed in Oliveira [74], where pairwise joint distributions are allowed to have some mass concentrated on the diagonal of the product space.

3.2 Strong Laws of Large Numbers

The first extensions of the SLLN to associated random variables follow the same approach as the classical one for independent variables. Thus we start by a result requiring the existence of second-order moments, leaving the identification of convergence rates for a later approach.

Theorem 3.1 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Assume that*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Cov}(X_n, S_n) < \infty. \tag{3.1}$$

Then $\frac{1}{n} S_n \longrightarrow 0$ almost surely.

Proof We will prove this result by arguments similar to the classic ones: prove the convergence along a suitably chosen subsequence and afterwards control the remaining terms.

Step 1. From Chebyshev's inequality we have that $\mathbf{P}(|S_{2^n}| > 2^n \varepsilon) \leq \frac{1}{4^n} \operatorname{Var}(S_{2^n})$. An elementary calculation gives that, as the covariances are nonnegative,

$$\operatorname{Var}(S_{2^n}) = \sum_{i,j=1}^{2^n} \operatorname{Cov}(X_i, X_j) \leq 2 \sum_{i=1}^{2^n} \operatorname{Cov}(X_i, S_i).$$

Thus, reversing the order of summation, again taking into account the nonnegativity of the covariances, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P}\big(|S_{2^n}| > \varepsilon 2^n\big) \\
&\quad \leq \sum_{n=1}^{\infty} \frac{1}{4^n \varepsilon^2} \operatorname{Var}(S_{2^n}) \\
&\quad \leq \frac{2}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{4^n} \sum_{i=1}^{2^n} \operatorname{Cov}(X_i, S_i) \leq \frac{2}{\varepsilon^2} \sum_{i=1}^{\infty} \Bigg(\sum_{n: 2^n \geq i} \frac{1}{4^n} \Bigg) \operatorname{Cov}(X_i, S_i) \\
&\quad \leq \frac{2}{\varepsilon^2} \sum_{i=1}^{\infty} \Bigg(\sum_{n \geq \log_2 i} \frac{1}{4^n} \Bigg) \operatorname{Cov}(X_i, S_i) \leq \frac{8}{3\varepsilon^2} \sum_{i=1}^{\infty} \frac{1}{i^2} \operatorname{Cov}(X_i, S_i) < \infty.
\end{aligned} \tag{3.2}$$

Applying now the Borel–Cantelli lemma, it follows that $\frac{1}{2^n} S_{2^n} \longrightarrow 0$ almost surely.

Step 2. To control the remaining terms, first notice that if $2^n < k \leq 2^{n+1}$, then

$$\left| \frac{S_k}{k} - \frac{S_{2^n}}{2^n} \right| \leq \frac{|S_k - S_{2^n}|}{2^n} + \frac{|S_{2^n}|}{2^n},$$

so the theorem follows from the convergence proved in Step 1 if we also prove that $2^{-n} \max_{2^n < k \leq 2^{n+1}} |S_k - S_{2^n}| \longrightarrow 0$ almost surely. For this, we use again the Borel–

Cantelli lemma: from Chebyshev's inequality and using Theorem 2.22, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P}\Big(\max_{2^n<k\leq 2^{n+1}} |S_k - S_{2^n}| > \varepsilon 2^n\Big) \\
&\quad \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2 4^n} \mathrm{E}\Big(\max_{2^n<k\leq 2^{n+1}} |S_k - S_{2^n}|\Big)^2 \\
&\quad \leq \sum_{n=1}^{\infty} \frac{2}{\varepsilon^2 4^n} \operatorname{Var}(S_{2^{n+1}} - S_{2^n}) \leq \sum_{n=1}^{\infty} \frac{2}{\varepsilon^2 4^n} \operatorname{Var}(S_{2^{n+1}}),
\end{aligned}$$

due to the association of the random variables. Finally, repeating the arguments in (3.2), the series above is convergent, so the theorem is proved. □

If we assume the variables to be stationary, we may replace (3.1) by an easier verifiable assumption.

Corollary 3.2 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable, stationary and associated variables. Assume that, for some $\alpha > 0$,*

$$a_n = \frac{1}{n} \sum_{j=1}^{n} \operatorname{Cov}(X_1, X_j) = O\big(\log^{-\alpha} n\big). \tag{3.3}$$

Then $\frac{1}{n} S_n \longrightarrow 0$ almost surely.

Proof It is easily seen that $\sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Cov}(X_n, S_n) = \sum_{n=1}^{\infty} \frac{a_n}{n}$, so (3.1) immediately follows from (3.3). □

Remark 3.3 A result similar to Corollary 3.2 is contained in Theorem 7 in Newman [70]. In this case only $a_n \longrightarrow 0$ is assumed, but it is required that the variables are strictly stationary, while in the previous result the covariance stationarity is enough.

We will next try to relax the existence of second-order moments and, simultaneously, identify convergence rates for the convergence. The approach follows a somewhat different path, looking at maxima of the partial sums instead of the partial sums themselves, avoiding the use of covariances. This method was first used by Louhichi [62] and Louhichi and Soulier [64], who proved the Marcinkiewicz–Zygmund Strong Law of Large Numbers, that is, the almost sure convergence to zero of $\frac{1}{n^{1/p}} S_n$ for $p \in [1, 2)$, thus giving some information about the rate of convergence in the strong law. This was proved under an assumption similar to (3.1), where the decrease rate of the covariances, or more precisely, of $u(n) = \sup_{k\in\mathbb{N}} \sum_{j:|j-k|\geq n} \operatorname{Cov}(X_j, X_k)$ is only characterized indirectly. Assuming a more precise behaviour, but still a rather mild restriction, on $u(n)$, the convergence rate for the Strong Law of Large Numbers has been improved by Yang, Su and Yu [108] to an almost optimal one, in the sense that it is arbitrarily close to the known rate for independent random variables, as characterized by the Law of Iterated Logarithm. Let us start by a general result concerning the almost sure convergence. Recall that we denote $M_n = \max(S_1, \ldots, S_n)$ and $M_n^* = \max(|S_1|, \ldots, |S_n|)$.

Theorem 3.4 *Let X_n, $n \in \mathbb{N}$, be centred random variables, and b_n, $n \in \mathbb{N}$, a nondecreasing sequence of real numbers such that $1 \leq \frac{b_{2n}}{b_n} \leq c \in \mathbb{R}$. Assume that, for every $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} \frac{1}{n} \mathbf{P}\big(M_n^* > \varepsilon b_n\big) < \infty. \tag{3.4}$$

Then $\frac{M_n^*}{b_n} \longrightarrow 0$ almost surely.

Proof We verify that we can apply the Borel–Cantelli lemma, first along a suitable subsequence and then control the remaining terms, thus following an approach similar to the argument used to prove Theorem 3.1. We start by considering the subsequence along the indices defined by powers of 2. For this step, we find an upper bound for

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}\big(M_{2^n}^* > \varepsilon b_{2^n}\big) &= \sum_{n=1}^{\infty} \sum_{k=2^n}^{2^{n+1}-1} \frac{1}{2^n} \mathbf{P}\big(M_{2^n}^* > \varepsilon b_{2^n}\big) \\
&\leq \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{1}{k} \mathbf{P}\big(M_{2k}^* > \varepsilon b_k\big) \\
&\quad + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{2}{2k+1} \mathbf{P}\big(M_{2k+1}^* > \varepsilon b_{k+1}\big),
\end{aligned}$$

taking into account, for the last inequality, that b_n is nondecreasing. As $\frac{b_{2n}}{b_n} \leq c$, we still have that

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P}\big(M_{2^n}^* > \varepsilon b_{2^n}\big) \\
&\quad \leq \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{1}{k} \mathbf{P}\left(M_{2k}^* > \frac{\varepsilon b_{2k}}{c}\right) + \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{2}{2k+1} \mathbf{P}\left(M_{2k+1}^* > \frac{\varepsilon b_{2k+1}}{c}\right) \\
&\quad \leq 2 \sum_{n=1}^{\infty} \sum_{k=2^n}^{2^{n+1}-1} \frac{1}{n} \mathbf{P}\left(M_n^* > \frac{\varepsilon b_n}{c}\right) \\
&\quad = \sum_{n=1}^{\infty} \frac{1}{n} \mathbf{P}\left(M_n^* > \frac{\varepsilon b_n}{c}\right) < \infty.
\end{aligned}$$

Thus, it follows that $\frac{1}{b_{2^n}} M^*_{2^n} \longrightarrow 0$ almost surely. As regards the remaining terms in the sequence, recalling again that b_n is nondecreasing and $\frac{b_{2n}}{b_n} \leq c$,

$$\max_{2^{n-1}<k\leq 2^n} \frac{M^*_k}{b_k} \leq \frac{M^*_{2^n}}{b_{2^{n-1}}} \leq \frac{M^*_{2^n}}{cb_{2^n}} \longrightarrow 0,$$

as proved in the first part of this proof. □

In order to prove the more general Marcinkiewicz–Zygmund Strong Law of Large Numbers, we need some preparatory results. For this, recall that $H_{j,k}(x,y) = \mathbf{P}(X_j > x, X_k > y) - \mathbf{P}(X_j > x)\mathbf{P}(X_k > y)$, $g_v(u) = \max(\min(u,v), -v)$ and

$$G_{j,k}(v) = \mathrm{Cov}\big(g_v(X_j), g_v(X_k)\big) = \int\int_{[-v,v]^2} H_{j,k}(x,y)\,dx\,dy,$$

as introduced before. We need a few technical results before proving the actual version of the strong law, dealing with some integrals needed in course of proof. Let $p \in [1,2)$, $j,k \in \mathbb{N}$, and $s \geq 1$. Then, as

$$\max\big(|x|,|y|,s^{1/p}\big)^{-2} = \int_0^1 \mathbb{I}_{[-u^{-1/2},u^{-1/2}]}(x)\mathbb{I}_{[-u^{-1/2},u^{-1/2}]}(y)\mathbb{I}_{[u,\infty]}\big(s^{-2/p}\big)\,du,$$

we can use Fubini's theorem to obtain

$$\begin{aligned}\int\int &\max\big(|x|,|y|,s^{1/p}\big)^{-2} H_{j,k}(x,y)\,dx\,dy \\ &= \int_0^1 \mathbb{I}_{[u,\infty]}\big(s^{-2/p}\big)G_{j,k}\big(u^{-1/2}\big)\,du = 2\int_{s^{1/p}}^{\infty} \frac{1}{v^3} G_{j,k}(v)\,dv, \qquad (3.5)\end{aligned}$$

assuming, of course, that the integrals are finite. The next lemma shows that the boundedness of the pth moments is enough for the finiteness of the integrals.

Lemma 3.5 *Let X_n, $n \in \mathbb{N}$, be associated random variables such that, for some $p \in [1,2)$, $\sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^p < \infty$. Then, for all $j,k \in \mathbb{N}$,*

$$\int_1^{\infty} v^{-3} G_{j,k}(v)\,dv \leq \frac{2}{p^2} \sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^p.$$

Proof Denote by $\mathrm{sgn}(x)$ the sign of x. Then, taking into account (3.5) and, due to the association of the random variables, the nonnegativity of $H_{j,k}$, we have

$$\begin{aligned}2\int_1^{\infty} v^{-3} G_{j,k}(v)\,dv &\leq \int\int \min\big(|x|^{-2},|y|^{-2},1\big) H_{j,k}(x,y)\,dx\,dy \\ &\leq \int\int |x|^{-2}|y|^{-2} H_{j,k}(x,y)\,dx\,dy.\end{aligned}$$

Using Theorem C.4, we have that

$$\begin{aligned}2\int_1^{\infty} v^{-3} G_{j,k}(v)\,dv &\leq \frac{4}{p^2}\mathrm{Cov}\big(\mathrm{sgn}(X_j)|X_j|, \mathrm{sgn}(X_k)|X_k|\big) \\ &\leq \frac{4}{p^2} \sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^p.\end{aligned}$$

□

Remark 3.6 The result above is better than what could be derived using the bound from Corollary 2.36. In fact, from (2.25) we find the uniform bound $H_{j,k}(x,y) \leq \frac{1}{M^*}\mathrm{Cov}^{1/3}(X,Y)$, which is not enough to derive the finiteness of the integral above.

Having obtained a convenient control on the $G_{j,k}$ functions, we can prove the Marcinkiewicz–Zygmund Strong Law of Large Numbers. The assumptions require a uniform domination of the sequence of random variables in terms of their distributions, but, differently from what happened in previous results, the covariances of the random variables do not appear explicitly. They are only indirectly considered through the functions $G_{j,k}$.

Theorem 3.7 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables such that there exists a random variable Y verifying $\sup_{n\in\mathbb{N}} \mathbf{P}(|X_n| > x) \leq \mathbf{P}(Y > x)$ for every $x > 0$ and $\mathrm{E}Y^p < \infty$ for some $p \in [1,2)$. If*

$$\sum_{\substack{j,k=1\\ j\neq k}}^{n} \int_{k^{1/p}}^{\infty} \frac{1}{v^3} G_{j,k}(v)\,dv < \infty, \tag{3.6}$$

then $\frac{M_n^*}{n^{1/p}} \longrightarrow 0$ almost surely.

Proof Taking into account Theorem 3.4, it is enough to prove that, for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \frac{1}{n} \mathbf{P}\big(M_n^* > \varepsilon n^{1/p}\big) < \infty.$$

The case $p \in (1,2)$. As the random variables are associated, we may apply Lemma 2.30 with $x = \varepsilon n^{1/p}$ and $m = n^{1/p}$ to find that

$$\begin{aligned}
&\frac{1}{n}\mathbf{P}\big(M_n^* > \varepsilon n^{1/p}\big)\\
&\quad\leq \frac{4}{n^{1/p}\varepsilon}\mathrm{E}(Y\mathbb{I}_{\{Y^p>n\}}) + \frac{4}{n^{2/p}\varepsilon^2}\mathrm{E}\big(Y^2\mathbb{I}_{\{Y^p\leq n\}}\big) + \frac{1}{\varepsilon^2}\mathbf{P}\big(Y > n^{1/p}\big)\\
&\qquad + \frac{8n^{-1-2/p}}{\varepsilon^2} \sum_{\substack{j,k=1\\ j\neq k}}^{n} G_{j,k}\big(n^{1/p}\big).
\end{aligned} \tag{3.7}$$

To bound the first three terms above, notice that

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^{1/p}}\mathrm{E}(Y\mathbb{I}_{\{Y^p>n\}}) &= \mathrm{E}\left(Y\sum_{n=1}^{\infty}\frac{1}{n^{1/p}}\mathbb{I}_{\{Y^p>n\}}\right)\\
&\leq \mathrm{E}\left(Y\int \frac{1}{x^{1/p}}\mathbb{I}_{\{Y^p>x\}}\,dx\right) \leq K\mathrm{E}Y^p,
\end{aligned}$$

using Fubini's theorem, for some $K > 0$. Notice that K above depends only on p. Proceeding in the same way, we find the bound

$$\sum_{n=1}^{\infty}\left(\frac{1}{n^{2/p}}\mathrm{E}\big(Y^2\mathbb{I}_{\{Y^p\leq n\}}\big)+\mathbf{P}\big(Y>n^{1/p}\big)\right)<K_1\mathrm{E}Y^p,$$

where again $K_1 > 0$ only depends on p. Finally,

$$\begin{aligned}
&\sum_{n=1}^{\infty}\frac{1}{n^{1+2/p}}\sum_{\substack{j,k=1\\ j\neq k}}^{n}G_{j,k}\big(n^{1/p}\big)\\
&\quad=\sum_{\substack{j,k=1\\ j\neq k}}^{n}\int\int\sum_{n=1}^{\infty}\frac{1}{n^{1+2/p}}\mathbb{I}_{\{\max(|x|^p,|y|^p,j)\leq n\}}H_{j,k}(x,y)\,dx\,dy\\
&\quad\leq K^*\sum_{\substack{j,k=1\\ j\neq k}}^{n}\int\int\frac{1}{\max(|x|^p,|y|^p,j)^2}H_{j,k}(x,y)\,dx\,dy\\
&\quad=2K^*\sum_{\substack{j,k=1\\ j\neq k}}^{n}\int_{j^{1/p}}^{\infty}\frac{1}{v^3}G_{j,k}(v)\,dv<\infty,
\end{aligned}$$

using (3.5) and taking into account (3.6). Thus, inserting all these inequalities into (3.7) yields the theorem for $p \in (1, 2)$.

The case $p = 1$. Let $\varepsilon > 0$ be fixed and define, for each $i, n \in \mathbb{N}$,

$$Y_{1,i,n}=g_n(X_i)=\max\big(\min(X_i,n),-n\big),\qquad Y_{2,i,n}=X_i-Y_{1,i,n}$$

and

$$T_n=\sup_{k\leq n}\sum_{i=1}^{k}(Y_{1,i,n}-\mathrm{E}Y_{1,i,n}).$$

Consider the set $\Gamma=\bigcup_{i=1}^{n}\{X_i\neq Y_{1,i,n}\}$. Then, obviously,

$$\mathbf{P}(\Gamma)\leq\sum_{i=1}^{n}\mathbf{P}\big(|X_i|>n\big)\leq\sum_{n=1}^{n}\mathbf{P}(Y>n).$$

On the other hand, if $\omega\notin\Gamma$, taking into account that $\mathrm{E}X_i = 0$, we may write $S_j = \sum_{i=1}^{j}(X_i-\mathrm{E}X_i)$, and it follows that $M_n^*(\omega)\leq T_n(\omega)+\sum_{i=1}^{n}\mathrm{E}|X_i-Y_{1,i,n}|$. But

$$\mathrm{E}|Y_{1,i,n}|\leq\mathrm{E}\big((|X_i|-n)\mathbb{I}_{\{|X_i|>n\}}\big)\leq\mathrm{E}\big((Y-n)\mathbb{I}_{\{Y>n\}}\big).$$

As Y is integrable for n large enough, we have that $\mathrm{E}|X_i-Y_{1,i,n}|\leq\frac{\varepsilon}{2}$. Hence, repeating the arguments used for the previous case, we have

$$\sum_{n=1}^{\infty}\frac{1}{n}\mathbf{P}\big(M_n^*>n\varepsilon\big)\leq\sum_{n=1}^{\infty}\mathbf{P}(Y>n)+\sum_{n=1}^{\infty}\frac{1}{n}\mathbf{P}\left(T_n>\frac{n\varepsilon}{2}\right)$$
$$\leq \mathrm{E}Y+K^*\mathrm{E}Y+\sum_{\substack{j,k=1\\ j\neq k}}^{n}\int_j^{\infty}\frac{1}{v^3}G_{j,k}(v)\,dv,$$

where K^* does not depend on n, thus concluding the proof. □

The next result gives a convergence rate of a general form.

Theorem 3.8 *Let X_n, $n\in\mathbb{N}$, be centred and associated random variables satisfying*

$$\sum_{i=1}^{\infty}u^{1/2}\big(2^i\big)<\infty.$$

Let φ be a nonnegative real function such that $\lim_{x\to+\infty}\varphi(x)=+\infty$, $\frac{\varphi(x)}{x}$ is monotonous, and $\frac{\varphi(x)}{x^2}$ is decreasing. Let b_n, $n\in\mathbb{N}$, be an increasing sequence of positive real numbers. Assume that the following conditions are satisfied:

$$0\leq\frac{b_{2n}}{b_n}\leq c<\infty, \tag{3.8}$$

$$\sum_{n=1}^{\infty}\frac{1}{b_n^2}<\infty, \tag{3.9}$$

$$\sum_{n=1}^{\infty}\frac{\mathrm{E}\varphi(|X_n|)}{\varphi(b_n)}<\infty, \tag{3.10}$$

$$\sum_{n=1}^{\infty}\frac{1}{b_n^2}\max_{k\leq n}b_j^2\mathbf{P}\big(|X_k|>b_k\big)<\infty, \tag{3.11}$$

$$\sum_{n=1}^{\infty}\frac{1}{b_n^2}\max_{k\leq n}\frac{b_j^2\mathrm{E}(\varphi(|X_k|)\mathbb{I}_{\{|X_k|\leq b_k\}})}{\varphi(b_k)}<\infty. \tag{3.12}$$

Then, $\frac{1}{b_n}S_n\longrightarrow 0$ almost surely.

Proof Define the truncated variables $Y_n=X_n\mathbb{I}_{\{|X_n|\leq b_n\}}-b_n\mathbb{I}_{\{X_n<-b_n\}}+b_n\mathbb{I}_{\{X_n>b_n\}}$, $n\in\mathbb{N}$. The variables Y_n, $n\in\mathbb{N}$, being monotonous transformations of the initial variables, are still associated. Now, if $\frac{\varphi(x)}{x}$ is nonincreasing, it follows that if $|X_n|\leq b_n$, then $\frac{|X_n|}{b_n}\leq\frac{\varphi(|X_n|)}{\varphi(b_n)}$, so, using Markov's inequality, we have

$$|\mathrm{E}Y_n|\leq\mathrm{E}|Y_n|\leq b_n\mathrm{E}\left(\frac{|X_n|}{b_n}\mathbb{I}_{\{|X_n|\leq b_n\}}\right)+b_n\mathbf{P}\big(|X_n|>b_n\big)\leq 2b_n\mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)}.$$

On the other hand, if $\frac{\varphi(x)}{x}$ is nondecreasing, it follows that if $|X_n|>b_n$, then $\frac{|X_n|}{b_n}\leq\frac{\varphi(|X_n|)}{\varphi(b_n)}$, so taking into account that $\mathrm{E}X_n=0$ and repeating the arguments of the previous case, we have

$$\begin{aligned}|\mathrm{E}Y_n| &= \left|\mathrm{E}(X_n \mathbb{I}_{\{|X_n|>b_n\}})\right| + b_n \mathbf{P}\big(|X_n|>b_n\big) \\ &\le b_n \mathrm{E}\left(\frac{|X_n|}{b_n}\mathbb{I}_{\{|X_n|>b_n\}}\right) + b_n \mathbf{P}\big(|X_n|>b_n\big) \\ &\le b_n \mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)} + b_n \mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)} \\ &= 2b_n \mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)}.\end{aligned}$$

That is, in either case we find that

$$\sum_n \frac{|\mathrm{E}Y_n|}{b_n} \le 2\sum_n \mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)}.$$

Thus, from (3.10) and using the Kronecker's lemma (see Lemma C.7), it follows that $\frac{1}{b_n}\sum_{j=1}^n |\mathrm{E}Y_j| \longrightarrow 0$ almost surely. Moreover, as

$$\sum_{n=1}^\infty \mathbf{P}(X_n \ne Y_n) = \sum_{n=1}^\infty \mathbf{P}\big(|X_n|>b_n\big) \le \sum_{n=1}^\infty \mathrm{E}\frac{\varphi(|X_n|)}{\varphi(b_n)} < \infty,$$

the theorem follows if we prove that $\frac{1}{b_n}\sum_{j=1}^n (Y_j - \mathrm{E}Y_j) \longrightarrow 0$ almost surely. Taking into account (3.8) and Theorem 3.4, it is enough to verify that, for each fixed $\varepsilon > 0$,

$$\sum_n \frac{1}{n}\mathbf{P}\left(\max_{k\le n}\left|\sum_{j=1}^k (Y_j - \mathrm{E}Y_j)\right| > \varepsilon b_n\right) < \infty. \tag{3.13}$$

Notice that if $|X_j| \le b_j$, as $\frac{\varphi(x)}{x^2}$ is decreasing, it follows that $\frac{X_j^2}{b_j^2} \le \frac{\varphi(|X_j|)}{\varphi(b_j)}$. Thus,

$$\begin{aligned}\mathrm{E}Y_j^2 &= \mathrm{E}\big(X_j^2 \mathbb{I}_{\{|X_j|\le b_j\}}\big) + b_j^2 \mathbf{P}\big(|X_j|>b_j\big) \\ &= b_j^2 \mathrm{E}\left(\frac{X_j^2}{b_j^2}\mathbb{I}_{\{|X_j|\le b_j\}}\right) + b_j^2 \mathbf{P}\big(|X_j|>b_j\big) \\ &\le b_j^2 \frac{\mathrm{E}(\varphi(|X_j|)\mathbb{I}_{\{|X_j|\le b_j\}})}{\varphi(b_j)} + b_j^2 \mathbf{P}\big(|X_j|>b_j\big).\end{aligned}$$

Then, as the random variables Y_n are associated, we use (2.17), to conclude that

$$\begin{aligned}&\sum_{n=1}^\infty \frac{1}{n}\mathbf{P}\left(\max_{k\le n}\left|\sum_{j=1}^k (Y_j - \mathrm{E}Y_j)\right| > \varepsilon b_n\right) \\ &\quad \le \frac{1}{\varepsilon^2}\sum_{n=1}^\infty \frac{1}{nb_n^2}\mathrm{E}\left(\max_{k\le n}\left|\sum_{j=1}^k (Y_j - \mathrm{E}Y_j)\right|^2\right)\end{aligned}$$

$$\leq C \sum_{n=1}^{\infty} \frac{1}{b_n^2} \Big(\max_{k \leq n} \mathrm{E} Y_k^2 + 1 \Big)$$

$$\leq C \sum_{n=1}^{\infty} \frac{1}{b_n^2} \left(\max_{k \leq n} b_k^2 \frac{\mathrm{E}(\varphi(|X_k|)\mathbb{I}_{\{|X_k| \leq b_k\}})}{\varphi(b_k)} + b_k^2 \mathbf{P}\big(|X_k| > b_k\big) + 1 \right) < \infty,$$

so it follows that $\frac{1}{b_n} \max_{k \leq n} S_k \longrightarrow 0$ almost surely, and hence the theorem is proved. □

Conditions (3.11) and (3.12) are difficult to verify, but we may replace them by convenient moment assumptions.

Corollary 3.9 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables satisfying*

$$\sum_{i=1}^{\infty} u^{1/2}\big(2^i\big) < \infty.$$

Let φ be a nonnegative real function such that $\lim_{x \to +\infty} \varphi(x) = +\infty$, $\frac{\varphi(x)}{x}$ is monotonous, and $\frac{\varphi(x)}{x^2}$ is decreasing. Let b_n, $n \in \mathbb{N}$, be an increasing sequence of positive real numbers. Assume that the following conditions are satisfied: (3.8), (3.9) *and*

$$\sum_{n=1}^{\infty} \frac{1}{\varphi(b_n)} < \infty, \tag{3.14}$$

$$\sup_{n \in \mathbb{N}} \mathrm{E}\varphi\big(|X_n|\big) < \infty. \tag{3.15}$$

Then, $\frac{1}{b_n} S_n \longrightarrow 0$ almost surely.

Proof It is enough to check that (3.10)–(3.12) are verified. Now, as

$$\frac{\mathrm{E}\varphi(|X_n|)}{\varphi(b_n)} \leq \frac{1}{\varphi(b_n)} \sup_{n \in \mathbb{N}} \mathrm{E}\varphi\big(|X_n|\big),$$

(3.10) immediately follows from (3.14). To verify (3.11), notice that $b_j^2 \mathbf{P}(|X_j| > b_j) \leq \frac{b_j^2}{\varphi(b_j)} \mathrm{E}\varphi(|X_j|) \leq \frac{b_j^2}{\varphi(b_j)} \sup_{n \in \mathbb{N}} \mathrm{E}\varphi(|X_n|)$, so, taking into account that $\frac{\varphi(x)}{x^2}$ is decreasing, we have that

$$\frac{1}{b_n^2} \max_{k \leq n} b_j^2 \mathbf{P}\big(|X_k| > b_k\big) \leq \frac{1}{\varphi(b_n)} \sup_{n \in \mathbb{N}} \mathrm{E}\varphi\big(|X_n|\big),$$

thus (3.11) follows from (3.14). Finally, (3.12) is verified in the same way. □

Choosing now a convenient function φ and a sequence b_n, we find a more explicit characterization of the convergence rate for the Strong Law of Large Numbers.

Corollary 3.10 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables such that* $\sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^p < \infty$ *for some* $p \geq 1$ *and*

$$\sum_{i=1}^{\infty} u^{1/2}(2^i) < \infty.$$

Then, for every $\delta > 1$,

$$\frac{n^{1-1/p}}{(\log n)^{1/p}(\log\log n)^{\delta/p}} \frac{1}{n} S_n \longrightarrow 0 \quad \textit{almost surely}.$$

Proof Choose, in Corollary 3.9, $\varphi(x) = |x|^p$ and $b_n = (n \log n (\log\log n)^{\delta})^{1/p}$. □

Remark 3.11 The most interesting use of the previous corollary corresponds, of course, to the case $p \in [1, 2]$. Notice further that, by choosing $p = 2$, the previous result identifies a convergence rate of order $n^{-1/2}(\log n)^{1/2}(\log\log n)^{\delta/2}$, thus close to the optimal one for independent variables, which is of order $n^{-1/2}(\log\log n)^{1/2}$.

It is possible to improve still the above-mentioned convergence rate at the cost of requiring the boundedness of higher-order moments. This can be achieved by a proper use of the maximal inequalities proved before (see Corollary 2.29) together with a suitable moment bound. We start by an extension of Theorem 3.8.

Theorem 3.12 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables, and φ be a nonnegative real function such that* $\lim_{x\to+\infty} \varphi(x) = +\infty$, $\frac{\varphi(x)}{x}$ *is monotonous, and* $\frac{\varphi(x)}{x^2}$ *is decreasing. Let b_n, $n \in \mathbb{N}$, be an increasing sequence of positive real numbers. Let $r > p \geq 2$ and assume that* $\sup_{n\in\mathbb{N}} \|X_n\|_r < \infty$ *and* $u(n) \leq Cn^{-\alpha}$, *where $C > 0$ is some constant, for some* $\alpha \geq \frac{r(p-2)}{2(r-p)}$, (3.8), (3.10) *are satisfied, and*

$$\sum_n \frac{n^{p/2-1}}{b_n^p} < \infty. \tag{3.16}$$

Then, $\frac{1}{b_n} S_n \longrightarrow 0$ *almost surely.*

Proof Follow the arguments in the proof of Theorem 3.8 until (3.13). An obvious upper bound is then obtained by using Markov's inequality, so that, taking into account Corollary 2.29, we have

$$\sum_{n=1}^{\infty} \frac{1}{n\varepsilon^p b_n^p} \mathrm{E}\left(\max_{k\leq n}\left|\sum_{j=1}^{k}(Y_j - \mathrm{E}Y_j)\right|^p\right) \leq c \sum_{n=1}^{\infty} \frac{1}{n b_n^p} \mathrm{E}\left(\sum_{j=1}^{n}(Y_j - \mathrm{E}Y_j)\right)^p,$$

where c does not depend on n. Thus, it is enough to prove the convergence of the series on the right. To control the pth-order moment, we apply Corollary 2.20, so we need to bound $u_Y(n)$ defined as in (2.13) but with respect to the Y_n variables. Now write, using Hoeffding's formula (1.2),

$$\mathrm{Cov}(Y_j, Y_k) = \int\int \mathbf{P}(Y_j > s, Y_k > t) - \mathbf{P}(Y_j > s)\mathbf{P}(Y_k > t)\, ds\, dt.$$

The integrand is easily checked to be null outside the rectangle $[-b_j, b_j] \times [-b_k, b_k]$ and equal to $\mathbf{P}(X_j > s, X_k > t) - \mathbf{P}(X_j > s)\mathbf{P}(X_k > t)$ inside this rectangle. So, due to the association of the random variables, it follows that $\mathrm{Cov}(Y_j, Y_k) \le \mathrm{Cov}(X_j, X_k)$. Thus, $u_Y(n) \le u(n) \le Cn^{-\alpha}$, and Corollary 2.20 is applicable to the Y_n variables to find that

$$\left| \mathrm{E}\left(\sum_{j=1}^{n} (Y_j - \mathrm{E}Y_j) \right)^p \right| \le \mathrm{E}\left| \sum_{j=1}^{n} (Y_j - \mathrm{E}Y_j) \right|^p \le K n^{p/2}.$$

Finally, from (3.16) the theorem follows. □

An improved convergence rate follows immediately by choosing, in the previous result, $\varphi(x) = x^p$ and $b_n = n^{1/2}(\log n (\log\log n)^\delta)^{1/p}$.

Corollary 3.13 *Let X_n, $n \in \mathbb{N}$, be centred and associated variables such that $\sup_{n\in\mathbb{N}} \|X_n\|_r < \infty$ for some $r > p \ge 2$. Assume that $u(n) \le Cn^{-\alpha}$, where $C > 0$ is some constant, for some $\alpha \ge \frac{r(p-2)}{2(r-p)}$. Then, for every $\delta > 1$,*

$$\frac{n^{1/2}}{(\log n (\log\log n)^\delta)^{1/p}} \frac{1}{n} S_n \longrightarrow 0 \quad \textit{almost surely}.$$

3.3 Large Deviations

After characterizing convergence rates for the Strong Law of Large Numbers, we will now study the convergence of tail probabilities $\mathbf{P}(\frac{1}{n} S_n \ge a)$, extending large distribution characterizations to associated variables. As usual, the typical result describes the asymptotic behaviour of $\frac{1}{n} \log \mathbf{P}(\frac{1}{n} S_n \ge a)$ using the moment generating function of the X_n variables. The approach followed below is taken from Henriques and Oliveira [44] and is largely inspired in the extension to various mixing structures proved by Bryc [20] and Bryc and Dembo [21] (see also Dembo and Zeitouni [29] for an account of the existing results on this direction), assuming a convenient decrease rate on the significant coefficients. The large deviation principle for associated variables requires a rather stringent decrease rate on the covariance structure, which corresponds to the translation of the assumptions used to prove similar results for mixing sequences.

Definition 3.14 A sequence of random variables X_n, $n \in \mathbb{N}$, is said to satisfy the *large deviation principle* with rate function r if:

(1) for every closed $F \subset \mathbb{R}$,

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n} S_n \in F \right) \le - \inf_{x\in F} r(x);$$

(2) for every open $G \subset \mathbb{R}$,

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n} S_n \in G \right) \ge - \inf_{x\in G} r(x).$$

The basic tool for proving large deviation principles is the well-known Gärtner–Ellis theorem (see Theorem B.1 and, for some more general results on large deviations, Appendix B). Taking into account Theorem B.1, the upper bound on the large deviation principle will follow if we prove, for each $u \in \mathbb{R}$, the existence and finiteness of

$$\Lambda(u) = \lim_{n\to+\infty} \frac{1}{n} \log \mathrm{E}\big(e^{uS_n}\big). \tag{3.17}$$

To prove the lower bound, one is compelled to some extra effort, as the direct verification of the differentiability, required by part (b) in Theorem B.1, is harder. The work around this difficulty uses Theorem B.6, to get the large deviation principle followed, by Theorem B.4, to identify the rate function.

One key assumption throughout this section concerns the behaviour of the density function of $\frac{1}{n}S_n$:

> There exist constants $c > 0$ and $B > 0$ such that, for each $n \in \mathbb{N}$, the density function of $\frac{1}{n}S_n$ is bounded by cB^n. (3.18)

This condition seems somewhat restrictive, but it is easily seen that, if the random variables X_n have distribution with common bounded support, then (3.18) is fulfilled.

In order to describe the asymptotic results below, we need another definition.

Definition 3.15 Given a function $\Lambda : \mathbb{R} \longrightarrow \mathbb{R}$, the *Fenchel–Legendre transform*, or convex conjugate, of Λ is the real-valued function Λ^* defined by

$$\Lambda^*(x) = \sup_{u\in\mathbb{R}} \big(ux - \Lambda(u)\big). \tag{3.19}$$

We first establish the upper bound of the large deviation principle by a direct application of Theorem B.1.

Theorem 3.16 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables such that there exists $M > 0$ for which $\mathbf{P}(|X_1| \le M) = 1$. Then, Λ as defined by* (3.17) *exists and is finite for every $u \in \mathbb{R}$. Moreover, for every closed $F \subset \mathbb{R}$,*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n}S_n \in F\right) = -\inf_{u\in F} \Lambda^*(u),$$

where Λ^* is the Fenchel–Legendre transform of Λ.

Proof As the variables are associated, given $u \in \mathbb{R}$ and $n, m \in \mathbb{N}$, we have

$$\mathrm{Cov}\big(e^{uS_n}, e^{u(S_{n+m}-S_n)}\big) \ge 0,$$

thus, from the stationarity it still follows that $\mathrm{E}(e^{uS_{n+m}}) \ge \mathrm{E}(e^{uS_n})\mathrm{E}(e^{uS_m})$ or, equivalently,

$$\log \mathrm{E}\big(e^{uS_{n+m}}\big) \ge \log \mathrm{E}\big(e^{uS_n}\big) + \log \mathrm{E}\big(e^{uS_m}\big).$$

Denote $h(n) = -\log \mathrm{E}(e^{uS_n})$. Then, $h(n+m) \leq h(n) + h(m)$, that is, h is subadditive, so the sequence $h(n)$, $n \in \mathbb{N}$, verifies the conditions of Lemma C.5 with $\varepsilon_n = 0$. Hence, according to this Lemma C.5, the following limit exists:

$$\Lambda(u) = \lim_{n \to +\infty} \frac{1}{n} \log \mathrm{E}\bigl(e^{uS_n}\bigr) = -\lim_{n \to +\infty} \frac{h(n)}{n}.$$

It remains to prove that the limit is finite. As the random variables are bounded by M, it follows that, for $u > 0$ and $n \in \mathbb{N}$, $e^{-unM} \leq \mathrm{E}(e^{uS_n}) \leq e^{unM}$, or, equivalently, $-uM \leq \frac{1}{n} \log \mathrm{E}(e^{uS_n}) \leq uM$, so, taking the limit as $n \to +\infty$, we have $-uM \leq \Lambda(u) \leq uM$. Repeating the argument for $u < 0$, we get in this case that $uM \leq \Lambda(u) \leq -uM$, so $\Lambda(u)$ is finite for every $u \in \mathbb{R}$. Finally, the second part of this theorem follows directly from the Gärtner–Ellis theorem. $\square$

The next step is the proof of the lower bound in Definition 3.14. As mentioned before, this will be accomplished using Theorems B.4 and B.6. The main steps to obtain this lower bound are the proof of the existence of a generalized limit analogous to (3.17), which will identify a candidate for the rate function, but is dependent on a suitable class of functions, followed by the proof of the convexity of this candidate to be a rate function. This later part needs some technical intermediate steps in the proof. We start by a lemma on sequences of real numbers taking care of some convergence aspects.

Lemma 3.17 *Let $t(n)$, $n \in \mathbb{N}$, be a sequence of nonnegative real numbers such that there exist $a, b > 0$ verifying $t(n) \leq a \exp(-n \log^{1+b} n)$. Then, for all $k < b$ and $c \in \mathbb{R}$,*

$$\lim_{n \to +\infty} nt\left(\frac{n}{\log^{1+k} n}\right) e^{cn} = 0.$$

Proof For each $n \in \mathbb{N}$, define $v(n) = \int_n^\infty \exp(-x \log^{1+b} x)\, dx$. As $0 \leq t(n) \leq av(n)$, it is enough to prove that

$$\lim_{n \to +\infty} nv\left(\frac{n}{\log^{1+k} n}\right) e^{cn} = 0.$$

Defining $h_1(x) = x \log^{-(1+k)} x$ and $h_2(x) = x^{-1} e^{-cx}$, the previous limit follows from $\lim_{x \to +\infty} \frac{v(h_1(x))}{h_2(x)} = 0$, and this is easily proved using Cauchy's rule. $\square$

Theorem 3.18 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables such that there exists $M > 0$ for which $\mathbf{P}(|X_1| \leq M) = 1$. Assume that there exist constants $a, b > 0$ such that*

$$u(n) \leq a \exp\bigl(-n \log^{1+b} n\bigr). \tag{3.20}$$

Then, for every continuous and concave function g such that $\sup_{x \in \mathbb{R}} g(x) < \infty$, the following limit exists:

$$\Lambda_g = \lim_{n \to +\infty} \frac{1}{n} \log \mathrm{E}\bigl(e^{ng(n^{-1}S_n)}\bigr).$$

Proof Let g be continuous, concave and such that $\sup_{x\in\mathbb{R}} g(x) < \infty$. Without loss of generality, we assume that $-\infty < -B \le g(x) \le 0$ for all $x \in [-M, M]$. Being concave and continuous, g is also Lipschitz continuous on $[-M, M]$, that is, there exists $L > 0$ such that, for all $x, y \in [-M, M]$, $|g(x) - g(y)| \le L|x - y|$.

For all $m, n, \ell \in \mathbb{N}$, as the random variables are bounded we have that, with probability 1,

$$\begin{aligned}
&\frac{1}{n+m}\big|S_{n+m} - \big(S_n + (S_{n+m+\ell} - S_{n+\ell})\big)\big| \\
&\quad \le \frac{1}{n+m}\left(\sum_{i=n+1}^{n+\ell} |X_i| + \sum_{i=n+m+1}^{n+m+\ell} |X_i|\right) \le \frac{2\ell M}{n+m}.
\end{aligned}$$

Thus, as g is Lipschitzian,

$$g\left(\frac{1}{n+m} S_{n+m}\right) - g\left(\frac{1}{n+m}\big(S_n + (S_{n+m+\ell} - S_{n+\ell})\big)\right) \ge -L\frac{2\ell M}{n+m},$$

and from the concavity of the function it follows that

$$\begin{aligned}
&g\left(\frac{1}{n+m} S_{n+m}\right) \\
&\quad \ge \frac{n}{n+m} g\left(\frac{1}{n} S_n\right) + \frac{m}{n+m} g\left(\frac{1}{m}\big(S_n + (S_{n+m+\ell} - S_{n+\ell})\big)\right) - \frac{2\ell LM}{n+m}.
\end{aligned}$$

For the sequel of this proof, let us denote $h(n) = -\log \mathrm{E}(e^{ng(n^{-1}S_n)})$. Then the inequality above is equivalent to

$$h(n+m) \le 2\ell LM - \log \mathrm{E}\big(e^{ng(n^{-1}S_n)} e^{mg(m^{-1}(S_{n+m+\ell} - S_{n+m}))}\big). \tag{3.21}$$

Define now, for all $n \in \mathbb{N}$ and $x \in [-M, M]$, $f_n(x) = e^{ng(x)}$. We have then that, for all $x, y \in [-M, M]$,

$$\big|f_n(x) - f_n(y)\big| \le \big|ng(x) - ng(y)\big| \le Ln|x - y|.$$

That is, each f_n is Lipschitz continuous and thus absolutely continuous and almost everywhere differentiable. Moreover, when the derivative exists, $|f_n'(x)| \le Ln$. Using integration by parts, we have that, due to the association of the random variables,

$$\begin{aligned}
&\mathrm{Cov}\big(e^{ng(n^{-1}S_n)}, e^{mg(m^{-1}(S_{n+m+\ell} - S_{n+m}))}\big) \\
&\quad = \bigg|\int_{[-M,M]^2} f_n'(x) f_m'(y) \\
&\qquad \times \mathrm{Cov}\left(\mathbb{I}_{(-\infty,x]}\left(\frac{1}{n} S_n\right), \mathbb{I}_{(-\infty,y]}\left(\frac{1}{m}(S_{n+m+\ell} - S_{n+m})\right)\right) dx\, dy\bigg| \\
&\quad \le L^2 nm \\
&\qquad \times \int_{[-M,M]^2} \mathrm{Cov}\left(\mathbb{I}_{(-\infty,x]}\left(\frac{1}{n} S_n\right), \mathbb{I}_{(-\infty,y]}\left(\frac{1}{m}(S_{n+m+\ell} - S_{n+m})\right)\right) dx\, dy.
\end{aligned}$$

Taking into account the stationarity of the random variables, we still have that

$$
\begin{aligned}
&\mathrm{Cov}\big(e^{ng(n^{-1}S_n)}, e^{mg(m^{-1}(S_{n+m+\ell}-S_{n+m}))}\big)\\
&\quad\le L^2 nm\,\mathrm{Cov}\bigg(\frac{1}{n}S_n, \frac{1}{m}(S_{n+m+\ell}-S_{n+m})\bigg) = L^2\sum_{i=1}^{n}\sum_{j=n+\ell+1}^{n+m+\ell}\mathrm{Cov}(X_i, X_j)\\
&\quad\le L^2 n\sum_{i=\ell+2}^{\infty}\mathrm{Cov}(X_1, X_i) = L^2 nu(\ell+1) \le L^2(n+m)u(\ell).
\end{aligned}
$$

Now, recalling that $g(x) \ge -B$ for every $x \in [-M, M]$, the previous inequality implies that

$$
\frac{\mathrm{E}(e^{ng(n^{-1}S_n)}e^{mg(m^{-1}(S_{n+m+\ell}-S_{n+m}))})}{\mathrm{E}(e^{ng(n^{-1}S_n)})\mathrm{E}(e^{mg(m^{-1}(S_{n+m+\ell}-S_{n+m}))})} \ge 1 - L^2(n+m)u(\ell)e^{(n+m)B}. \tag{3.22}
$$

Define, for all $\ell, n \in \mathbb{N}$, $\Theta(\ell, n) = 1 - L^2 nu(\ell)e^{nB}$. From (3.22) it follows that

$$
\begin{aligned}
&\log \mathrm{E}\big(e^{ng(n^{-1}S_n)}e^{mg(m^{-1}(S_{n+m+\ell}-S_{n+m}))}\big)\\
&\qquad \ge -h(n) - h(m) + \log\big(\max\big(\Theta(\ell, n+m), 0\big)\big),
\end{aligned}
$$

so, inserting this into (3.21), we find

$$
h(n+m) \le 2\ell LM + h(n) + h(m) - \log\big(\max\big(\Theta(\ell, n+m), 0\big)\big). \tag{3.23}
$$

Choose now $\delta \in (0, b)$, where b is defined by (3.20). It follows from Lemma 3.17 that, as $n \longrightarrow +\infty$,

$$
\Theta\bigg(\frac{n}{\log^{1+\delta} n}, n\bigg) = 1 - L^2 nu\bigg(\frac{n}{\log^{1+\delta} n}\bigg)e^{nB} \longrightarrow 1.
$$

If we choose $\ell = [(n+m)\log^{-(1+\delta)}(n+m)]$, we have that, for $n+m$ large enough, $-\log(\max(\Theta(\ell, n+m), 0)) \le \ell$; thus, from (3.23) it follows that

$$
\begin{aligned}
h(n+m) &\le h(n) + h(m) + (2LM+1)\ell\\
&= h(n) + h(m) + (2LM+1)\frac{n+m}{\log^{1+\delta}(n+m)}.
\end{aligned}
$$

Finally, recalling Lemma C.5, we have that the limit below exists:

$$
\lim_{n\to+\infty}\frac{h(n)}{n} = \lim_{n\to+\infty}\frac{-\mathrm{E}(e^{ng(n^{-1}S_n)})}{n},
$$

which concludes the proof of the theorem. □

We still need one auxiliary result to prove the convexity of the rate function on the large deviation principle.

Lemma 3.19 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables such that there exists $M > 0$ for which $\mathbf{P}(|X_1| \le M) = 1$. Assume that (3.18) and (3.20) are satisfied. If $x_1, x_2 \in \mathbb{R}$ are such that, for each $\delta > 0$,*

$$
\liminf_{n\to+\infty}\frac{1}{n}\log \mathbf{P}\bigg(\bigg|\frac{1}{n}S_n - x_i\bigg| < \delta\bigg) > -\infty, \quad i = 1, 2,
$$

then

$$\inf_{\delta>0}\liminf_{n\to+\infty}\frac{1}{n}\log\frac{\mathbf{P}(|(2n)^{-1}S_{2n}-(x_1+x_2)/2|<\delta)}{\mathbf{P}(|n^{-1}S_n-x_1|<\delta/2)\mathbf{P}(|n^{-1}S_n-x_2|<\delta/2)}\geq 0.$$

Proof Let $\delta>0$ be fixed. From the assumptions it follows that there exists $c_1>0$ such that, for every sufficiently large n,

$$\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2}\right)\mathbf{P}\left(\left|\frac{1}{n}S_n-x_2\right|<\frac{\delta}{2}\right)\geq\exp(-nc_1). \tag{3.24}$$

Now, for every $n,\ell\in\mathbb{N}$,

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2},\left|\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})-x_2\right|<\frac{\delta}{2}\right)\\
&\quad-\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2}\right)\mathbf{P}\left(\left|\frac{1}{n}S_n-x_2\right|<\frac{\delta}{2}\right)\\
&\leq\left|\mathbf{P}\left(\frac{1}{n}S_n<x_1+\frac{\delta}{2},\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})<x_2+\frac{\delta}{2}\right)\right.\\
&\quad\left.-\mathbf{P}\left(\frac{1}{n}S_n<x_1+\frac{\delta}{2}\right)\mathbf{P}\left(\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})<x_2+\frac{\delta}{2}\right)\right|\\
&\quad+\left|\mathbf{P}\left(\frac{1}{n}S_n<x_1+\frac{\delta}{2},\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})\leq x_2+\frac{\delta}{2}\right)\right.\\
&\quad\left.-\mathbf{P}\left(\frac{1}{n}S_n<x_1+\frac{\delta}{2}\right)\mathbf{P}\left(\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})\leq x_2+\frac{\delta}{2}\right)\right|\\
&\quad+\left|\mathbf{P}\left(\frac{1}{n}S_n\leq x_1+\frac{\delta}{2},\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})<x_2+\frac{\delta}{2}\right)\right.\\
&\quad\left.-\mathbf{P}\left(\frac{1}{n}S_n\leq x_1+\frac{\delta}{2}\right)\mathbf{P}\left(\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})<x_2+\frac{\delta}{2}\right)\right|\\
&\quad+\left|\mathbf{P}\left(\frac{1}{n}S_n\leq x_1+\frac{\delta}{2},\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})\leq x_2+\frac{\delta}{2}\right)\right.\\
&\quad\left.-\mathbf{P}\left(\frac{1}{n}S_n\leq x_1+\frac{\delta}{2}\right)\mathbf{P}\left(\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})\leq x_2+\frac{\delta}{2}\right)\right|,
\end{aligned}$$

so, taking into account Corollary 2.36, we have

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2},\left|\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})-x_2\right|<\frac{\delta}{2}\right)\\
&\quad-\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2}\right)\mathbf{P}\left(\left|\frac{1}{n}S_n-x_2\right|<\frac{\delta}{2}\right)\\
&\leq 4B_n\operatorname{Cov}^{1/3}\left(\frac{1}{n}S_n,\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})\right),
\end{aligned}$$

where $B_n=2\max(2\pi^{-2},45cB^n)$, as the variables $\frac{1}{n}S_n$ and $\frac{1}{n}(S_{2n+\ell}-S_{n+\ell})$ have a density function bounded above by cB^n, according to (3.18). Thus, recalling that the variables are stationary, we get

$$\mathbf{P}\left(\left|\frac{1}{n}S_n - x_1\right| < \frac{\delta}{2}, \left|\frac{1}{n}(S_{2n+\ell} - S_{n+\ell}) - x_2\right| < \frac{\delta}{2}\right)$$
$$- \mathbf{P}\left(\left|\frac{1}{n}S_n - x_1\right| < \frac{\delta}{2}\right)\mathbf{P}\left(\left|\frac{1}{n}S_n - x_2\right| < \frac{\delta}{2}\right)$$
$$\leq 4B_n\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=n+\ell+1}^{\infty}\operatorname{Cov}(X_i, X_j)\right)^{1/3} = 4B_n n^{-1/3}u^{1/3}(\ell).$$

Taking again into account the stationarity and (3.24), we find that, for all $\ell \in \mathbb{N}$ and sufficiently large n,

$$\frac{\mathbf{P}(|n^{-1}S_n - x_1| < \delta/2, |n^{-1}(S_{2n+\ell} - S_{n+\ell}) - x_2| < \delta/2)}{\mathbf{P}(|n^{-1}S_n - x_1| < \delta/2)\mathbf{P}(|n^{-1}S_n - x_2| < \delta/2)}$$
$$\geq 1 - 4B_n n^{-1/3}u^{1/3}(\ell)\exp(c_1 n). \tag{3.25}$$

As the variables are bounded by M, it easily derived that

$$\left|\frac{1}{2n}S_{2n} - \frac{x_1+x_2}{2}\right|$$
$$\leq \frac{1}{2}\left|\left(\frac{1}{n}S_n - x_1\right) + \left(\frac{1}{n}(S_{2n+\ell} - S_{n+\ell}) - x_2\right)\right| + \frac{1}{2n}\left|\sum_{i=n+1}^{n+\ell}X_i + \sum_{i=2n+1}^{2n+\ell}X_i\right|$$
$$\leq \frac{1}{2}\left|\left(\frac{1}{n}S_n - x_1\right) + \left(\frac{1}{n}(S_{2n+\ell} - S_{n+\ell}) - x_2\right)\right| + \frac{\ell M}{n},$$

so that the following holds:

$$\mathbf{P}\left(\left|\frac{1}{2n}S_{2n} - \frac{x_1+x_2}{2}\right| < \delta\right)$$
$$\geq \mathbf{P}\left(\frac{1}{2}\left|\left(\frac{1}{n}S_n - x_1\right) + \left(\frac{1}{n}(S_{2n+\ell} - S_{n+\ell}) - x_2\right)\right| < \delta - \frac{\ell M}{n}\right)$$
$$\geq \mathbf{P}\left(\left|\frac{1}{n}S_n - x_1\right| < \delta - \frac{\ell M}{n}, \left|\frac{1}{n}(S_{2n+\ell} - S_{n+\ell}) - x_2\right| < \delta - \frac{\ell M}{n}\right).$$

Choose now $\ell = \frac{\delta n}{2M}$ and insert the resulting inequality into (3.25) to obtain

$$\frac{\mathbf{P}(|(2n)^{-1}S_{2n} - (x_1+x_2)/2| < \delta)}{\mathbf{P}(|n^{-1}S_n - x_1| < \delta/2)\mathbf{P}(|n^{-1}S_n - x_2| < \delta/2)}$$
$$\geq 1 - 4B_n n^{-1/3}u^{1/3}\left(\frac{\delta n}{2M}\right)\exp(c_1 n).$$

Thus, finally,

$$\liminf_{n\to+\infty}\frac{1}{n}\log\frac{\mathbf{P}(|(2n)^{-1}S_{2n} - (x_1+x_2)/2| < \delta)}{\mathbf{P}(|n^{-1}S_n - x_1| < \delta/2)\mathbf{P}(|n^{-1}S_n - x_2| < \delta/2)}$$
$$\geq \liminf_{n\to+\infty}\frac{1}{n}\log\left(\max\left(1 - 4B_n n^{-1/3}u^{1/3}\left(\frac{\delta n}{2M}\right)\exp(c_1 n), 0\right)\right).$$

The limit on the right is, according to Lemma 3.17, equal to 0, and so, as $\delta > 0$ is arbitrary, the theorem is proved. □

We are finally in position to state and prove the large deviation principle that we have been announcing throughout this section.

Theorem 3.20 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables such that there exists $M > 0$ for which $\mathbf{P}(|X_1| \leq M) = 1$. Assume that (3.18) and (3.20) are satisfied. Then, the random variables X_n, $n \in \mathbb{N}$, satisfy the large deviation principle with rate function $\Lambda^*(x)$, the Fenchel–Legendre transform of*

$$\Lambda(u) = \lim_{n\to+\infty} \frac{1}{n} \log \mathrm{E}\big(e^{uS_n}\big).$$

Proof As the variables are strictly stationary and uniformly bounded, the distributions of S_n, $n \in \mathbb{N}$, are exponentially tight (see page 178). Also, according to the comment at the end of Appendix B, the family of continuous, concave and bounded above functions is well separated (see page 179). Moreover, according to Theorem 3.18, the limit $\Lambda_g = \lim_{n\to+\infty} \frac{1}{n} \log \mathrm{E}(e^{ng(n^{-1}S_n)})$ exists for every g continuous, concave and bounded above. That is, the conditions of Theorem B.8 are fulfilled, so it follows that X_n, $n \in \mathbb{N}$, verifies the large deviation principle with a good rate function $r(\cdot)$ (see page 178). The proof of this theorem will be concluded, that is, the rate function will be identified as the Fenchel–Legendre transform of Λ if we prove that this rate function is convex. In fact, as we have proved the finiteness of Λ, defined by (3.17), we may apply Theorem B.4 to conclude. According to Theorem B.2, for every $x \in \mathbb{R}$,

$$\begin{aligned} r(u) &= -\inf_{\delta>0,\, y:|y-x|<\delta} \left\{ \liminf_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\left| \frac{1}{n} S_n - y \right| < \delta \right) \right\} \\ &= -\inf_{\delta>0,\, y:|y-x|<\delta} \left\{ \limsup_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\left| \frac{1}{n} S_n - y \right| < \delta \right) \right\}. \end{aligned}$$

As, given that $y \in (x-\delta, x+\delta)$, there exists some δ' such that $(x-\delta', x+\delta') \subset (y-\delta, y+\delta)$, we may write

$$\begin{aligned} r(u) &= -\inf_{\delta>0} \left\{ \liminf_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\left| \frac{1}{n} S_n - x \right| < \delta \right) \right\} \\ &= -\inf_{\delta>0} \left\{ \limsup_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\left| \frac{1}{n} S_n - x \right| < \delta \right) \right\}. \end{aligned}$$

Consider now $x_1, x_2 \in \mathbb{R}$ such that $r(x_1), r(x_2) < \infty$. Thus, x_1 and x_2 both satisfy the assumptions of Lemma 3.19, so it follows that

$$\inf_{\delta>0} \liminf_{n\to+\infty} \frac{1}{n} \log \frac{\mathbf{P}(|(2n)^{-1} S_{2n} - (x_1+x_2)/2| < \delta)}{\mathbf{P}(|n^{-1} S_n - x_1| < \delta/2)\mathbf{P}(|n^{-1} S_n - x_2| < \delta/2)} \geq 0.$$

We have then that

$$
\begin{aligned}
&-r\left(\frac{x_1+x_2}{2}\right)\\
&\quad=\inf_{\delta>0}\left(\limsup_{n\to+\infty}\frac{1}{n}\log\left(\mathbf{P}\left(\left|\frac{1}{n}S_n-\frac{x_1+x_2}{2}\right|<\delta\right)\right)\right)\\
&\quad\geq\inf_{\delta>0}\left(\liminf_{n\to+\infty}\frac{1}{2n}\log\left(\frac{\mathbf{P}(|(2n)^{-1}S_{2n}-(x_1+x_2)/2|<\delta)}{\mathbf{P}(|n^{-1}S_n-x_1|<\delta/2)\mathbf{P}(|n^{-1}S_n-x_2|<\delta/2)}\right)\right)\\
&\qquad+\inf_{\delta>0}\left(\liminf_{n\to+\infty}\frac{1}{2n}\log\left(\mathbf{P}\left(\left|\frac{1}{n}S_n-x_1\right|<\frac{\delta}{2}\right)\right)\right)\\
&\qquad+\inf_{\delta>0}\left(\liminf_{n\to+\infty}\frac{1}{2n}\log\left(\mathbf{P}\left(\left|\frac{1}{n}S_n-x_2\right|<\frac{\delta}{2}\right)\right)\right)\\
&\quad\geq-\frac{1}{2}r(x_1)-\frac{1}{2}r(x_2).
\end{aligned}
$$

That is, we have concluded that, for every $x_1, x_2 \in \mathbb{R}$, $r(\frac{x_1+x_2}{2}) \leq \frac{1}{2}r(x_1) + \frac{1}{2}r(x_2)$. Iterating now this inequality, it is easily proved that, for every $n \in \mathbb{N}$ and $k = 0, \ldots, 2^n$,

$$
r\left(\frac{k}{2^n}x_1+\left(1-\frac{k}{2^n}\right)x_2\right)\leq\frac{k}{2^n}r(x_1)+\left(1-\frac{k}{2^n}\right)r(x_2),
$$

and the convexity follows from the continuity of r. Thus, according to Theorem B.4, the rate function is the Fenchel–Legendre transform of Λ, that is, $r(u) = \Lambda^*(u)$. $\square$

3.4 Kernel Density Estimation

The technical treatment of the kernel density or regression estimators is similar. As could be expected, the regression problem is somewhat more intricate as we need to handle the dependence between the X and Y variables that are natural in this problem. We will start by proving the asymptotic results for the density estimator, introducing the techniques in a simpler framework, and then extend them to the treatment of the regression.

In this section we start looking at the classical statistical problems of, based on a sample, estimating the density function, assuming, naturally, that it exists. We will, of course, be interested in characterizing the asymptotics of the estimates assuming that the samples verify an association condition. As expected, the main difficulty will be to handle the variance of the estimators. This will be achieved assuming a mild condition on joint distributions, not even requiring the existence of joint densities, as is frequently done throughout the literature. The first results proving the consistency of nonparametric estimators were obtained by Roussas [88] assuming, among other technical conditions, that joint densities should be close enough to the product densities, thus, controlling the deviance from independence in a way rather similar to what is done while handling strong mixing samples. Even simple models generating associated variables construct sequences of random variables for which their joint distributions do not have a joint density, just having some mass concentrated on the diagonal (see Example 1.28). We will prove the consistency of kernel estimates allowing for this mass concentration on the diagonal. It is curious to notice right away that it is exactly this diagonal mass that will appear in the characterization of the asymptotic results concerning the variances.

3.4.1 Definitions and Preliminary Results

We now describe the framework for our statistical problem, introducing the estimator that will be studied in the sequel. Let $X_1, X_2, \ldots$ be an associated sequence of random variables with the same distribution as X for which there exists a density function f. Let K be a fixed probability density, and h_n a sequence of real numbers converging to zero.

Definition 3.21 The *kernel estimator of the density* function f is, of course, defined as

$$\widehat{f_n}(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right). \tag{3.26}$$

This estimator is well known to be asymptotically unbiased if there exists a suitable version of the density, as a consequence of the Bochner's lemma (see Bochner [19]).

Lemma 3.22 *If f is bounded and continuous, then $\mathrm{E}\widehat{f_n}(x) \longrightarrow f(x)$ uniformly on any compact set.*

This means also that, in order to establish the convergence of $\widehat{f_n}(x)$, it is enough to prove that $\widehat{f_n}(x) - \mathrm{E}\widehat{f_n}(x) \longrightarrow 0$ in the appropriate mode of convergence.

We now introduce the assumptions that describe the appropriate control on the diagonal decomposition of the joint distributions. Let λ^2 be, as usual, the Lebesgue measure on $\mathbb{R}^2$. Let Δ be the diagonal of $\mathbb{R} \times \mathbb{R}$, and represent by λ^* the measure on Δ defined by $\lambda^*\{(u, u), u \in A\} = \lambda(A)$, where A is Borel subset of $\mathbb{R}$, and λ is the Lebesgue measure on $\mathbb{R}$. The first set of conditions may now be described:

(D)

- **(D.1)** For each $j, k \in \mathbb{N}$, the distribution of (X_j, X_k) is the sum of a measure $m_{1,j,k}$ on $\mathbb{R} \times \mathbb{R} \setminus \Delta$ with a measure $m_{2,j,k}$ on Δ such that $m_{1,j,k} \ll \lambda^2$ and $m_{2,j,k} \ll \lambda^*$.
- **(D.2)** For each $j, k \in \mathbb{N}$, there exists a bounded version $g_{1,j,k}$ of $\frac{dm_{1,j,k}}{d\lambda^2}$.
- **(D.3)** For all $j, k \in \mathbb{N}$, there exists a bounded and continuous version $g_{2,j,k}$ of $\frac{dm_{2,j,k}}{d\lambda^*}$.

Notice that the pairs are allowed to have some mass concentrated on the diagonal, and thus they need not to be absolutely continuous with respect to λ^2. A simple example where this property is required arises from a method used to construct sequences of associated variables from an independent sequence of variables, as described in Example 1.28: assuming that $m \in \mathbb{N}$ is some fixed integer, define, for each $n \in \mathbb{N}$, $X_n = \min(Y_n, \ldots, Y_{n+m})$. The variables X_n, $n \in \mathbb{N}$, are associated, have a common absolutely continuous distribution, but the random pairs (X_n, X_{n+j}), $j = 1, \ldots, m$, are not absolutely continuous. Nevertheless, their joint distribution satisfies **(D)**.

In order to prove the convergence results, we will need some control on Radon–Nikodym derivatives introduced in **(D)**, thus we introduce a second set of conditions:

(A)
- **(A.1)** **(D)** is satisfied.
- **(A.2)** $\frac{1}{n}\sum_{j,k=1}^{n} |g_{1,j,k}(x,y) - f(x)f(y)|$ converges uniformly to a bounded function g_1.
- **(A.3)** $\frac{1}{n}\sum_{j,k=1}^{n} g_{2,j,k}(x,x)$ converges uniformly to a bounded and continuous function g_2.

It is now convenient to introduce the following notation.

Definition 3.23 Given functions $h_1 : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h_2 : \mathbb{R}^m \longrightarrow \mathbb{R}$, we denote by $h_1 \otimes h_2$ the function defined on $\mathbb{R}^{n+m}$ by

$$h_1 \otimes h_2(x_1, \ldots, x_n, y_1, \ldots, y_m) = h_1(x_1, \ldots, x_n) h_2(y_1, \ldots, y_m).$$

With respect to **(A)**, remark that, if the variables $X_1, X_2, \ldots$ are strictly stationary, it is easy to check that

$$\frac{1}{n}\sum_{j,k=1}^{n} |g_{1,j,k} - f \otimes f| \le \sum_{j=1}^{n-1} |g_{1,1,j} - f \otimes f|,$$

and analogously for the summation in **(A.3)**. In such a case, it is enough to assume the convergence of these upper bounds.

The main tool for proving the convergence is stated in the following lemma.

Lemma 3.24 *Let X_n, $n \in \mathbb{N}$, be random variables such that* **(A)** *is satisfied. If the kernel K is square integrable, then*

$$\frac{1}{nh_n}\sum_{j,k=1}^{n} \operatorname{Cov}\left(K\left(\frac{x - X_j}{h_n}\right), K\left(\frac{x - X_k}{h_n}\right)\right) \longrightarrow g_2(x,x) \int K^2(u)\, du$$

uniformly on any compact set.

Proof Write

$$\begin{aligned}
&\frac{1}{h_n}\operatorname{Cov}\left(K\left(\frac{x - X_j}{h_n}\right), K\left(\frac{x - X_k}{h_n}\right)\right) \\
&\quad = \frac{1}{h_n}\int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) \big(g_{1,j,k}(u,v) - f(u)f(v)\big)\, du\, dv \\
&\qquad + \frac{1}{h_n}\int_{\Delta} K^2\left(\frac{x-u}{h_n}\right) g_{2,j,k}(u,u)\, du. \qquad (3.27)
\end{aligned}$$

The first integral in (3.27) is bounded above by

$$\int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) \left|g_{1,j,k}(u,v) - f(u)f(v)\right| du\, dv,$$

so that, for $n \in \mathbb{N}$ large enough,

$$\begin{aligned}
&\frac{1}{nh_n} \sum_{j,k=1}^{n} \int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) \left|g_{1,j,k}(u,v) - f(u)f(v)\right| du\, dv \\
&\quad \le 2h_n \frac{1}{h_n^2} \int_{\mathbb{R}^2} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) g_1(u,v)\, du\, dv \\
&\quad \le 2h_n \sup_{u,v} \left|g_1(u,v)\right| \longrightarrow 0,
\end{aligned}$$

as $h_n \longrightarrow 0$. For the second integral in (3.27), we have that

$$\frac{1}{h_n} \int_{\Delta} K^2\left(\frac{x-u}{h_n}\right) \frac{1}{n} \sum_{j,k=1}^{n} g_{2,j,k}(u,u)\, du \longrightarrow g_2(x,x) \int K^2(u)\, du$$

using Bochner's lemma after renormalizing K^2 to find a density. □

Remark 3.25 The previous result states a general convergence for the covariances not requiring the association of the random variables.

Remark 3.26 Notice that the limit obtained in Lemma 3.24 is the diagonal density multiplied by a constant depending only on the kernel. So, if we had assumed the absolute continuity of the random pairs (X_j, X_k), this limit would have been zero. As this term appears when dealing with the convergence of the estimator, in case it converges to zero most of the convergences would follow with relaxed assumptions on the bandwidth sequence h_n.

The convergence in probability of $\widehat{f}_n(x)$ is a simple and immediate consequence of Lemma 3.24, as stated in the next result.

Theorem 3.27 *Let X_n, $n \in \mathbb{N}$, be random variables such that* **(A)** *is satisfied. If the kernel K is square integrable and*

$$nh_n \longrightarrow +\infty, \tag{3.28}$$

then $\widehat{f}_n(x)$ converges in probability to $f(x)$ for every $x \in \mathbb{R}$.

Proof As

$$\begin{aligned}
&\mathbf{P}\left(\left|\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)\right| > \varepsilon\right) \\
&\quad \le \frac{1}{\varepsilon^2 nh_n} \frac{1}{nh_n} \sum_{j,k=1}^{n} \mathrm{Cov}\left(K\left(\frac{x-X_j}{h_n}\right), K\left(\frac{x-X_k}{h_n}\right)\right),
\end{aligned}$$

the theorem follows immediately taking into account Lemmas 3.22 and 3.24. □

3.4.2 Almost Sure Consistency

The almost sure convergence requires some further assumptions on the kernel K and on the bandwidth sequence. This is due to the fact that the variables $K(\frac{x-X_j}{h_n})$ are, in general, not associated. In order to keep the association, we should apply only monotone transformations to the original variables, which is not the case with a general kernel. This may be resolved by assuming the kernel K to be of bounded variation, which includes most of the popular choices for the kernel function. There are some additional technical assumptions required by the method of proof, which follows closely the proof of Theorem 2.7.1 in Ferrieux [39]. For a more general setting, it is interesting to have a look at Ferrieux [38], although, as in [39], the author is interested in estimation problems for point processes, a framework that is not the one we are studying here, as it follows from Sect. 1.4. The result below appears in this form in Oliveira [74].

Theorem 3.28 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume that* **(A)** *is satisfied, K is of bounded variation such that, for fixed $x \in \mathbb{R}$, $K(\alpha x)$ is a decreasing function of $\alpha > 0$, and*

$$h_n \searrow 0, \qquad \sum_{n=1}^{\infty} \frac{1}{n^2 h_{n^2}} < \infty, \qquad \frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1.$$

Then $\widehat{f}_n(x)$ converges almost surely to $f(x)$ for every $x \in \mathbb{R}$.

Proof We first check that the subsequence corresponding to terms of order n^2 converges almost surely. This is an immediate consequence of the Borel–Cantelli lemma, Lemma 3.24 and the assumptions made, as

$$\begin{aligned}
&\mathbf{P}\left(\left|\widehat{f}_{n^2}(x) - \mathrm{E}\widehat{f}_{n^2}(x)\right| > \varepsilon\right) \\
&\quad \leq \frac{1}{\varepsilon^2 n^2 h_{n^2}} \frac{1}{n^2 h_{n^2}} \sum_{j,k=1}^{n^2} \mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right), K\left(\frac{x-X_k}{h_{n^2}}\right)\right).
\end{aligned}$$

For the remaining terms, we write, for an integer $k \in (n^2, (n+1)^2]$,

$$\begin{aligned}
&\left|\left(\widehat{f}_k(x) - \mathrm{E}\widehat{f}_k(x)\right) - \left(\widehat{f}_{n^2}(x) - \mathrm{E}\widehat{f}_{n^2}(x)\right)\right| \\
&\quad \leq \frac{1}{n^2} \max_{n^2 < k \leq (n+1)^2} \left| \frac{1}{h_k} \sum_{j=1}^{k} \left[K\left(\frac{x-X_j}{h_k}\right) - \mathrm{E}K\left(\frac{x-X_j}{h_k}\right)\right] \right. \\
&\qquad \left. - \frac{1}{h_{n^2}} \sum_{j=1}^{n^2} \left[K\left(\frac{x-X_j}{h_{n^2}}\right) - \mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right] \right| \\
&\qquad + \frac{1}{n^2 h_{n^2}} \left| \sum_{j=1}^{n^2} \left[K\left(\frac{x-X_j}{h_{n^2}}\right) - \mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right] \right|,
\end{aligned}$$

thus leaving the first term to be treated. Now,

$$\frac{1}{h_k}\sum_{j=1}^{k}\left[K\left(\frac{x-X_j}{h_k}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_k}\right)\right]$$

$$-\frac{1}{h_{n^2}}\sum_{j=1}^{n^2}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right]$$

$$=\sum_{j=1}^{k}\left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_k}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_k}\right)\right]$$

$$+\sum_{j=1}^{k}\frac{1}{h_{n^2}}\left(\left[K\left(\frac{x-X_j}{h_k}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_k}\right)\right]\right.$$

$$\left.-\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right]\right)$$

$$+\sum_{j=n^2+1}^{k}\frac{1}{h_{n^2}}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right].$$

Let us denote by a_n, b_n and c_n the maxima over $k \in (n^2,(n+1)^2]$ of each of the three terms on the right of the preceding expression, respectively. The consistency of the estimator (3.26) follows if we prove that $\frac{a_n}{n^2}$, $\frac{b_n}{n^2}$ and $\frac{c_n}{n^2}$ all converge almost surely to 0.

Convergence of $\frac{a_n}{n^2}$. As h_n is decreasing and using the decreasing assumption on the kernel, it follows that

$$0 \le \left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)K\left(\frac{x-u}{h_k}\right) \le \left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)K\left(\frac{x-u}{h_{n^2}}\right).$$

Now, for every $k \in (n^2,(n+1)^2]$,

$$\frac{1}{n^2}\left|\sum_{j=1}^{k}\left(\frac{1}{h_k}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_k}\right)-\mathrm{E}K\left(\frac{x-X_j}{h_k}\right)\right]\right|$$

$$\le \frac{1}{n^2}\sum_{j=1}^{(n+1)^2}\left(\frac{1}{h_{(n+1)^2}}-\frac{1}{h_{n^2}}\right)\left[K\left(\frac{x-X_j}{h_{n^2}}\right)+\mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)\right]. \tag{3.29}$$

We first look at the terms with the mathematical expectation. On one hand, we have

$$\frac{1}{n^2 h_{(n+1)^2}}\sum_{j=1}^{(n+1)^2}\mathrm{E}K\left(\frac{x-X_j}{h_{n^2}}\right)=\frac{(n+1)^2}{n^2}\frac{h_{n^2}}{h_{(n+1)^2}}\frac{1}{h_{n^2}}\mathrm{E}K\left(\frac{x-X}{h_{n^2}}\right)\longrightarrow f(x),$$

using Lemma 3.22. Analogously, it follows that

$$\frac{1}{n^2 h_{n^2}} \sum_{j=1}^{(n+1)^2} \mathrm{E}K\left(\frac{x - X_j}{h_{n^2}}\right) \longrightarrow f(x).$$

Thus, in (3.29) the terms corresponding to the mathematical expectations converge to zero. This allows replacing the sign "+" inside the square brackets on the right-hand side of (3.29) by the sign "−" as, given $\delta > 0$, for n large enough, the right-hand side in (3.29) becomes smaller than

$$\frac{1}{n^2} \sum_{j=1}^{(n+1)^2} \left(\frac{1}{h_{(n+1)^2}} - \frac{1}{h_{n^2}}\right)\left[K\left(\frac{x - X_j}{h_{n^2}}\right) - \mathrm{E}K\left(\frac{x - X_j}{h_{n^2}}\right)\right] + \delta.$$

So, it is enough to verify that the summation above converges almost surely to zero. In the present form, Chebyshev's inequality gives an upper bound with a variance term: for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbf{P}\left(\frac{1}{n^2}\left|\sum_{j=1}^{(n+1)^2} \left(\frac{1}{h_{(n+1)^2}} - \frac{1}{h_{n^2}}\right)\left[K\left(\frac{x - X_j}{h_{n^2}}\right) - \mathrm{E}K\left(\frac{x - X_j}{h_{n^2}}\right)\right]\right| > \varepsilon\right) \\
&\quad \leq \frac{1}{\varepsilon^2 n^4}\left(\frac{1}{h_{(n+1)^2}} - \frac{1}{h_{n^2}}\right)^2 \sum_{j,j'=1}^{(n+1)^2} \mathrm{Cov}\left(K\left(\frac{x - X_j}{h_{n^2}}\right), K\left(\frac{x - X_{j'}}{h_{n^2}}\right)\right) \\
&\quad = \frac{1}{\varepsilon^2 n^2 h_{n^2}}\left(\frac{h_{n^2}}{h_{(n+1)^2}} - 1\right)^2 \frac{1}{n^2 h_{n^2}} \sum_{j,j'=1}^{(n+1)^2} \mathrm{Cov}\left(K\left(\frac{x - X_j}{h_{n^2}}\right), K\left(\frac{x - X_{j'}}{h_{n^2}}\right)\right),
\end{aligned}$$

which, taking account of Lemma 3.24 and the assumptions made on the bandwidth sequence, defines a convergent series.

Convergence of $\frac{b_n}{n^2}$. Using the decreasing assumption on the kernel, it follows that, for every $k \in (n^2, (n+1)^2]$,

$$\begin{aligned}
\frac{|b_n|}{n^2} \leq \frac{1}{n^2 h_{n^2}} \sum_{j=1}^{(n+1)^2} &\left(\left[K\left(\frac{x - X_j}{h_{n^2}}\right) + \mathrm{E}K\left(\frac{x - X_j}{h_{n^2}}\right)\right]\right. \\
&\left. - \left[K\left(\frac{x - X_j}{h_{(n+1)^2}}\right) + \mathrm{E}K\left(\frac{x - X_j}{h_{(n+1)^2}}\right)\right]\right).
\end{aligned}$$

As for the convergence of the terms $\frac{a_n}{n^2}$, it is easy to check that the terms with the mathematical expectations cancel each other in the limit. So, we are left with checking the almost sure convergence to zero of

$$\frac{1}{n^2 h_{n^2}} \sum_{j=1}^{(n+1)^2} \left[K\left(\frac{x - X_j}{h_{n^2}}\right) - K\left(\frac{x - X_j}{h_{(n+1)^2}}\right)\right].$$

Given $\varepsilon > 0$,

$$\mathbf{P}\left(\frac{1}{n^2 h_{n^2}}\left|\sum_{j=1}^{(n+1)^2}\left[K\left(\frac{x-X_j}{h_{n^2}}\right)-K\left(\frac{x-X_j}{h_{(n+1)^2}}\right)\right]\right|>\varepsilon\right)$$
$$\leq \frac{1}{\varepsilon^2 n^4 h_{n^2}^2}\sum_{j,j'=1}^{(n+1)^2}\mathrm{Cov}\left(K\left(\frac{x-X_j}{h_{n^2}}\right)-K\left(\frac{x-X_j}{h_{(n+1)^2}}\right),\right.$$
$$\left.K\left(\frac{x-X_{j'}}{h_{n^2}}\right)-K\left(\frac{x-X_{j'}}{h_{(n+1)^2}}\right)\right).$$

Taking into account Lemma 3.24 and $\frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1$, we get that the sum of these covariances, divided by $n^2 h_{n^2}$, is convergent to $4g_2(x,x)\int K^2(u)\,du$. As there remains the term $\frac{1}{n^2 h_{n^2}^2}$, we have in fact a convergent series defined by the probabilities.

Convergence of $\frac{c_n}{n^2}$. Write $K = K_1 - K_2$, with K_1, K_2 increasing functions. The random variables $K_1(\frac{x-X_j}{h_{n^2}})$, $j = 1, 2, \ldots$, being monotone transformations of associated variables, are associated. Then, we may apply the generalization of Kolmogorov's inequality for associated variables proved in Theorem 2.22 to obtain that, given $\varepsilon > 0$,

$$\mathbf{P}\left(\frac{1}{n^2 h_{n^2}}\max_{n^2<k\leq(n+1)^2}\left|\sum_{j=n^2+1}^{k}\left[K_1\left(\frac{x-X_j}{h_{n^2}}\right)-\mathrm{E}K_1\left(\frac{x-X_j}{h_{n^2}}\right)\right]\right|>\varepsilon\right)$$
$$\leq \frac{2}{\varepsilon^2 n^4 h_{n^2}^2}\sum_{j,j'=n^2+1}^{(n+1)^2}\mathrm{Cov}\left(K_1\left(\frac{x-X_j}{h_{n^2}}\right),K_1\left(\frac{x-X_{j'}}{h_{n^2}}\right)\right).$$

Because of the association, this sum is bounded above by the sum with j, j' ranging from 1 to $(n+1)^2$, and then the proof is completed repeating the arguments used for the two previous cases. The terms corresponding to K_2 are treated analogously. □

Remark 3.29 Notice that the association of the random variables is only used in the final step of the proof.

Remark 3.30 The assumptions made on the kernel function, namely, that K is of bounded variation and such that $K(\alpha x)$, for fixed $x \in \mathbb{R}$, is a decreasing function of $\alpha > 0$, are met by most of the usual kernels considered in the statistical literature.

3.4.3 Almost Sure Convergence Rates

Characterization of strong convergence rates relies on appropriate inequalities for sums of variables. The main tools have been exponential inequalities such as the one proved in Theorem 2.52 or Rosenthal inequalities like (2.3). Results based on the later inequalities require milder assumptions on the covariance decay rate of the variables. In fact, as already mentioned before, exponential inequalities are not available unless the covariances decrease geometrically. There is, though, an interest on the exponential inequality approach: the rates derived are better, closer to the optimal $(\frac{\log\log n}{nh_n})^{1/2}$ rate known for independent random variables. We will next prove a convergence rate based on an adaptation of (2.3) that is just slightly slower than the optimal rate.

Theorem 3.31 *Let $p \geq 2$, $r > p + \frac{2}{3}$, and X_n, $n \in \mathbb{N}$, be associated random variables with density function f such that $\sup_{x\in\mathbb{R}} |f(x)| \leq B < \infty$ and $u(n) \leq Cn^{-\theta}$ with $\theta \geq \frac{r(p-2)}{2(r-p)}$. Assume that the kernel function K is differentiable with bounded derivative and that K^r is integrable. Then, for every $\gamma \in (0,1)$, $\widehat{f_n}(x)$ converges almost surely to $f(x)$ with rate $(\frac{\log\log n}{nh_n})^{\gamma/2}$.*

Proof The main argument for this proof is to find a suitable version of the moment inequality (2.5) in Theorem 2.18. For this purpose, define, for all $i, n \in \mathbb{N}$,

$$Y_{j,n} = \frac{1}{h_n}\left(K\left(\frac{x - X_j}{h_n}\right) - \mathrm{E}K\left(\frac{x - X_j}{h_n}\right)\right).$$

With respect to the proof of Theorem 2.18, we refer here only what should be changed, as the arguments are completely parallel. Denote $r_p = r(p-1) - 1$ and define, similarly to proof of Theorem 2.18, $T_n = \sum_{i=1}^n Y_{i,n}$. Given $a \in (0, \frac{1}{2})$, put $m = [na] + 1$ and $k_n = [\frac{n}{2m}] + 1$. Decompose T_n into the blocks:

$$\xi_\ell = \sum_{j=2(\ell-1)m+1}^{n\wedge(2\ell-1)m} Y_{j,n} \quad \text{and} \quad \eta_\ell = \sum_{j=(2\ell-m)+1}^{n\wedge 2\ell m} Y_{j,n}, \quad \ell = 1, \ldots, k_n,$$

and define the sums of alternating blocks: $T_{1,n} = \sum_{\ell=1}^{k_n} \xi_\ell$ and $T_{2,n} = \sum_{\ell=1}^{k_n} \eta_\ell$. The initial arguments are the same as in the proof of Theorem 2.18. Notice now that, taking into account the boundedness of the density f, we have

$$\frac{1}{h_n^r}\int \left|K\left(\frac{x-u}{h_n}\right)\right|^r f(u)\,du \leq \frac{B}{h_n^{r-1}}\int |K(z)|^r\,dz,$$

so the bound corresponding to (2.7) becomes

$$\begin{aligned} A_2 \leq & \sum_{\ell=1}^{k_n} \mathrm{E}|\xi_\ell|\mathrm{E}|T_{1,n} - \xi_\ell|^{p-1} \\ & + \sum_{\ell=1}^{k_n} p\big(\mathrm{E}|T_{1,n} - \xi_\ell|^p\big)^{(r-1)(p-2)/r_p} \\ & \times \left(\frac{2Bm}{h_n^{1-1/r}}\int |K(z)|^r\,dz\right)^{r(p-2)/r_p}\left(\frac{m\|K'\|_\infty}{h_n^4}u(m)\right)^{(r-p)/r_p}. \end{aligned}$$

Thus, we may continue with the argument as in the proof of Theorem 2.18, redefining

$$C_1 = \frac{(2B\int |K(z)|^r\,dz)^{r(p-2)/r_p}\,\|K'\|_\infty^{2(r-p)/(r-2)}}{h_n^{1+(3r-4p+2)/r_p}}\big(u(m)\big)^{(r-p)/(r-2)}.$$

Repeating the arguments, we need to redefine the constant b_p in the bound corresponding to (2.9) as

$$b_p = \frac{(p2^p + 4p2^{((2p-3)/(r-2))r_p})}{h_n^{(1+(3r-4p+2)/r_p)r_p/(r-2)}}\left(\left(2B\int \left|K(z)\right|^r dz\right)^{r(p-2)/r_p}\right.$$
$$\left.\times \left\|K'\right\|_\infty^{2(r-p)/(r-2)}\big(u(m)\big)^{(r-p)/(r-2)}\right)^{r_p/(r-2)}.$$

Remark that the denominator of this last expression may be written as $h_n^{2+p-3(p-2)/(r-2)}$. Now, taking into account that $u(n) \le Cn^{-\theta}$ and that $\theta \ge \frac{r(p-2)}{2(r-p)}$, repeating the arguments of Corollary 2.20, we find that there exist two constants K_1 and K_2, depending only on r, p and $u(0)$, such that

$$\mathrm{E}|T_n|^p \le n^{p/2}\left(K_1 + \frac{K_2}{h_n^{2+p-3(p-2)/(r-2)}}\right),$$

which is the sought convenient moment inequality. Now, the convergence rate for $\widehat{f_n}(x)$ follows using standard arguments based on the Borel–Cantelli lemma. In fact, we want to choose a sequence ε_n such that $\sum_{n=1}^\infty \mathbf{P}(\frac{1}{n}|T_n| > \varepsilon_n) < \infty$. As $\mathbf{P}(\frac{1}{n}|T_n| > \varepsilon_n) \le \frac{1}{n^p\varepsilon_n^p}\mathrm{E}|T_n|^p$, by the previous inequality, it is enough to verify the convergence of both series

$$\sum_{n=1}^\infty \frac{1}{n^{p/2}\varepsilon_n^p} \quad \text{and} \quad \sum_{n=1}^\infty \frac{1}{n^{p/2}h_n^{2+p-3(p-2)/(r-2)}\varepsilon_n^p}.$$

It is, of course, enough to find the convergence of the second series as $h_n \longrightarrow 0$. To find a convergent series, we set

$$n^{p/2}h_n^{2+p-3(p-2)/(r-2)}\varepsilon_n^p = n\log^\delta n \quad \text{for some } \delta > 1.$$

If we choose $\varepsilon_n = (\frac{\log\log n}{nh_n})^{\gamma/2}$, this means that, since $p \ge 2$,

$$h_n^{2+(1-\gamma/2)p-3(p-2)/(r-2)} = n^{1-((1-\gamma)/2)p}\frac{\log^\delta n}{(\log\log n)^{\delta p/2}} \longrightarrow 0,$$

so the exponent of h_n should be positive, but this is true as long as $\gamma < \frac{3p}{2}\frac{r-p}{r-2}$. Finally it is easy to check that $\frac{3p}{2}\frac{r-p}{r-2} > 1$ is equivalent to $r > p + \frac{2}{3} - \frac{8}{3}\frac{1}{3p-2}$, so we may really choose γ arbitrarily close to 1. □

Remark 3.32 The argument used in the final part of the proof above fails if we choose $\varepsilon_n = (\frac{\log\log n}{nh_n})^{1/2}$. In fact, repeating the arguments, we would arrive at

$$h_n^{2+p/2-3(p-2)/(r-2)} = \frac{n\log^\delta n}{(\log\log n)^{p/2}} \longrightarrow +\infty,$$

so the exponent of h_n should now be negative. It is easily verified that, as $r > p \geq 2$, this cannot be true.

3.5 Kernel Regression Estimation

We now concentrate on the estimation of the regression function and the asymptotic results corresponding to those studied in the preceding subsections. We start by extending the framework appropriately. We begin by stating clearly our framework. Let $(X_1, Y_1), (X_2, Y_2), \dots$ be an associated sequence of random vectors with the same distribution as (X, Y), for which there exists a density function f^*. We will denote by f the marginal density corresponding to X and by h the marginal density corresponding to Y. Let K be a fixed probability density, and h_n a sequence of real numbers converging to zero.

Definition 3.33 The *kernel estimator of the regression* function $r(x) = \mathrm{E}(Y|X = x)$ is defined as

$$\widehat{r}_n(x) = \frac{\sum_{j=1}^n Y_j K((x - X_j)/h_n)}{\sum_{j=1}^n K((x - X_j)/h_n)} = \frac{(1/(nh_n)) \sum_{j=1}^n Y_j K((x - X_j)/h_n)}{\widehat{f}_n(x)}, \tag{3.30}$$

where $\widehat{f}_n(x)$ is the density estimator (3.26). If we denote

$$\widehat{m}_n(x) = \frac{1}{nh_n} \sum_{j=1}^n Y_j K\left(\frac{x - X_j}{h_n}\right),$$

then, obviously, $\widehat{r}_n(x) = \frac{\widehat{m}_n(x)}{\widehat{f}_n(x)}$.

The technical treatment of the density or regression estimators is similar to that of the expression for the corresponding estimators. As could be expected, the regression problem is somewhat more intricate as we need to handle the dependence between the X and Y variables that are natural in this problem. We will extend the techniques used in the previous section to prove the almost sure asymptotics for the regression estimator (3.30).

The proofs will be based in the inclusion proved in Theorem C.3, which enables the separation of the variables X and Y in the definition of (3.30). This separation result depends on a nonnegativity assumption on the variables. For the present setting, it is enough to assume the nonnegativity of the variable Y, as explained next. We now use the representation of (3.30) as $\widehat{r}_n(x) = \frac{\widehat{m}_n(x)}{\widehat{f}_n(x)}$. For each $x \in \mathbb{R}$, the random variables $\widehat{f}_n(x)$ and $\widehat{m}_n(x)$ are obviously nonnegative if we assume that Y is nonnegative valued. We need to control $|\widehat{r}_n(x) - r(x)|$, but this will be achieved indirectly through $\mathrm{E}\widehat{f}_n(x)$ and $\mathrm{E}\widehat{m}_n(x)$. We have already remarked that $\mathrm{E}\widehat{f}_n(x) \longrightarrow f(x)$ (see Lemma 3.22). So, in order to apply Theorem C.3, it is enough to assume that $f(x) > 0$. We still need to describe the asymptotic behaviour of $\mathrm{E}\widehat{m}_n(x)$.

Lemma 3.34 *Let Y be a random variable such that the regression function $r(x) = \mathrm{E}(Y|X=x)$ is continuous and assume that the density f of X is continuous. If the kernel K has compact support, we have* $\mathrm{E}\widehat{m}_n(x) \longrightarrow r(x)f(x)$.

Proof Write

$$\mathrm{E}\widehat{m}_n(x) = \frac{1}{h_n}\mathrm{E}\left(YK\left(\frac{x-X}{h_n}\right)\right) = \frac{1}{h_n}\int r(u)K\left(\frac{x-u}{h_n}\right)f(u)\,du$$

and apply Lemma 3.22 to conclude the proof. □

So, if we assume Y to be nonnegative, the regression function r is also nonnegative, and the conditions of Theorem C.3 are verified, at least for n large enough, and thus the following inclusion holds.

Lemma 3.35 *Assume that Y is a nonnegative random variable and the regression function $r(x) = \mathrm{E}(Y|X=x)$ is continuous. If the kernel K has compact support, we have, for $\varepsilon > 0$ small enough and n large enough, that*

$$\begin{aligned}\left\{\left|\widehat{r}_n(x) - \frac{\mathrm{E}\widehat{m}_n(x)}{\mathrm{E}\widehat{f}_n(x)}\right| > \varepsilon\right\} &\\ \subset \left\{\left|\widehat{m}_n(x) - \mathrm{E}\widehat{m}_n(x)\right| > \frac{\varepsilon}{4}\mathrm{E}\widehat{f}_n(x)\right\}&\\ \cup \left\{\left|\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)\right| > \frac{\varepsilon}{4}\frac{(\mathrm{E}\widehat{f}_n(x))^2}{\mathrm{E}\widehat{m}_n(x)}\right\}&. \end{aligned} \tag{3.31}$$

This lemma allows us to reduce the proof of the consistency of $\widehat{r}_n(x)$ to the convergence to zero of $\widehat{m}_n(x) - \mathrm{E}\widehat{m}_n(x)$ and $\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)$. Naturally, the later convergence has been established in the previous section, so we need to prove the other convergence.

To proceed, we need an extension of assumptions **(D)** and **(A)**, describing an adequate control on an extension of the diagonal decomposition to a higher dimension. Recall that we have denoted by λ^2 be the Lebesgue measure on $\mathbb{R}$. Now let λ_2^2 be the Lebesgue measure on $\mathbb{R}^2 \times \mathbb{R}^2$, that is, the four-dimensional Lebesgue measure. Represent by $\Delta^* = \{(u,v,u,v), (u,v) \in \mathbb{R}^2\}$ the diagonal of $\mathbb{R}^2 \times \mathbb{R}^2$ and introduce the measure λ_2^* defined by $\lambda_2^*\{(u,v,u,v), (u,v) \in A\} = \lambda^2(A)$, where A is a Borel subset of $\mathbb{R}^2$. We consider the following extension of **(D)**:

(D′)

- **(D′.1)** For all $j,k \in \mathbb{N}$, the distribution of $((X_j,Y_j),(X_k,Y_k))$ is the sum of a measure $m_{1,j,k}$ on $\mathbb{R}^2 \times \mathbb{R}^2 \setminus \Delta$ and a measure $m_{2,j,k}$ on Δ such that $m_{1,j,k} \ll \lambda_2^2$ and $m_{2,j,k} \ll \lambda_2^*$.
- **(D′.2)** For all $j,k \in \mathbb{N}$, there exists a bounded version $b_{1,j,k}$ of $\frac{dm_{1,j,k}}{d\lambda_2^2}$.
- **(D′.3)** For all $j,k \in \mathbb{N}$, there exists a bounded and continuous version $b_{2,j,k}$ of $\frac{dm_{2,j,k}}{d\lambda_2^*}$.

What concerns **(A)**, we have the following extension:

(A′)

- **(A′.1)** **(D′)** is satisfied.
- **(A′.2)** $\frac{1}{n}\sum_{j,k=1}^{n} |b_{1,j,k} - f^* \otimes f^*|$ converges uniformly to a function b_1 such that $b_1^*(v_1, v_2)\int_{\mathbb{R}^2} u_1 u_2 b_1(v_1, u_1, v_2, u_2)\,du_1\,du_2$ is a bounded function of (v_1, v_2).
- **(A′.3)** $\frac{1}{n}\sum_{j,k=1}^{n} b_{2,j,k}$ converges uniformly to b_2 such that $b_2^*(v) = \int_{\mathbb{R}} u^2 b_2(v,u)\,du$ is bounded and continuous.

Reproducing the arguments of the proof of Lemma 3.24, it is now a simple matter to prove the following result.

Lemma 3.36 *Let (X_n, Y_n), $n \in \mathbb{N}$, be random vectors such that* **(A′)** *is satisfied. If the kernel K is square integrable, then*

$$\frac{1}{nh_n}\sum_{j,k=1}^{n} \mathrm{Cov}\left(Y_j K\left(\frac{x - X_j}{h_n}\right), Y_k K\left(\frac{x - X_k}{h_n}\right)\right) \longrightarrow b_2^*(x,x)\int K^2(u)\,du$$

uniformly on any compact set.

With this tool in hand, the convergence of $\widehat{m}_n(x) - \mathrm{E}\widehat{m}_n(x)$ is immediate, again by reproducing the corresponding arguments for the proof of the convergence of $\widehat{f}_n(x) - \mathrm{E}\widehat{f}_n(x)$, as in Theorem 3.28.

Theorem 3.37 *Let (X_n, Y_n), $n \in \mathbb{N}$, be associated random vectors. Assume that* **(A′)** *is satisfied, K is of bounded variation such that $K(\alpha x)$, for fixed $x \in \mathbb{R}$, is a decreasing function of $\alpha > 0$, and*

$$h_n \searrow 0, \qquad \sum_{n=1}^{\infty} \frac{1}{n^2 h_{n^2}} < \infty, \qquad \frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1.$$

Then $\widehat{m}_n(x)$ converges almost surely to $r(x) f(x)$ for every $x \in \mathbb{R}$.

Finally, we may state the complete result with the almost sure convergence of the regression estimator (3.30), taking into account Lemma 3.35.

Theorem 3.38 *Let (X_n, Y_n), $n \in \mathbb{N}$, be associated random vectors. Assume that $r(x) = \mathrm{E}(Y|X = x)$ is continuous,* **(A)** *and* **(A′)** *are satisfied, K is of bounded variation such that $K(\alpha x)$, for fixed $x \in \mathbb{R}$, is a decreasing function of $\alpha > 0$, and*

$$h_n \searrow 0, \qquad \sum_{n=1}^{\infty} \frac{1}{n^2 h_{n^2}} < \infty, \qquad \frac{h_{(n+1)^2}}{h_{n^2}} \longrightarrow 1.$$

Then $\widehat{r}_n(x)$ converges almost surely to $r(x)$ for every $x \in \mathbb{R}$.

It is worth noticing that this theorem is an explicit version of the results included in Ferrieux [38, 39] and in Jacob and Oliveira [50–52], although our statement seems somewhat more intricate than the corresponding ones in the above-mentioned references. This is due the fact that [38, 39, 50–52] explore a more complex and enlarged framework, using point processes, which requires a quite condensed form of writing in order to produce readable results. We notice also that [38, 39, 52], which deal with statistical problems based on associated samples of point processes, do not overlap with the results in these two last sections. In fact, as follows from Sect. 1.4, association of point processes is not equivalent to association of random variables or vectors, due to a different order structure on the base space (see Example 1.4).

Chapter 4
Convergence in Distribution

Abstract This chapter addresses central limit theorems, invariance principles and then proceeds to the convergence of empirical processes. The pathway will be to start with versions based on stationary variables and drop this assumption introducing the necessary control on the covariance structure. The techniques will be based on approximations of independent variables relying on a few inequalities established in Chap. 2. Once we have proved the first results, we will find characterizations of convergence rates with respect to the usual supnorm metric between distribution functions. A few applications to statistical estimation problems will be addressed in the final section of this chapter, as done in the previous one.

4.1 Introduction

Central Limit Theorems are at the heart of every probability model, so it is not surprising that this problem was one of the first to be addressed in the literature for associated random variables. In fact, after the early developments, mainly concerned with the dependency structure itself, the first asymptotic result was a Central Limit Theorem and an invariance principle proved in Newman and Wright [72, 71] for strictly associated sequences of random variables. The stationarity assumption was dropped by Cox and Grimmett [25], who introduced the coefficients $u(n)$ (see Definition 2.13) that control the covariance structure of the variables. Naturally, extensions of the classical Berry–Esséen inequalities characterizing convergence rates were addressed in Wood [104], assuming the stationarity of the sequence, and Birkel [14] for general sequences of associated variables. The approach used either for the Central Limit Theorems or for the convergence rate characterization is based on block decompositions of the sums and approximations by independent variables. This step is essentially controlled using Newman's inequality (2.26), thus being a key ingredient for this kind of results. Functional results characterizing the convergence in distribution of the partial sums process or the empirical process have also been proved, but these are deferred to Chap. 5.

In statistical estimation based on associated samples, the first problems involving convergence in distribution that were treated concerned the approximation of distribution functions in Roussas [85] and survival functions in Bagai and Prakasa Rao [4]. The results obtained proved the almost sure consistency and asymptotic

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8_4,

normality for stationary sequences under a polynomial decrease rate of the coefficient $v(n) = \sum_{i=n+1}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_i)$. This is not the usual $u(n)$ coefficient, but its appearance is easily explained noticing that distributions functions involve transformations depending on indicator functions, and recalling inequality (2.25). Roussas [85] even proved a polynomial convergence rate for the estimator assuming, besides a suitable decay rate of $v(n)$, the existence of joint densities and a control of the distance of this with respect to the product density, thus controlling the distance of joint distributions to independence. Extensions of the above asymptotic characterizations addressing associated random fields appeared in Roussas [86, 87], to quantile functions in Cai and Roussas [24] and to multidimensional distribution functions in Azevedo and Oliveira [3] and Henriques and Oliveira [41]. The estimation of density functions was treated in Roussas [88], who proved the consistency and asymptotic normality under the above-mentioned control between the distance of joint densities and product densities. The control of this distance was dropped in Roussas [89, 90], but still assuming the existence of pairwise joint densities. An extension of these results appeared in Oliveira [74], allowing for nonabsolutely continuous joint distributions. Again, one of the key tools to prove the asymptotic normality results for the estimators is Newman's inequality (2.26).

There is a much more general approach to estimation problems followed by Jacob and Oliveira [50–52], Ferrieux [38, 39], Bensaïd and Fabre [6] and Bensaïd and Oliveira [7], where one deals with models characterized by point processes, and the interest is in the estimation of Radon–Nikodym derivatives. These models are shown to be an extension of several classical estimation problems, as explained, for example, in Jacob and Oliveira [50, 51]. Nevertheless, as follows from Sect. 1.4, association of point processes is not the same as association of the random variables that define the point processes. However, some of the techniques used in this point process framework are adaptable to the treatment of associated random variables allowing, for example, for the extension to nonabsolutely continuous joint distributions proved in Oliveira [74].

4.2 Central Limit Theorems

The first Central Limit Theorem for associated variables will be proved assuming the stationarity of the sequence. The requirement on the covariance structure is just that defines a summable series. The technique of proof that will be used again later in a broader context consists of decomposing sums of variables into sums of blocks of variables and treating these as if they were independent. Naturally, we will need some control on the approximation between the sums of the dependent blocks and their independent counterparts. This control is achieved using characteristic functions and is based on inequality (2.26), following the approach used in Newman [68].

Theorem 4.1 *Let X_n, $n \in \mathbb{N}$, be centred, strictly stationary, square-integrable and associated random variables such that*

$$\sigma^2 = \mathrm{Var}(X_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty. \tag{4.1}$$

Then,

$$\frac{1}{\sigma\sqrt{n}} S_n \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Proof Denote the characteristic function of $\frac{1}{\sqrt{n}} S_n$ by $\varphi_n(t)$. We will decompose S_n into the sum of blocks of size $\ell \in \mathbb{N}$. The number of such blocks is $m = [\frac{n}{\ell}]$, the largest integer less than or equal to n/ℓ. Define now the blocks

$$Y_{j,\ell} = \sum_{i=(j-1)\ell+1}^{j\ell} X_i, \quad j = 1, \ldots, m, \quad \text{and} \quad Y_{m+1,\ell} = \sum_{i=m\ell+1}^{n} X_i.$$

The proof will be divided into four steps, proceeding with fixed ℓ in the first three steps and allowing $\ell \longrightarrow +\infty$ in the last one.

Step 1. We start by checking that it is enough to consider n a multiple of ℓ. Using $|e^{it} - e^{is}| \le |t - s|$ for $t, s \in \mathbb{R}$ and the Cauchy inequality, it follows that

$$\begin{aligned}
\left|\varphi_n(t) - \varphi_{m\ell}(t)\right| &\le |t| \mathrm{E}^{1/2}\left(\frac{S_n}{\sqrt{n}} - \frac{S_{m\ell}}{\sqrt{m\ell}}\right)^2 \\
&\le |t|\left(\frac{1}{\sqrt{m\ell}} - \frac{1}{\sqrt{n}}\right)\mathrm{E}^{1/2}\left(\sum_{j=1}^{m\ell} X_j\right)^2 \\
&\quad + \frac{|t|}{\sqrt{m\ell}}\mathrm{E}^{1/2}\left(\sum_{j=m\ell+1}^{n} X_j\right)^2.
\end{aligned}$$

Expanding the squares in the mathematical expectations it easily follows that, given the definition of σ^2,

$$\mathrm{E}\left(\sum_{j=1}^{m\ell} X_j\right)^2 \le \sigma^2 m\ell \quad \text{and} \quad \mathrm{E}\left(\sum_{j=m\ell+1}^{n} X_j\right)^2 \le \sigma^2 \ell.$$

As ℓ is fixed, both $n \longrightarrow +\infty$ and $m \longrightarrow +\infty$, so

$$\left|\varphi_n(t) - \varphi_{m\ell}(t)\right| \le |t|\sigma\left(1 - \frac{\sqrt{m\ell}}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right) \longrightarrow 0. \tag{4.2}$$

Step 2. We now control the difference between the joint distribution of the blocks and what we would find if they were independent. We obviously have $\frac{1}{\sqrt{m\ell}} S_{m\ell} = \frac{1}{\sqrt{m}} \frac{1}{\sqrt{\ell}} Y_{j,\ell}$, and, given that the variables are stationary, the distribution of $Y_{j,\ell}$ coincides with the distribution of S_ℓ, thus the characteristic function of $\frac{1}{\sqrt{\ell}} Y_{j,\ell}$

is $\varphi_\ell(t)$. Moreover, the blocks $Y_{j,\ell}$ are increasing functions of the original variables, so are still associated, and we can apply inequality (2.26) to their characteristic functions to find:

$$\left|\varphi_{m\ell}(t) - \varphi_\ell^m\left(\frac{t}{\sqrt{m}}\right)\right| \leq \frac{t^2}{2m} \sum_{\substack{j,k=1 \\ j\neq k}}^{m} \operatorname{Cov}(Y_{j,\ell}, Y_{k,\ell}).$$

Due to the stationarity, we obviously have that

$$\sum_{\substack{j,k=1 \\ j\neq k}}^{m} \operatorname{Cov}(Y_{j,\ell}, Y_{k,\ell}) = \frac{1}{\ell}\big(\operatorname{Var}(S_{m\ell}) - m \operatorname{Var}(S_\ell)\big).$$

Using again the stationarity, $\operatorname{Var}(S_n) = n \operatorname{Var}(X_1) + 2\sum_{j=2}^{n}(n-1-j)\operatorname{Cov}(X_1, X_j)$, so $\frac{1}{n}\operatorname{Var}(S_n) \longrightarrow \sigma^2$, and finally, as $n \longrightarrow +\infty$,

$$\left|\varphi_{m\ell}(t) - \varphi_\ell^m\left(\frac{t}{\sqrt{m}}\right)\right| \longrightarrow 0. \tag{4.3}$$

Step 3. In this step we assume that the blocks are independent. If we define $\sigma_\ell^2 = \frac{1}{\sqrt{\ell}}\operatorname{Var}(S_\ell)$, then the classical Central Limit Theorem for independent variables implies that, as $n \longrightarrow +\infty$ (which also implies that $m \longrightarrow +\infty$),

$$\left|\varphi_\ell^m\left(\frac{t}{\sqrt{m}}\right) - \exp\left(-\frac{t^2\sigma_\ell^2}{2}\right)\right| \longrightarrow 0. \tag{4.4}$$

Step 4. Using again the inequality

$$\left|\exp\left(-\frac{t^2\sigma^2}{2}\right) - \exp\left(-\frac{t^2\sigma_\ell^2}{2}\right)\right| \leq \frac{t^2}{2}|\sigma - \sigma_\ell|$$

and collecting inequalities (4.2), (4.3) and (4.4), we get that, for each fixed $\ell > 0$,

$$\limsup_{n\to+\infty}\left|\varphi_n(t) - \exp\left(-\frac{t^2\sigma^2}{2}\right)\right| \leq t^2|\sigma - \sigma_\ell|.$$

As the left side of this inequality does not depend on ℓ, we may allow $\ell \to +\infty$ to conclude that

$$\limsup_{n\to+\infty}\left|\varphi_n(t) - \exp\left(-\frac{t^2\sigma^2}{2}\right)\right| = 0,$$

so the theorem is proved. □

Remark 4.2 Notice that association is crucial in the way we control the approximation Steps 1 and 2 in the previous proof. In fact, it is association, as it implies that covariances are always nonnegative, that allows for the conclusion (4.2). For the deviance from independence, again it is association that allows the conclusion, as the argument used relies on Newman's inequality (2.26) (in fact, for this part of the proof, it would be enough to assume the variables to be LPQD).

Remark 4.3 The main argument in the proof above is the approximation of the distribution of S_n by the product distribution of the blocks defined in the proof. This sort of argument has been used before, sometimes just by including this approximation in the assumptions, as done, for example, in Lemma 3 in Birkel [12]. Of course, such a result did not include the association in its assumptions, otherwise the statement would have been quite similar to Theorem 4.1.

Taking into account the previous remark, it is an easy matter to prove a general Central Limit Theorem, repeating the arguments above, with the obvious adaptations to control the approximation to the independent coupling variables.

Theorem 4.4 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. For each $n \in \mathbb{N}$, let $\ell_n \in \mathbb{N}$ and $m_n = [\frac{n}{\ell_n}]$. Define, for $j = 1, \ldots, m_n$ $Y_{j,\ell_n} = \sum_{i=(j-1)\ell_n+1}^{j\ell_n} X_i$ and $Y_{m_n+1,\ell_n} = \sum_{m_n\ell_n+1}^{n} X_i$. Assume that $m_n \longrightarrow +\infty$,*

$$\frac{1}{s_n^2}\sum_{j=1}^{m_n} \mathrm{Var}(Y_{j,\ell_n}) \longrightarrow 1, \tag{4.5}$$

$$\left| \mathrm{E}\exp\left(\frac{iu}{s_n} S_n\right) - \prod_{j=1}^{m_n} \mathrm{E}\exp\left(\frac{iu}{s_n} Y_{j,\ell_n}\right)\right| \longrightarrow 0, \quad u \in \mathbb{R}, \tag{4.6}$$

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2}\sum_{j=1}^{m_n} \int_{\{|Y_{j,\ell_n}| \geq \varepsilon s_n\}} Y_{j,\ell_n}^2 \, d\mathbf{P} \longrightarrow 0. \tag{4.7}$$

Then

$$\frac{1}{s_n} S_n \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

As mentioned above, the proof of the previous Central Limit Theorem is based on (2.26), with (4.5) playing the role of (4.1), (4.6) allowing for the argument corresponding to Step 1 in the proof of Theorem 4.1 and (4.7), a Lindeberg condition, implying the Central Limit Theorem corresponding to Step 3 of the proof of Theorem 4.1.

We have proved an extension to possibly nonmonotone transformations of the random variables in Theorem 2.40. So, it possible to have an extended version of Theorem 4.1, by adapting the proof of Theorem 2.40 in an obvious way.

Theorem 4.5 *Let Y_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables. Define, for each $n \in \mathbb{N}$, $X_n = f(Y_n, Y_{n+1}, \ldots)$ and $X_n^* = g(Y_n, Y_{n+1}, \ldots)$ where $f \preceq g$ (see Definition 2.4). Assume that*

$$\sum_{k=2}^{\infty} \mathrm{Cov}\big(X_1^*, X_k^*\big) < \infty. \tag{4.8}$$

Then,

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mathrm{E}S_n) \xrightarrow{d} Z \sim \mathcal{N}(0,1),$$

where σ^2 is defined by (4.1).

It is possible to be somewhat more precise about the convergence assumptions and the characterization of the asymptotic variance σ^2 in the previous statements. For this purpose, let us define $H^*_{j,k}$, the functions corresponding to the usual $H_{j,k}$ but with respect to the Y_n variables:

$$\begin{aligned} H^*_{j,k}(x,y) &= \mathbf{P}(Y_j > x, Y_k > y) - \mathbf{P}(Y_j > x)\mathbf{P}(Y_k > y) \\ &= \mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(Y_j), \mathbb{I}_{(y,+\infty)}(Y_k)\big) \\ &= \mathrm{Cov}\big(\mathbb{I}_{(-\infty,x]}(Y_j), \mathbb{I}_{(-\infty,y]}(Y_k)\big). \end{aligned} \tag{4.9}$$

Assume the Y_n are strictly stationary and define

$$\Gamma(x,y) = H_{1,1}(x,y) + \sum_{k=2}^{\infty}\big(H_{1,k}(x,y) + H_{1,k}(y,x)\big). \tag{4.10}$$

If we assume the variables Y_n to be associated, as is the case where the initial X_n are associated, it follows that $0 \le H_1 \le \Gamma \le +\infty$. Define further,

$$H_{(n)}(x,y) = \frac{1}{n}\sum_{j,k=1}^{n} \mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(Y_j), \mathbb{I}_{(y,+\infty)}(Y_k)\big).$$

Then, given a real-valued function g,

$$\begin{aligned} &\int\int g(x)H_{(n)}(x,y)g(y)\,dx\,dy \\ &\quad = \frac{1}{n}\sum_{j,k=1}^{n}\int\int g(x)H_{j,k}(x,y)g(y)\,dx\,dy = \frac{1}{n}\mathrm{Var}\left(\sum_{j=1}^{n} g(Y_j)\right) \ge 0. \end{aligned}$$

Taking into account the strict stationarity of the random variables, it still follows that

$$\begin{aligned} H_{(n)}(x,y) = \mathrm{Cov}\big(&\mathbb{I}_{(x,+\infty)}(Y_1), \mathbb{I}_{(y,+\infty)}(Y_1)\big) \\ &+ \frac{1}{n}\sum_{j=2}^{n-1}(n-j)\big(\mathrm{Cov}\big(\mathbb{I}_{(x,+\infty)}(Y_1), \mathbb{I}_{(y,+\infty)}(Y_j)\big) \\ &+ \mathrm{Cov}\big(\mathbb{I}_{(y,+\infty)}(Y_1), \mathbb{I}_{(x,+\infty)}(Y_j)\big)\big). \end{aligned}$$

Thus, taking into account the nonnegativity of each term, we have

$$\lim_{n\to+\infty} H_{(n)}(x,y) = H_{1,1}(x,y) + \sum_{j=2}^{\infty}\big(H_{1,j}(x,y) + H_{1,j}(y,x)\big),$$

so, if the integrals are finite, we have that, for any real-valued function g,

$$\int\int g(x)\Gamma(x,y)g(y)\,dx\,dy = \lim_{n\to+\infty}\int\int g(x)H_{(n)}(x,y)g(y)\,dx\,dy \ge 0.$$

Notice further that, if we take $g \equiv 1$, we may rewrite the asymptotic variance in Theorem 4.1 as $\sigma^2 = \int\int g(x)\Gamma(x,y)g(y)\,dx\,dy$.

The inequality above expresses one fundamental property that should be required if Γ is to be a covariance operator, namely the semidefinite positiveness. Of course, the above integral only makes sense for g such that $g(x)H_{(n)}(x,y)g(y)$ and $g(x)\Gamma(x,y)g(y)$ are integrable. To simplify referring to this integrability property, given a real-valued symmetric function Ψ on $\mathbb{R}^2$, we denote

$$D_\Psi = \bigl\{g : g(x)\Psi(x,y)g(y) \text{ is integrable}\bigr\}.$$

Theorem 4.6 *Let Y_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables, and $X_n = F(Y_n)$, where F is an absolutely continuous function. Put $X_n^* = F^*(Y_n)$ where $F^*(t) = \int_{-\infty}^t |F'(u)|\,du$. Assume that, for Γ defined by (4.10), $F' \in D_\Gamma$. Then, X_1^* is square integrable, and*

$$\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n (X_j - \mathrm{E}X_j) \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Proof Using integration by parts, it follows easily that, for absolutely continuous functions g_1 and g_2,

$$\mathrm{Cov}\bigl(g_1(Y_1), g_2(Y_j)\bigr) = \int\int g_1'(x)H_{1,j}(x,y)g_2'(y)\,dx\,dy \tag{4.11}$$

whenever this covariance exists. Rewriting (4.11) for $j = 1$ and $g_1 = g_2 = F^*$, we get that X_1^* is square-integrable. As, due to the association of the variables, for every $j \in \mathbb{N}$, $0 \le H_{1,j} \le \Gamma$, it is obvious that $F' \in D_{H_{1,j}}$. The proof is now completed referring to Theorem 4.5: by choosing $g_1 = g_2 = F^*$ and summing for $j = 2, 3, \ldots$, (4.8) follows from (4.11), while the characterization of the asymptotic variance σ^2 follows using again (4.11) with $g_1 = g_2 = F$. □

Remark 4.7 Assume, in the previous result, that $\sum_j \mathrm{Cov}(Y_1, Y_j) < +\infty$. Then, if F' is bounded, it follows that $F' \in D_\Gamma$:

$$\begin{aligned}\int\int F'(x)\Gamma(x,y)F'(y)\,dx\,dy &\le \|F'\|_\infty \int\int \Gamma(x,y)\,dx\,dy \\ &= \|F'\|_\infty \sum_{j=2}^\infty \mathrm{Cov}(Y_1, Y_j) < +\infty.\end{aligned}$$

To deal with nonstationary variables, let us recall the definition of

$$u(n) = \sup_{k\in\mathbb{N}} \sum_{j:|j-k|\ge n} \mathrm{Cov}(X_j, X_k).$$

Theorem 4.8 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Assume that*

$$u(n) \longrightarrow 0, \qquad u(1) < \infty, \tag{4.12}$$

$$\inf_{n \in \mathbb{N}} \frac{1}{n} s_n^2 > 0, \tag{4.13}$$

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{j=1}^{n} \int_{\{|X_j| > \varepsilon s_n\}} X_j^2 \, d\mathbf{P} \longrightarrow 0. \tag{4.14}$$

Then

$$\frac{1}{s_n} S_n \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Proof For the proof, we will check that the assumptions of Theorem 4.4 are verified. For this purpose, reproduce the decomposition of S_n into blocks: for each $n \in \mathbb{N}$, let $\ell_n \in \mathbb{N}$, $m_n = [\frac{n}{\ell_n}]$ and define $Y_{j,\ell_n} = \sum_{i=(j-1)\ell_n+1}^{j\ell_n} X_i$, $j = 1, \ldots, m_n$, and $Y_{m_n+1,\ell_n} = \sum_{m_n\ell_n+1}^{n} X_i$. We first choose a sequence ℓ_n such that, for every $\varepsilon > 0$,

$$\frac{\ell_n^2}{s_n^2} \sum_{j=1}^{n} \int_{\{|X_j| > \varepsilon s_n/\ell_n\}} X_j^2 \, d\mathbf{P} \longrightarrow 0. \tag{4.15}$$

Such a sequence does exist, as follows from (4.14). In fact, put $n_1 = 1$ and define, for $k \geq 2$, $n_k \in \mathbb{N}$ such that $2n_k \leq n_{k+1}$ and

$$\frac{1}{s_n^2} \sum_{j=1}^{n} \int_{\{|X_j| > \varepsilon s_n/k^2\}} X_j^2 \, d\mathbf{P} \leq \frac{1}{k^3}, \qquad n \geq n_k.$$

Take now $\ell_n = k$ if $n_k \leq n < n_{k+1}$. So, we have that $\ell_n \longrightarrow +\infty$ and also $m_n \longrightarrow +\infty$. As the random variables are associated, we may repeat the arguments of the proof of Theorem 4.1 to find that

$$\left| \mathrm{E} \exp\left(\frac{iu}{s_n} S_n\right) - \prod_{j=1}^{m_n} \mathrm{E} \exp\left(\frac{iu}{s_n} Y_{j,\ell_n}\right) \right| \leq t^2 \left| 1 - \frac{1}{s_n^2} \sum_{j=1}^{m_n} \mathrm{Var}(Y_{j,\ell_n}) \right|.$$

So, to complete the proof, it is enough to verify that (4.5) and (4.7) hold. Taking into account that $S_n = Y_{1,\ell_n} + \cdots + Y_{m_n,\ell_n}$, we have

$$1 - \frac{1}{s_n^2} \sum_{j=1}^{m_n} \mathrm{Var}(Y_{j,\ell_n}) = \frac{2}{s_n^2} \sum_{j=1}^{m_n-1} \sum_{k=j+1}^{m_n} \mathrm{Cov}(Y_{j,\ell_n}, Y_{k,\ell_n}) \geq 0.$$

As, due to the association of variables, all the covariances are nonnegative, it follows that

$$1 - \frac{1}{s_n^2} \sum_{j=1}^{m_n} \mathrm{Var}(Y_{j,\ell_n}) \leq \frac{2m_n}{s_n^2} \sum_{j=1}^{\ell_n} u(j) \leq \frac{2n}{s_n^2} \frac{1}{\ell_n} \sum_{j=1}^{\ell_n} u(j) \longrightarrow 0$$

as $\ell_n \longrightarrow +\infty$, by Cesàro convergence (see Lemma C.6). We have thus verified that (4.5) is satisfied. We now verify that (4.7) also holds. We have, for each $j = 1, \ldots, m_n$, $Y_{j,\ell_n}^2 \le \ell_n \sum_{i=(j-1)\ell_n+1}^{j\ell_n} X_i^2$, so,

$$\begin{aligned}
&\int_{\{|Y_{j,\ell_n}|>\varepsilon s_n\}} Y_{j,\ell_n}^2 \, d\mathbf{P} \\
&\quad \le \ell_n \sum_{i=(j-1)\ell_n+1}^{j\ell_n} \int_{\{\sum_k X_k^2 > \varepsilon^2 s_n^2/\ell_n\}} X_i^2 \, d\mathbf{P} \\
&\quad \le \ell_n \sum_{i=(j-1)\ell_n+1}^{j\ell_n} \sum_{k=(j-1)\ell_n+1}^{j\ell_n} \int_{\{X_k^2 > \varepsilon^2 s_n^2/\ell_n^2\}} X_i^2 \, d\mathbf{P} \\
&\quad \le \ell_n(\ell_n - 1) \sum_{i=(j-1)\ell_n+1}^{j\ell_n} \int_{\{X_j^2 > \varepsilon^2 s_n^2/\ell_n^2\}} X_i^2 \, d\mathbf{P}.
\end{aligned}$$

It follows then that

$$\frac{1}{s_n^2} \sum_{j=1}^{m_n} \int_{\{|Y_{j,\ell_n}|>\varepsilon s_n\}} Y_{j,\ell_n}^2 \, d\mathbf{P} \le \frac{\ell_n^2}{s_n^2} \sum_{j=1}^{n} \int_{\{|X_j|>\varepsilon s_n/\ell_n\}} X_j^2 \, d\mathbf{P} \longrightarrow 0$$

taking into account (4.15), so (4.7) also holds. □

Remark 4.9 Assumptions (4.13) and (4.14) are the minimum one could expect, even for independent random variables without the stationarity. Thus, in order to prove the Central Limit Theorem, all that is required is that the covariances decrease fast enough. Notice further, that for strictly stationary random variables, $u(n) = 2\sum_{j=n}^{\infty} \operatorname{Cov}(X_1, X_n)$, thus (4.12) is equivalent to the convergence of $\sum_n \operatorname{Cov}(X_1, X_n)$, which is implicitly required in (4.1).

4.3 Convergence Rates

In this section we prove some extensions to associated random variables of the classical Berry–Esséen bounds for the distance between the asymptotic Gaussian distribution function and the distribution function of $\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$. We will start by proving a bound with respect to the usual supnorm, assuming only the existence of third-order moments. A first result in this direction, assuming the strict stationarity of the variables, was proved by Wood [104]. Is well known that, for the supnorm distance, the Berry–Esséen inequality provides a convergence rate of order $n^{-1/2}$, assuming the existence of moments of order 2. If we assume higher-order moments, this convergence rate can be improved, as expected from an approach based on Taylor expansions. The bounds that are proved for associated random variables can only provide much slower convergence rates. This seems to depend on the approach, still based on Taylor expansions now complemented with the block decomposition of $X_1 + \cdots + X_n$ as used in the previous section.

The proof of the Berry–Esséen bound for associated variables follows the same path of arguments as used for proving Central Limit Theorems: we decompose the partial sums into blocks and couple these with independent variables keeping the distributions of these blocks. The essential of the proof is then controlling the approximation to independence. So, in order to prepare for the proof under association, we prove a convenient version of the Berry–Esséen bound for independent variables. Denote in the sequel by Φ_a the distribution function of the Gaussian distribution with mean 0 and variance $a > 0$.

Theorem 4.10 *Let $X_1, \ldots, X_n$ be independent centred random variables with finite third-order absolute moments $\beta_j = \mathrm{E}|X_j|^3$. Assume that*

$$\inf_{n\in\mathbb{N}} \frac{s_n^2}{n} \geq c_0. \tag{4.16}$$

Then

$$\sup_{x\in\mathbb{R}} \left|\mathbf{P}(S_n \leq \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \leq \frac{24\sum_j \beta_j}{\pi s_n^2 n^{1/2}} + \frac{96\sum_j \beta_j}{c_0\pi\sqrt{2\pi}\, s_n n}. \tag{4.17}$$

Proof The proof consists of deriving convenient bounds for the difference between the characteristic functions of the distributions involved, that is, bounding $|\varphi_{S_n}(tn^{-1/2}) - \exp(-\frac{t^2 s_n^2}{2n})|$, and then applying (A.1). Denote in the sequel $v_j^2 = \mathrm{E}X_j^2$, so that $s_n^2 = v_1^2 + \cdots + v_n^2$.

Assume first that

$$\frac{n^{1/2}}{2(\sum_j \beta_j)^{1/3}} \leq |t| \leq \frac{c_0 n^{3/2}}{4\sum_j \beta_j}. \tag{4.18}$$

For each $j \in \mathbb{N}$, consider random variables Y_j independent of X_j and with the same distribution as X_j. So, $\mathrm{E}(X_j - Y_j) = 0$, $\mathrm{Var}(X_j - Y_j) = 2\mathrm{E}X_j^2 = 2v_j^2$ and $\mathrm{E}|X_j - Y_j|^3 \leq \mathrm{E}(|X_j| + |Y_j|)^3 \leq 2\mathrm{E}|X_j|^3 + 6\mathrm{E}|X_j|\mathrm{E}Y_j^2 \leq 8\beta_j$. The characteristic function of $X_j - Y_j$ is $|\varphi_{X_j}|^2$. Now, a Taylor expansion gives, for some $\theta \in (-1, 1)$,

$$\begin{aligned}\left|\varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right)\right|^2 &= 1 - \frac{v_j^2 t^2}{n} + \theta\frac{t^3\mathrm{E}(X_j - Y_j)^3}{6n^{3/2}} \\ &\leq 1 - \frac{v_j^2 t^2}{n} + \theta\frac{4|t|^3\beta_j}{3n^{3/2}} \leq \exp\left(-\frac{v_j^2 t^2}{n} + \frac{4|t|^3\beta_j}{3n^{3/2}}\right).\end{aligned}$$

Taking account of the upper bound for t and (4.16), it follows that

$$\begin{aligned}\left|\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right)\right|^2 &= \prod_{j=1}^n \left|\varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right)\right|^2 \leq \exp\left(-\frac{t^2}{n}\sum_j v_j^2 + \frac{4|t|^3}{3n^{3/2}}\sum_j \beta_j\right) \\ &\leq \exp\left(-\frac{t^2 s_n^2}{n} + \frac{t^2 s_n^2}{3n}\right) = \exp\left(-\frac{2t^2 s_n^2}{3n}\right).\end{aligned}$$

Finally, given (4.18), we have

$$\left|\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{t^2 s_n^2}{2n}\right)\right| \le \exp\left(-\frac{2t^2 s_n^2}{3n}\right) + \exp\left(-\frac{t^2 s_n^2}{2n}\right) \le 2\exp\left(-\frac{t^2 s_n^2}{3n}\right).$$

Assume now that $|t| \le \frac{n^{1/2}}{2(\sum_j \beta_j)^{1/3}}$ and write

$$\varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) - 1 = -\frac{v_j^2 t^2}{2n} + \theta \frac{t^3 \beta_j}{6n^{3/2}}$$

for some $\theta \in (-1, 1)$. It follows, using Hölder's inequality and the fact that the β_j are nonnegative, that

$$\left|\varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) - 1\right| \le \frac{v_j^2 t^2}{2n} + \frac{|t|^3 \beta_j}{6n^{3/2}} \le \frac{v_j^2}{8(\sum_j \beta_j)^{2/3}} + \frac{\beta_j}{48\sum_j \beta_j} \le \frac{7}{48}.$$

Hence, in the interval $|t| \le \frac{n^{1/2}}{2(\sum_j \beta_j)^{1/3}}$, the characteristic function $\varphi_{X_j}(tn^{-1/2})$ is bounded away from zero. On the other hand,

$$\begin{aligned}
\left|\varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) - 1\right|^2 &\le 2\frac{v_j^4 t^4}{4n^2} + 2\frac{t^6 \beta_j^2}{36n^3} \\
&\le |t|^3 \beta_j \left(\frac{v_j^4}{4n^{3/2}\beta_j(\sum_j \beta_j)^{1/3}} + \frac{\beta_j}{144n^{3/2}\sum_j \beta_j}\right) \\
&\le |t|^3 \beta_j \frac{37}{144n^{3/2}},
\end{aligned}$$

from which follows that

$$\log \varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) = -\frac{v_j^2 t^2}{2n} + \theta \frac{|t|^3 \beta_j}{6n^{3/2}} + \gamma \frac{37|t|^3 \beta_j}{144n^{3/2}} = -\frac{v_j^2 t^2}{2n} + \eta \frac{|t|^3 \beta_j}{2n^{3/2}},$$

where $\gamma \in (-1, 1)$ and $\eta = \frac{\theta}{3} + \frac{37\gamma}{72} \in (-1, 1)$. Thus, we find the expansion

$$\log \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \sum_{j=1}^{n} \log \varphi_{X_j}\left(\frac{t}{\sqrt{n}}\right) = -\frac{t^2}{2n}\sum_j v_j^2 + \eta \frac{|t|^3}{2n^{3/2}}\sum_j \beta_j,$$

from which it follows, recalling that $s_n^2 = \sum_j v_j^2$, that

$$\begin{aligned}
&\left|\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) - \exp\left(-\frac{t^2 s_n^2}{2n}\right)\right| \\
&\quad\le \exp\left(-\frac{t^2 s_n^2}{2n}\right)\left|\exp\left(\eta \frac{|t|^3}{2n^{3/2}}\sum_j \beta_j\right) - 1\right| \\
&\quad\le \exp\left(-\frac{t^2 s_n^2}{2n}\right)\frac{|t|^3}{2n^{3/2}}\sum_j \beta_j \exp\left(\frac{|t|^3}{2n^{3/2}}\sum_j \beta_j\right) \\
&\quad\le \frac{e^{1/16}}{2}\exp\left(-\frac{t^2 s_n^2}{2n}\right)\frac{|t|^3}{n^{3/2}}\sum_j \beta_j.
\end{aligned}$$

So, putting together the two upper bounds derived in each interval for t, we have that

$$\left|\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right)-\exp\left(-\frac{t^2s_n^2}{2n}\right)\right| \leq 16\frac{|t|^3}{n^{3/2}}\sum_j \beta_j \exp\left(-\frac{t^2s_n^2}{3n}\right)$$

for every $|t| \leq \frac{c_0 n^{3/2}}{4\sum_j \beta_j}$. To finish the proof, just use (A.1) with $T = \frac{c_0 n^{3/2}}{4\sum_j \beta_j}$ to find

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \leq \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \\
&\quad\leq \frac{16\sum_j \beta_j}{\pi n^{3/2}}\int_{-T}^{T} t^2 \exp\left(-\frac{t^2s_n^2}{3n}\right)dt + \frac{96\sqrt{n}\sum_j \beta_j}{c_0\pi\sqrt{2\pi}\sigma_n n^{3/2}} \\
&\quad\leq \frac{24\sum_j \beta_j}{\pi s_n^2 n^{1/2}} + \frac{96\sum_j \beta_j}{c_0\pi\sqrt{2\pi}s_n n}.
\end{aligned}$$

□

Notice that, for the case of independent and identically distributed variables, (4.17) gives the usual $n^{-1/2}$ rate.

Corollary 4.11 *Let $X_1, \ldots, X_n$ be independent and identically distributed centred random variables with finite third-order absolute moments $\beta = \mathrm{E}|X_1|^3$. Then, with $s_1^2 = \mathrm{E}X_1^2$,*

$$\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \leq \sqrt{n}x) - \Phi_{s_1^2}(x)\right| \leq \frac{24\beta}{\pi s_1^2 n^{1/2}} + \frac{96\beta}{\pi\sqrt{2\pi}s_1^3 n^{1/2}}. \tag{4.19}$$

Based on the previous bounds and inequality (2.26), we may get a convergence rate for the Central Limit Theorem for associated variables.

Theorem 4.12 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables. For each $\ell \in \mathbb{N}$, let m be the largest integer less than or equal to n/ℓ, and define $Y_{j,\ell} = \frac{1}{\sqrt{\ell}}\sum_{i=(j-1)\ell+1}^{j\ell} X_i$, $j = 1, \ldots, m$, $\sigma_{j,\ell}^2 = \mathrm{E}Y_{j,\ell}^2$ and $\tau_{j,\ell} = \mathrm{E}|Y_{j,\ell}|^3$. Assume that*

$$\inf_{m\in\mathbb{N}} \frac{1}{m}\sum_{j=1}^{m} \sigma_{j,\ell}^2 \geq c_0 > 0. \tag{4.20}$$

Then, for $n = m \times \ell$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \leq \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \\
&\quad\leq \frac{c_0^2 m^3}{(\sum_j \tau_{j,\ell})^2}\left|\frac{s_n^2}{n} - \frac{1}{m}\sum_j \sigma_{j,\ell}^2\right| \\
&\qquad + \frac{24\sum_j \tau_{j,\ell}}{\pi m^{1/2}\sum_j \sigma_{j,\ell}^2} + \frac{96\sum_j \tau_{j,\ell}}{c_0\pi\sqrt{2\pi}m(\sum_j \sigma_{j,\ell}^2)^{1/2}}.
\end{aligned} \tag{4.21}$$

Proof Using (A.1), we have

$$\begin{aligned}&\left|\mathbf{P}(S_n \le \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \\ &\quad \le \frac{1}{\pi}\int_{-T}^{T}\frac{1}{|t|}\left|\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) - e^{-t^2 s_n^2/(2n)}\right| dt + \frac{24\sqrt{n}}{\pi\sqrt{2\pi}s_n T}.\end{aligned}$$

Notice that $\frac{1}{\sqrt{n}}S_n = \frac{1}{\sqrt{m}}\sum_{j=1}^m Y_{j,\ell}$, so the integral may be decomposed into the sum

$$\begin{aligned}I_1 + I_2 + I_3 &= \int_{-T}^{T}\frac{1}{|t|}\left|\mathrm{E}\exp\left(i\frac{t}{\sqrt{m}}\sum_j Y_{j,\ell}\right) - \prod_{j=1}^m \mathrm{E}\exp\left(i\frac{t}{\sqrt{m}}Y_{j,\ell}\right)\right| dt \\ &\quad + \int_{-T}^{T}\frac{1}{|t|}\left|\prod_{j=1}^m \mathrm{E}\exp\left(i\frac{t}{\sqrt{m}}Y_{j,\ell}\right) - \exp\left(-\frac{t^2}{2m}\sum_j \sigma_{j,\ell}^2\right)\right| dt \\ &\quad + \int_{-T}^{T}\frac{1}{|t|}\left|\exp\left(-\frac{t^2}{2m}\sum_j \sigma_{j,\ell}^2\right) - e^{-t^2 s_n^2/(2n)}\right| dt.\end{aligned}$$

The third integral is bounded by using the inequality $|e^{-t} - e^{-s}| \le |t-s|$:

$$I_3 \le \frac{1}{2}\left|\frac{s_n^2}{n} - \frac{1}{m}\sum_j \sigma_{j,\ell}^2\right| \int_{-T}^{T}|t|\,dt = \frac{T^2}{2}\left|\frac{s_n^2}{n} - \frac{1}{m}\sum_j \sigma_{j,\ell}^2\right|.$$

As the random variables are associated, the integral I_1 is bounded using (2.26):

$$I_1 \le \frac{1}{2m}\sum_{\substack{j,k=1\\ j\ne k}}^m \mathrm{Cov}(Y_{j,\ell}, Y_{k,\ell})\int_{-T}^{T}|t|\,dt = \frac{T^2}{2}\left|\frac{s_n^2}{n} - \frac{1}{m}\sum_j \sigma_{j,\ell}^2\right|,$$

where, according to the proof of Theorem 4.10, we may choose $T = \frac{c_0 m^{3/2}}{4\sum_j \tau_{j,\ell}}$. Finally, to bound I_2, use (4.17) to find

$$I_2 \le \frac{24\sum_j \tau_{j,\ell}}{\pi m^{1/2}\sum_j \sigma_{j,\ell}^2} + \frac{96\sum_j \tau_{j,\ell}}{c_0\pi\sqrt{2\pi}m(\sum_j \sigma_{j,\ell}^2)^{1/2}}.$$

Inequality (4.21) now follows immediately by summing up these upper bounds. □

It is straightforward to write a version of (4.21) assuming the strict stationarity of the random variables. The inequality that follows is essentially the same as derived in Theorem 1 in Wood [104], although our constants are not the same. But this is due to a different method used by Wood [104] to control the upper bound for the result corresponding to (4.17) for independent random variables. In fact, in Wood [104] the stationarity was assumed from the beginning, allowing for some further simplification.

Corollary 4.13 *Let X_n, $n \in \mathbb{N}$, be centred, strictly stationary and associated random variables. For each $k \in \mathbb{N}$, let $\sigma_k^2 = \frac{1}{k}\mathrm{E}S_k^2$ and $\tau_k = \frac{1}{k^{3/2}}\mathrm{E}|S_k|^3$. Then, for $n = m \times k$, we have that*

$$\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \le \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \\ \le \frac{\sigma_k^4 m}{\tau_k^2}\left(\frac{s_n^2}{n} - \sigma_k^2\right) + \frac{24\tau_k}{\pi m\sigma_k^2} + \frac{96\tau_k}{\pi\sqrt{2\pi}m^{1/2}\sigma_k^3}. \tag{4.22}$$

This upper bound allows for a simple identification of the convergent rate that follows from this Berry–Esséen bound for associated variables. The rate, as could be expected, is slower that the $n^{-1/2}$ rate for independent variables.

Corollary 4.14 *Let X_n, $n \in \mathbb{N}$, be centred, strictly stationary and associated random variables such that* $\mathrm{E}|X_1|^r < \infty$ *for some* $r > 2$. *For each* $k \in \mathbb{N}$, *let* $\sigma_k^2 = \frac{1}{k}\mathrm{E}S_k^2$ *and* $\tau_k = \frac{1}{k^{3/2}}\mathrm{E}|S_k|^3$. *Assume that* $u(0) < \infty$ *and* $u(n) < Cn^{-\theta}$ *for some* $\theta > 2$. *Then, there exists a positive constant* $K > 0$, *independent from* n, *such that*

$$\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \le \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \le Kn^{-1/5}. \tag{4.23}$$

Proof It is obvious that, due to the association of the random variables, $\sigma_k^2 \le u(0) < \infty$. Put $\sigma^2 = \mathrm{E}X_1^2 + 2\sum_{i=2}^{\infty}\mathrm{Cov}(X_1, X_i)$. Then, we have that

$$\begin{aligned}\sigma^2 - \frac{s_n^2}{n} &= 2\sum_{i=2}^{\infty}\mathrm{Cov}(X_1, X_i) - \frac{2}{n}\sum_{i=2}^{n}\mathrm{Cov}(n-i+1)\,\mathrm{Cov}(X_1, X_i) \\ &= 2\sum_{i=n+1}^{\infty}\mathrm{Cov}(X_1, X_i) + \frac{2}{n}\sum_{i=2}^{n}\mathrm{Cov}(i-1)\,\mathrm{Cov}(X_1, X_i) \\ &\le 2u(n) + \frac{2}{n}\sum_{i=1}^{\infty}u(i) \le \frac{C_1}{n}\end{aligned}$$

for some $C_1 > 0$, independent of n. Recall that we have $\sigma_k^2 = \frac{1}{k}s_k^2$, so, the same bound holds for $\sigma^2 - \sigma_k^2$. Thus (4.22) rewrites as

$$\sup_{x\in\mathbb{R}}\left|\mathbf{P}(S_n \le \sqrt{n}x) - \Phi_{n^{-1}s_n^2}(x)\right| \\ \le \frac{\sigma_k^4 m}{\tau_k^2}\left(\sigma - \sigma_k^2 - \frac{C_1}{n}\right) + \frac{24\tau_k}{\pi m\sigma_k^2} + \frac{96\tau_k}{\pi\sqrt{2\pi}m^{1/2}\sigma_k^3}.$$

It follows from Corollary 2.21 that both σ_k^2 and τ_k are bounded, so choosing $k = [n^{3/5}]$ and $m = [n^{2/5}]$ concludes the proof. □

4.4 A Law of Iterated Logarithm

The upper bounds established in the previous section allow us to prove results that actually characterize convergence rates for the Strong Law of Large Numbers in a more precise way than the results obtained in the final part of Sect. 3.2, recovering now the same convergence rate as for independent random variables. In fact, using the approximations to the distribution function of a Gaussian variable described in Corollary 4.14, it is possible to prove a Law of Iterated Logarithm (LIL). The first results in this direction were obtained by Dabrowski [27], Dabrowski and Dehling [28], even proving functional versions of the LIL, and Yu [110]. Their method relied on a version of Corollary 4.13 proved by Wood [104] and some control on the upper bound of Newman's inequality (2.26), together with a characterization of the set of limit points due to Berkes [8]. We will follow here a later approach to this problem of Li and Wang [59] that uses more direct arguments.

We need a technical lemma that gives us the work around to using an exponential inequality that we do not have in a sufficiently strong form for associated variables.

Lemma 4.15 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable, strictly stationary and associated random variables such that*

$$\mathrm{E}|X_1|^p < \infty \quad \textit{for some } p > 2, \tag{4.24}$$

$$u(n) \le C n^{-\theta} \quad \textit{for some } C > 0 \textit{ and } \theta > 2, \tag{4.25}$$

$$\sigma^2 = \mathrm{E}X_1^2 + 2\sum_{k=2}^{\infty} \mathrm{Cov}(X_1, X_k) < \infty. \tag{4.26}$$

Let c_n, $n \in \mathbb{N}$, be a nondecreasing sequence of positive numbers, and n_k, $k \in \mathbb{N}$, a nondecreasing sequence of positive integers such that $\sum_{k=1}^{\infty} n_k^{-1/5} < \infty$. The following are equivalent:

(a) $\sum_{k=1}^{\infty} \mathbf{P}(S_{n_k} > c_{n_k}\sqrt{n_k}\sigma) < \infty$.
(b) $\sum_{k=1}^{\infty} \mathbf{P}(|S_{n_k}| > c_{n_k}\sqrt{n_k}\sigma) < \infty$.
(c) $\sum_{k=1}^{\infty} \frac{1}{c_{n_k}} \exp(-\frac{1}{2}c_{n_k}^2) < \infty$.

Proof Notice that, taking into account Theorem 4.1, we get that $\frac{1}{\sigma\sqrt{n}}S_n$ converges weakly to a standard Gaussian variable, so, denoting by Φ the distribution function of a standard Gaussian distribution, it follows from Corollary 4.14 that

$$\left|\mathbf{P}(S_{n_k} \le c_{n_k}\sqrt{n_k}\sigma) - \Phi(c_{n_k})\right| \le K n_k^{-1/5}.$$

Thus, given the assumption on the sequence n_k,

$$\sum_{k=1}^{\infty} \mathbf{P}(S_{n_k} > c_{n_k}\sqrt{n_k}\sigma) < \infty \quad \Leftrightarrow \quad \sum_{k=1}^{\infty} \bigl(1 - \Phi(c_{n_k})\bigr) < \infty.$$

Now, taking into account Lemma A.4, we get that this last series converges or diverges as does the series

$$\sum_{k=1}^{\infty} \frac{1}{c_{n_k}} \Phi'(c_n n_k) = \frac{1}{\sqrt{2\pi}} \sum_{k=}^{\infty} \frac{1}{c_{n_k}} \exp\left(-\frac{c_{n_k}^2}{2}\right). \qquad \square$$

Theorem 4.16 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable, strictly stationary and associated random variables such that* (4.24), (4.25) *and* (4.26) *are verified. Then, with probability* 1,

$$\limsup_{n\to+\infty} \frac{S_n}{\sqrt{2\sigma^2 n \log\log n}} = 1. \tag{4.27}$$

Proof Assume, for simplicity, that $\sigma^2 = 1$. To prove the theorem, it is enough to show that with probability one, for $\varepsilon > 0$ small enough, we have:

$$\limsup_{n\to+\infty} \frac{|S_n|}{\sqrt{2n \log\log n}} \le 1 + 4\varepsilon, \tag{4.28}$$

$$\limsup_{n\to+\infty} \frac{S_n}{\sqrt{2n \log\log n}} \ge 1 - 4\varepsilon. \tag{4.29}$$

Proof of (4.28). Choose $\alpha > 0$ such that $\alpha(1+4\varepsilon)^2 > 1$ and define, for each $k \ge 1$, $n_k = [e^{k^\alpha}]$. Then,

$$\sum_{k=1}^{\infty} \frac{\exp(-(1+4\varepsilon)^2 \log\log n_k)}{(1+4\varepsilon)\sqrt{2\log\log n_k}} \le \left(\frac{\alpha}{2}\right)^{1/2} \sum_{k=1}^{\infty} \frac{1}{k^{\alpha(1+4\varepsilon)^2}} < \infty;$$

hence, taking into account Lemma 4.15, we have that

$$\sum_{k=1}^{\infty} \mathbf{P}\big(|S_{n_k}| > (1+4\varepsilon)\sqrt{2n_k \log\log n_k}\big) < \infty,$$

so (4.28) follows along the subsequence n_k, that is,

$$\limsup_{k\to+\infty} \frac{|S_{n_k}|}{\sqrt{2n_k \log\log n_k}} \le 1 + 4\varepsilon.$$

We still need to control the remaining terms of the sequence. For this step, define, for each $k \ge 1$,

$$M_k = \sup_{n_k \le n < n_{k+1}} \frac{|S_n - S_{n_k}|}{\sqrt{2n_k \log\log n_k}}.$$

Then, obviously, for each $n \in [n_k, n_{k+1})$, we have that

$$\frac{|S_n|}{\sqrt{2n_k \log\log n}} \le \frac{|S_{n_k}|}{\sqrt{2n_k \log\log n_k}} + M_k.$$

The first term on the right has just been handled, so we need to prove the almost sure convergence to 0 of the sequence M_k. Assume that $p(1-\alpha) \ge 2$. As $p > 2$, this assumption on α is compatible with the previous one, $\alpha(1+4\varepsilon)^2 > 1$, whenever $\varepsilon < \frac{1}{8}$. Then, taking into account Corollary 2.21, there exists a constant $K > 0$, depending only on p, such that

$$\sum_{k=1}^{\infty} \mathrm{E} M_k^p \le K \sum_{k=1}^{\infty} \frac{(n_{k+1} - n_k)^{p/2}}{\sqrt{2n_k \log\log n_k}} \le K \sum_{k=1}^{\infty} \frac{1}{k^{p(1-\alpha)/2} (\log n_k)^{p/2}} < \infty,$$

so, again the Borel–Cantelli lemma implies that, with probability one, $M_k \longrightarrow 0$, as $k \longrightarrow +\infty$.

Proof of (4.29). Let $N > 2$ be fixed and define, for each $k \geq 1$, $\tau_k = S_{N^k} - S_{N^{k-1}+[N^{k/2}]}$ and

$$C_k = \left\{\omega : \tau_k(\omega) > (1-2\varepsilon)\psi\left(N^k - N^{k-1} - \left[N^{k/2}\right]\right)\right\},$$

where $\psi(n) = \sqrt{2n \log\log n}$. We start by showing that $\sum_{k=1}^{\infty} \mathbf{P}(C_k) = +\infty$. For this, notice that, for N_0 large enough and some suitable constant $C_1 > 0$,

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{\exp(-(1-2\varepsilon)^2 \log\log(N^k - N^{k-1} - [N^{k/2}]))}{(1-2\varepsilon)\sqrt{2\log\log(N^k - N^{k-1} - [N^{k/2}])}} \\
&\quad \geq C_1 + \sum_{k=N_0}^{\infty} \exp\left(-(1-\varepsilon)^2 \log\log\left(N^k - N^{k-1} - \left[N^{k/2}\right]\right)\right) \\
&\quad \geq C_1 + \sum_{k=N_0}^{\infty} \exp\left(-(1-\varepsilon)^2 \log\log N^k\right) \\
&\quad = C_1 + \frac{1}{(\log N)^{(1-\varepsilon)^2}} \sum_{k=1}^{\infty} \frac{1}{k^{(1-\varepsilon)^2}} = +\infty.
\end{aligned}$$

It follows now from Lemma 4.15 and the stationarity of the random variables that

$$\sum_{k=1}^{\infty} \mathbf{P}(C_k) = \sum_{k=1}^{\infty} \mathbf{P}\left(S_{N^k - N^{k-1} - [N^{k/2}]} > (1-2\varepsilon)\psi\left(N^k - N^{k-1} - \left[N^{k/2}\right]\right)\right) = +\infty.$$

Choose a real-valued function g such that $\sup_{x\in\mathbb{R}} |g'(x)| \leq \gamma < \infty$ and

$$\mathbb{I}_{((1-2\varepsilon)\psi(N^k - N^{k-1} - [N^{k/2}]), +\infty)}(x) \leq g(x) \leq \mathbb{I}_{((1-3\varepsilon)\psi(N^k - N^{k-1} - [N^{k/2}]), +\infty)}(x).$$

Then, we obviously have that

$$\sum_{k=1}^{\infty} \mathrm{E} g(\tau_k) \geq \sum_{k=1}^{\infty} \mathbf{P}(C_k) = +\infty. \tag{4.30}$$

On the other hand, we have

$$\begin{aligned}
&\mathbf{P}\left(\sum_{k=1}^{\infty} g(\tau_k) \leq \frac{1}{2}\sum_{k=1}^{n} \mathrm{E} g(\tau_k)\right) \\
&\quad \leq \mathbf{P}\left(\left|\sum_{k=1}^{\infty} g(\tau_k) - \sum_{k=1}^{n} \mathrm{E} g(\tau_k)\right| \geq \frac{1}{2}\sum_{k=1}^{n} \mathrm{E} g(\tau_k)\right) \\
&\quad \leq 4\frac{\mathrm{Var}(\sum_{k=1}^{n} g(\tau_k))}{(\sum_{k=1}^{n} \mathrm{E} g(\tau_k))^2} \\
&\quad \leq \frac{4}{\sum_{k=1}^{n} \mathrm{E} g(\tau_k)} + 8\frac{1}{(\sum_{k=1}^{n} \mathrm{E} g(\tau_k))^2} \sum_{k=1}^{\infty} \sum_{j=k+1}^{\infty} \left|\mathrm{Cov}\left(g(\tau_k), g(\tau_j)\right)\right|.
\end{aligned}$$

Taking into account Bulinsky's inequality (2.2), (4.25) and the stationarity of the variables, we still have that

$$
\begin{aligned}
\sum_{k=1}^{\infty}\sum_{j=k+1}^{\infty}\left|\mathrm{Cov}\big(g(\tau_k),g(\tau_j)\big)\right| &\le \gamma^2\sum_{k=1}^{\infty}\sum_{j=k+1}^{\infty}\left|\mathrm{Cov}(\tau_k,\tau_j)\right| \\
&\le C\gamma^2\sum_{k=1}^{\infty}\big(N^k-N^{k-1}-\big[N^{k/2}\big]\big)u\big(\big[N^{(k+1)/2}\big]\big) \\
&\le \frac{C\gamma}{N^{1/2}}\sum_{k=1}^{\infty}\frac{1}{N^{k(\theta/2-1)}}<\infty,
\end{aligned}
$$

as $\theta>2$. Recalling now (4.30) and letting $n\longrightarrow+\infty$, it follows that

$$
\mathbf{P}\left(\sum_{k=1}^{\infty}g(\tau_k)<\infty\right)=0,
$$

and hence

$$
\mathbf{P}\Big(\limsup_{k\to\infty}\big\{\tau_k>(1-3\varepsilon)\psi\big(N^k-N^{k-1}-\big[N^{k/2}\big]\big)\big\}\Big)=1. \tag{4.31}
$$

Finally, consider, for each $k\ge1$, the sets

$$
C'_k=\big\{\omega:\tau_k(\omega)>(1-3\varepsilon)\psi\big(N^k-N^{k-1}-\big[N^{k/2}\big]\big)\big\}
$$

and

$$
B_k=\big\{\omega:S_{N^{k-1}+[N^{k/2}]}(\omega)>-2\psi\big(N^{k-1}+\big[N^{k/2}\big]\big)\big\}.
$$

So, taking into account (4.28) and (4.31), we have

$$
\mathbf{P}\Big(\limsup_{k\to+\infty}B_k\cap C'_k\Big)=1.
$$

If we now choose N large enough, it follows that

$$
\begin{aligned}
\mathbf{P}\big(S_{N^k}>(1-4\varepsilon)\psi\big(N^k\big)\big) &\ge \mathbf{P}\Big(\limsup_{k\to\infty}\big\{S_{N^k}\ge(1-3\varepsilon)\psi\big(N^k-N^{k-1}-\big[N^{k/2}\big]\big) \\
&\qquad -2\psi\big(N^{k-1}+\big[N^{k/2}\big]\big)\big\}\Big) \\
&\ge \mathbf{P}\Big(\limsup_{k\to+\infty}B_k\cap C'_k\Big)=1,
\end{aligned}
$$

which concludes the proof. □

4.5 Density Estimation

We now look at the asymptotic normality of the kernel estimator for the density (3.26),

$$
\widehat{f_n}(x)=\frac{1}{nh_n}\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right).
$$

We will prove the asymptotic normality under conditions somewhat weaker than second-order stationarity, assuming that the distribution of pairs of random variables verify a diagonal decomposition as described by condition **(D)**, introduced for the treatment of the almost sure consistency of the estimator in Sect. 3.4 (see page 88). The method of proof is an extension of the approach used to prove Theorem 5.22, which is based on the blocking decomposition used in the proof of Theorem 4.1. Just as for proving Theorem 5.22, this implies a quite long and technical proof that we will present here divided into several steps.

We will need some notation for the proof of the main result in this section, Theorem 4.19 below. Let ℓ be an integer smaller than n and $m=\lfloor \frac{n}{\ell} \rfloor$, the greatest integer less than or equal to n/ℓ. Let us further define the random variables

$$T_{n,i}=\frac{1}{\sqrt{h_n}}\left(K\left(\frac{x-X_i}{h_n}\right)-\mathrm{E}K\left(\frac{x-X_i}{h_n}\right)\right), \qquad i=1,\dots,n, n\in\mathbb{N},$$

$$T_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}T_{n,i} \quad \text{and} \quad Y_{n,k}=\frac{1}{\sqrt{\ell}}\sum_{r=(k-1)\ell+1}^{jk}T_{n,r}, \qquad j=1,\dots,m.$$

Note that the random variables $Y_{n,k}$ just introduced are the analogs to the blocks defined in Theorem 5.22, now referring to the corresponding terms for the present framework. Moreover, notice also that $T_{m\ell}=\frac{1}{\sqrt{m}}\sum_{k=1}^{m}Y_{m\ell,k}=\frac{1}{\sqrt{m\ell}}\sum_{i=1}^{m\ell}T_{m\ell,i}$. An obvious adaptation of Lemma 3.24 describes the behaviour of the variances of the variables just defined.

Lemma 4.17 *Assume that* **(A)** *is satisfied. Then, for each fixed* $\ell\in\mathbb{N}$,

$$\lim_{m\to+\infty}\sigma_{m\ell}^2=\lim_{m\to+\infty}\mathrm{Var}(T_{m\ell})=g_2(x,x)\int K^2(u)\,du. \tag{4.32}$$

We now introduce the following notation: given $x\in\mathbb{R}$, write

$$\sigma^2(x)=x\int K^2(u)\,du.$$

In order to complete our proof of the asymptotic normality, we need some extra assumptions on the kernel function, due to the peculiarities of association, as already discussed when proving the almost sure consistency of the kernel estimator (see page 91). Throughout this section, we will assume that K is of bounded variation. Thus, there exist increasing functions K_1 and K_2 such that $K=K_1-K_2$. Introduce the following assumption:

(K) K_1, K_2 are bounded and $\lim_{|u|\to+\infty}K_1(u)=0$, $\lim_{|u|\to+\infty}K_2(u)=0$.

Remark that, obviously, (4.32) still holds with respect to K_1 and K_2. Corresponding to these two functions, we define, for each $n\in\mathbb{N}$, the analog of the block decomposition introduced above:

$$T_{n,i,q} = \frac{1}{\sqrt{h_n}}\left(K_q\left(\frac{x-X_i}{h_n}\right) - \mathrm{E}K_q\left(\frac{x-X_i}{h_n}\right)\right), \quad q=1,2, i=1,\dots,n,$$

$$Y_{n,k,q} = \frac{1}{\sqrt{\ell}}\sum_{r=(k-1)\ell+1}^{j\ell} T_{n,r,q}, \quad q=1,2, j=1,\dots,m.$$

Let us first prove a technical lemma that appeared in Utev [101], handling integrals of summations of random variables over sets also defined by conditions on summations of variables, separating each term. This will help us dealing with the Lindeberg conditions in the proof of the main result.

Lemma 4.18 *Let X_n, $n \in \mathbb{N}$, be random variables. Then, for all $\varepsilon > 0$ and $n \in \mathbb{N}$,*

$$\int_{\{|\sum_{i=1}^n X_i| \geq \varepsilon n\}} \left(\sum_{i=1}^n X_i\right)^2 d\mathbf{P} \leq 2n \sum_{i=1}^n \int_{\{|X_i| \geq \varepsilon/2\}} X_i^2 \, d\mathbf{P}. \tag{4.33}$$

Proof For all $\varepsilon > 0$ and $n \in \mathbb{N}$, define $g(x) = \max(2x^2 - \varepsilon^2 n^2, 0)$. Then, obviously, if $\varepsilon n \leq |x|$, one has $g(x) \geq x^2$, so,

$$\int_{\{|\sum_{i=1}^n X_i| \geq \varepsilon n\}} \left(\sum_{i=1}^n X_i\right)^2 d\mathbf{P} \leq \int g\left(\sum_{i=1}^n X_i\right) d\mathbf{P}.$$

It is also obvious that g is convex, and thus,

$$\int g\left(\sum_{i=1}^n X_i\right) d\mathbf{P} \leq \frac{1}{n}\sum_{i=1}^n \int g(nX_i)\, d\mathbf{P} \leq 2n \sum_{i=1}^n \int_{\{|X_i| \geq \varepsilon/2\}} X_i^2 \, d\mathbf{P}. \qquad \square$$

We may now state the main result of this section.

Theorem 4.19 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume* **(A)** *and* **(K)** *are satisfied. Assume that*

$$h_n \longrightarrow 0, \qquad nh_n \longrightarrow +\infty, \qquad \frac{h_{n+1}}{h_n} \longrightarrow 1, \tag{4.34}$$

$$\lim_{m\to+\infty} \frac{1}{m\ell}\sum_{k=1}^m \sum_{r,r'=(k-1)\ell+1}^{j\ell} g_{2,r,r'} = g_{2,\ell}, \tag{4.35}$$

$$\lim_{\ell\to+\infty} g_{2,\ell} = g_2 \quad \textit{uniformly}. \tag{4.36}$$

Then

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n \left(K\left(\frac{x-X_i}{h_n}\right) - \mathrm{E}K\left(\frac{x-X_i}{h_n}\right)\right)$$

converges in distribution to a centred Gaussian random variable with variance $\sigma^2(g_2(x,x))$.

Proof Write

$$
\begin{aligned}
\left|\mathrm{E}e^{iuT_n} - e^{-(u^2/2)\sigma^2(g_2(x,x))}\right| &\leq \left|\mathrm{E}e^{iuT_n} - \mathrm{E}e^{iuT_{m\ell}}\right| + \left|\mathrm{E}e^{iuT_{m\ell}} - \prod_{k=1}^{m}\mathrm{E}e^{i(u/\sqrt{m})Y_{m\ell,k}}\right| \\
&\quad + \left|\prod_{k=1}^{m}\mathrm{E}e^{i(u/\sqrt{m})Y_{m\ell,k}} - e^{-(u^2/2)\sigma^2(g_{2,\ell}(x,x))}\right| \\
&\quad + \left|e^{-(u^2/2)\sigma^2(g_{2,\ell}(x,x))} - e^{-(u^2/2)\sigma^2(g_2(x,x))}\right|. \qquad (4.37)
\end{aligned}
$$

The proof is now completed in four steps. The first three find convenient upper bounds for each of the terms in the right of the inequality above. The final step puts everything together and takes care of the final details.

Step 1. As $|\mathrm{E}e^{iuT_n} - \mathrm{E}e^{iuT_{m\ell}}| \leq |u|\,\mathrm{Var}^{1/2}(T_n - T_{m\ell})$, it is enough to prove the convergence to zero of this variance. Write

$$
\begin{aligned}
&\mathrm{Var}^{1/2}(T_n - T_{m\ell}) \\
&\leq \mathrm{Var}^{1/2}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{m\ell}(T_{n,k} - T_{m\ell,k})\right] + \mathrm{Var}^{1/2}\left[\left(\frac{1}{\sqrt{m\ell}} - \frac{1}{\sqrt{n}}\right)\sum_{k=1}^{m\ell}T_{m\ell,k}\right] \\
&\quad + \mathrm{Var}^{1/2}\left[\frac{1}{\sqrt{n}}\sum_{k=m\ell+1}^{n}T_{n,k}\right]. \qquad (4.38)
\end{aligned}
$$

We now prove that each of these terms converges to zero. As for the first term,

$$
\begin{aligned}
&\frac{1}{n}\mathrm{Var}\left[\sum_{k=1}^{m\ell}(T_{n,k} - T_{m\ell,k})\right] \\
&= \frac{1}{n}\sum_{k,k'=1}^{m\ell}\left[\mathrm{Cov}\left(\frac{1}{\sqrt{h_n}}K\left(\frac{x-X_k}{h_n}\right), \frac{1}{\sqrt{h_n}}K\left(\frac{x-X_{k'}}{h_n}\right)\right)\right. \\
&\quad - \mathrm{Cov}\left(\frac{1}{\sqrt{h_n}}K\left(\frac{x-X_k}{h_n}\right), \frac{1}{\sqrt{h_{m\ell}}}K\left(\frac{x-X_{k'}}{h_{m\ell}}\right)\right) \\
&\quad - \mathrm{Cov}\left(\frac{1}{\sqrt{h_{m\ell}}}K\left(\frac{x-X_k}{h_{m\ell}}\right), \frac{1}{\sqrt{h_n}}K\left(\frac{x-X_{k'}}{h_n}\right)\right) \\
&\quad \left. + \mathrm{Cov}\left(\frac{1}{\sqrt{h_{m\ell}}}K\left(\frac{x-X_k}{h_{m\ell}}\right), \frac{1}{\sqrt{h_{m\ell}}}K\left(\frac{x-X_{k'}}{h_{m\ell}}\right)\right)\right].
\end{aligned}
$$

From Lemma 3.24, as $\frac{n}{m\ell} \longrightarrow 1$, it follows that the summation over the first and last terms of this previous expansion is convergent to $2\sigma^2(g_2(x,x))$. The remaining terms are of the form

$$
\begin{aligned}
&\frac{1}{n}\sum_{k,k'=1}^{m\ell}\frac{1}{\sqrt{h_n h_{m\ell}}}\left(\int K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-v}{h_{m\ell}}\right)(g_{1,k,k'} - f\otimes f)\,du\,dv\right. \\
&\quad \left. + \int K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-u}{h_{m\ell}}\right)g_{2,k,k'}(u,u)\,du\right).
\end{aligned}
$$

The part corresponding to the first integral converges to zero as it is easily checked reproducing the arguments in the proof of Lemma 3.24. As for the second integral, rewrite it as

$$\frac{1}{n}\sqrt{\frac{h_n}{h_{m\ell}}}\sum_{k,k'=1}^{m\ell}\int K(z)K\left(z\frac{h_n}{h_{m\ell}}\right)g_{2,k,k'}(x-h_nz,x-h_nz)\,dz,$$

which, by the Lebesgue dominated convergence theorem and taking into account assumptions **(D)**, that are included in the set of assumptions **(A)**, and $\frac{h_{n+1}}{h_n}\longrightarrow 1$, converges to $\sigma^2(g_2(x,x))$. So, we have proved that

$$\frac{1}{n}\operatorname{Var}\left[\sum_{k=1}^{m\ell}(T_{n,k}-T_{m\ell,k})\right]\longrightarrow 0.$$

The second term in (4.38) may be rewritten as

$$\begin{aligned}&\left(\frac{1}{\sqrt{m\ell}}-\frac{1}{\sqrt{n}}\right)^2\operatorname{Var}\left[\sum_{k=1}^{m\ell}T_{m\ell,k}\right]\\&\quad=\left(1-\frac{\sqrt{m\ell}}{\sqrt{n}}\right)^2\frac{1}{m\ell h_{m\ell}}\sum_{k,k'=1}^{m\ell}\operatorname{Cov}\left(K\left(\frac{x-X_k}{h_{m\ell}}\right),K\left(\frac{x-X_{k'}}{h_{m\ell}}\right)\right).\end{aligned}$$

As $\frac{\sqrt{m\ell}}{\sqrt{n}}\longrightarrow 1$, the convergence to zero of this term follows from Lemma 3.24. Finally, for the third term in (4.38), we have

$$\begin{aligned}&\frac{1}{n}\operatorname{Var}\left[\sum_{k=m\ell+1}^{n}T_{n,k}\right]\\&\quad=\frac{1}{nh_n}\sum_{k,k'=m\ell+1}^{n}\left|\operatorname{Cov}\left(K\left(\frac{x-X_k}{h_n}\right),K\left(\frac{x-X_{k'}}{h_n}\right)\right)\right|\\&\quad\le\frac{1}{nh_n}\sum_{k,k'=1}^{n}\left(\int_{\mathbb{R}^2}K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-v}{h_n}\right)\left|g_{1,k,k'}(u,v)-f(u)f(v)\right|du\,dv\right.\\&\qquad\left.+\int_{\Delta}K^2\left(\frac{x-u}{h_n}\right)g_{2,k,k'}(u,u)\,du\right),\end{aligned}$$

which converges to zero due to the nonnegativity of the terms and $\frac{h_{n+1}}{h_n}\longrightarrow 1$. Thus, for each fixed $\ell\in\mathbb{N}$,

$$\left|\mathrm{E}e^{iuT_n}-\mathrm{E}e^{iuT_{m\ell}}\right|\longrightarrow 0. \tag{4.39}$$

Step 2. We find an upper bound for the second term in (4.37). As the variables $Y_{m\ell,k}$, $k=1,\ldots,m$, are not associated but, nevertheless, functions of associated variables, we apply the extended version of Newman's inequality proved in Theorem 2.40, to obtain

$$\left| \mathrm{E}e^{iuT_{m\ell}} - \prod_{k=1}^{m} \mathrm{E}e^{i(u/\sqrt{m})Y_{m\ell,k}} \right|$$
$$\leq \frac{u^2}{2m} \sum_{\substack{k,k'=1\\k\neq k'}}^{m} \mathrm{Cov}(Y_{m\ell,k,1} + Y_{m\ell,k,2}, Y_{m\ell,k',1} + Y_{m\ell,k',2}).$$

After expanding this covariance, we find four terms that are controlled in the same way as the following one:

$$\frac{1}{m} \sum_{\substack{k,k'=1\\k\neq k'}}^{m} \mathrm{Cov}(Y_{m\ell,k,1}, Y_{m\ell,k',1})$$
$$= \frac{1}{m} \sum_{k,k'=1}^{m} \mathrm{Cov}(Y_{m\ell,k,1}, Y_{m\ell,k',1}) - \frac{1}{m} \sum_{k=1}^{m} \mathrm{Var}(Y_{m\ell,k,1}).$$

From Lemma 3.24 it follows that

$$\frac{1}{m} \sum_{k,k'=1}^{m} \mathrm{Cov}(Y_{m\ell,k,1}, Y_{m\ell,k',1}) \longrightarrow g_2(x,x) \int K_1^2(u)\,du.$$

As for the other term,

$$\frac{1}{m} \sum_{k=1}^{m} \mathrm{Var}(Y_{m\ell,k,1})$$
$$= \frac{1}{m\ell h_{m\ell}} \sum_{k=1}^{m} \sum_{r,r'=(k-1)\ell+1}^{k\ell} \mathrm{Cov}\left(K_q\left(\frac{x-X_r}{h_{m\ell}}\right), K_q\left(\frac{x-X_{r'}}{h_{m\ell}}\right)\right).$$

Now, using **(D)**, we may write these covariances as a sum of an integral over $\mathbb{R}^2$ with an integral over the diagonal Δ of $\mathbb{R}^2$, as was done in the proof of Lemma 3.24. The first integral thus appearing is bounded above by

$$\frac{1}{m\ell h_{m\ell}} \sum_{r,r'=1}^{m\ell} \int_{\mathbb{R}^2} K_q\left(\frac{x-u}{h_{m\ell}}\right) K_q\left(\frac{x-v}{h_{m\ell}}\right) \left| g_{1,r,r'}(u,v) - f(u)f(v)\right| du\,dv \longrightarrow 0,$$

as in the proof of Lemma 3.24. The second integral appearing in this decomposition is equal to

$$\frac{1}{m\ell h_{m\ell}} \sum_{k=1}^{m} \sum_{r,r'=(k-1)\ell+1}^{k\ell} \int_{\Delta} K_q^2\left(\frac{x-u}{h_{m\ell}}\right) g_{2,r,r'}(u,u)\,du$$
$$\longrightarrow g_{2,\ell}(x,x) \int K_q^2(u)\,du,$$

taking into account (4.35). Thus,

$$\limsup_{m\to+\infty}\left|\mathrm{E}e^{iuT_{m\ell}}-\prod_{k=1}^{m}\mathrm{E}e^{i(u/\sqrt{m})Y_{m\ell,k}}\right|\leq Bu^2\big(g_2(x,x)-g_{2,\ell}(x,x)\big),\tag{4.40}$$

where $B=\int K_1^2(u)+K_2^2(u)+2K_1(u)K_2(u)\,du$.

Step 3. Controlling the third term in the upper bound on (4.37) is equivalent to proving a Central Limit Theorem for the variables $Y_{m\ell,k}$, $k=1,\ldots,m$, treating them as if they where independent (to be formally completely correct, we should introduce a new collection of independent variables with the same distributions as the ones we have; we shall not do so to avoid further notation). We shall accomplish this step by proving that the triangular array of random variables $Y_{m\ell,k}$, $k=1,\ldots,m$, satisfies the Lindeberg condition. Using Lemma 3.24, one easily checks that

$$\frac{1}{m}\operatorname{Var}\left(\sum_{k=1}^{m}Y_{m\ell,k}\right)\longrightarrow g_{2,\ell}(x,x)\int K^2(u)\,du.$$

So, the Lindeberg condition reduces to verifying that

$$\sum_{k=1}^{m}\int_{\{|Y_{m\ell,k}|>cg_{2,\ell}(x,x)\sqrt{m}\}}\frac{1}{m}Y_{m\ell,k}^2\,d\mathbf{P}\longrightarrow 0.$$

Taking into account (4.33), we get that an upper bound for the integral above is

$$\begin{aligned}&\frac{2}{m}\sum_{k=1}^{m}\sum_{r=(k-1)\ell+1}^{k\ell}\int_{\{|T_{m\ell,r}|>(cg_{2,\ell}(x,x)/2)\sqrt{m/\ell}\}}T_{m\ell,r}^2\,d\mathbf{P}\\&=\frac{2}{m}\sum_{k=1}^{m\ell}\int_{\{|T_{m\ell,k}|>(cg_{2,\ell}(x,x)/2)\sqrt{m/\ell}\}}T_{m\ell,k}^2\,d\mathbf{P}.\end{aligned}$$

Write, for notational simplicity, $c'_\ell=\frac{c}{2}g_{2,\ell}(x,x)$. Recalling the definition of $T_{m\ell,k}$ and taking into account that the kernel K is bounded, it follows that

$$\frac{2}{m}\sum_{k=1}^{m\ell}\int_{\{|T_{m\ell,k}|>c'_\ell\sqrt{m/\ell}\}}T_{m\ell,k}^2\,d\mathbf{P}\leq\frac{2}{m}\sum_{k=1}^{m\ell}\int_{\{2\|K\|_\infty>c'_\ell\sqrt{m\ell h_{m\ell}}/\ell\}}\frac{4}{h_{m\ell}}\|K\|_\infty^2\,d\mathbf{P}.$$

Assumption (4.34) implies that $m\ell h_{m\ell}\longrightarrow+\infty$ as $m\longrightarrow+\infty$ so, recalling that ℓ is fixed, the integration set becomes, for m large enough, empty, thus the integrals are 0, and the Lindeberg condition is trivially verified. Hence $m^{-1/2}\sum_{k=1}^{m}Y_{m\ell,k}$ converges in distribution to a centred Gaussian random variable with variance $\sigma^2(g_{2,\ell}(x,x))$.

Step 4. It follows from the preceding steps that, for each fixed ℓ,

$$\begin{aligned}&\limsup_{n\to+\infty}\left|\mathrm{E}e^{iuT_n}-e^{-(u^2/2)\sigma^2(g_2(x,x))}\right|\\&\quad\leq Bu^2\big(g_2(x,x)-g_{2,\ell}(x,x)\big)+\left|e^{-(u^2/2)\sigma^2(g_{2,\ell}(x,x))}-e^{-(u^2/2)\sigma^2(g_2(x,x))}\right|,\end{aligned}$$

where B is defined in (4.40). Letting now $\ell\longrightarrow+\infty$ and taking into account (4.36), we have that this upper bound converges to zero. $\square$

4.6 Regression Estimation

We may now adapt the previous approaches to prove the convergence in distribution of the regression estimator $\widehat{r}_n(x)$, as defined in (3.30):

$$\widehat{r}_n(x) = \frac{\sum_{j=1}^n Y_j K((x - X_j)/h_n)}{\sum_{j=1}^n K((x - X_j)/h_n)}.$$

The approach follows the same lines as in Sect. 3.5. We do not have an asymptotic normality result, as usual in such problems, but only a finite-dimensional normality of the regression estimator, that is, we will prove that the random vector

$$\left(\widehat{r}_n(x_1) - \mathrm{E}\widehat{r}_n(x_1), \ldots, \widehat{r}_n(x_q) - \mathrm{E}\widehat{r}_n(x_q)\right) \tag{4.41}$$

is asymptotically normal for every choice of $x_1, \ldots, x_q \in \mathbb{R}$. This will be obtained in an indirect way, applying a suitable transformation Θ, defined later, to

$$\begin{aligned} \Psi(x_1, \ldots, x_q) = \big(&\widehat{f}_n(x_1) - \mathrm{E}\widehat{f}_n(x_1), \ldots, \widehat{f}_n(x_q) - \mathrm{E}\widehat{f}_n(x_q), \\ &\widehat{m}_n(x_1) - \mathrm{E}\widehat{m}_n(x_1), \ldots, \widehat{m}_n(x_q) - \mathrm{E}\widehat{m}_n(x_q)\big), \end{aligned} \tag{4.42}$$

where $\widehat{f}_n$ and $\widehat{m}_n$ have been defined in Sects. 3.4 and 3.5, respectively (see page 97 for the later definition). As for the density estimator studied in the previous section, we need to prove a Central Limit Theorem for an arbitrary linear combination of the coordinates of this random vector. Analogously to what was done for the almost sure convergence in Sect. 3.5, this will be accomplished by adapting the proof of the density estimator to appropriately handle all the terms appearing now. The proof of Theorem 4.19 is based on the manipulation of covariances using Lemma 3.24. This correctly takes care of the terms depending only on expressions of the form $\widehat{f}_n(x_s) - \mathrm{E}\widehat{f}_n(x_s)$. As we have already proved a two-dimensional extension of this result in Lemma 3.36, we can reproduce the same arguments to handle the sums depending only on terms of the form $\widehat{m}_n(x_s) - \mathrm{E}\widehat{m}_n(x_s)$. We are thus left with the cross terms depending both on $\widehat{f}_n(x_s) - \mathrm{E}\widehat{f}_n(x_s)$ and $\widehat{m}_n(x_s) - \mathrm{E}\widehat{m}_n(x_s)$. As could be expected, the extension of assumptions **(D)** and **(A)** introduced in Sect. 3.5 offers a solution for this problem, providing an extension of Lemmas 3.24 and 3.36. We just need to complete **(A′)** to handle the cross terms, adding a fourth assumption:

(A′) **(A′.4)** $b_2^{**}(v) = \int_{\mathbb{R}} u b_2(v, u)\, du$ is bounded and continuous.

Lemma 4.20 *Assume that* **(D′)** *and* **(A′.1)**–**(A′.4)** *are satisfied and that the kernel* K *is bounded. Then*

$$\frac{1}{nh_n} \sum_{j,k=1}^n \mathrm{Cov}\left(K\left(\frac{x - X_j}{h_n}\right), Y_k K\left(\frac{x - X_k}{h_n}\right)\right) \longrightarrow b_2^{**}(x, x) \int K^2(u)\, du$$

uniformly on any compact set.

The proof of the result is, as before, a simple repetition of the arguments of the proof of Lemma 3.24.

We may now state the result concerning the convergence in distribution of (4.41). The proof follows by repeating the steps of the proof of Theorem 4.19 and taking the terms corresponding to the same x_s, $\widehat{f}_n(x_s) - \mathrm{E}\widehat{f}_n(x_s)$ and $\widehat{m}_n(x_s) - \mathrm{E}\widehat{m}_n(x_s)$, in pairs, followed by taking into account a Taylor expansion of a suitable transformation (this is also known as the δ-method).

Theorem 4.21 *Let X_n, $n \in \mathbb{N}$, be associated random variables. Assume* **(A)**, **(A′)** *and* **(K)** *are satisfied. If*

$$h_n \longrightarrow 0, \qquad nh_n \longrightarrow +\infty, \qquad \frac{h_{n+1}}{h_n} \longrightarrow 1, \tag{4.43}$$

$$\lim_{m\to+\infty} \frac{1}{m\ell} \sum_{k=1}^{m} \sum_{r,r'=(k-1)\ell+1}^{k\ell} g_{2,r,r'} = g_{2,\ell}, \tag{4.44}$$

$$\lim_{m\to+\infty} \frac{1}{m\ell} \sum_{k=1}^{m} \sum_{r,r'=(k-1)\ell+1}^{k\ell} b_{2,r,r'} = b_{2,\ell}, \tag{4.45}$$

$$\lim_{\ell\to+\infty} g_{2,\ell} = g_2 \quad \textit{uniformly}, \tag{4.46}$$

$$\lim_{\ell\to+\infty} b_{2,\ell} = b_2 \quad \textit{uniformly}. \tag{4.47}$$

Then, for all $x_1, \ldots, x_q \in \mathbb{R}$, the random vector

$$\sqrt{nh_n}\left(\widehat{r}_n(x_1) - \frac{\mathrm{E}\widehat{m}_n(x_1)}{\mathrm{E}\widehat{f}_n(x_1)}, \ldots, \widehat{r}_n(x_q) - \frac{\mathrm{E}\widehat{m}_n(x_q)}{\mathrm{E}\widehat{f}_n(x_q)}\right)$$

converges in distribution to a centred Gaussian random vector with covariance matrix

$$\begin{aligned} \Gamma^* = \mathrm{diag}\Bigg(& \frac{r^2(x_1)g_2(x_1,x_1) - 2r(x_1)b_2^{**}(x_1,x_1) + b_2^*(x_1)}{f(x_1)}, \ldots, \\ & \frac{r^2(x_q)g_2(x_q,x_q) - 2r(x_q)b_2^{**}(x_q,x_q) + b_2^*(x_q)}{f(x_q)}\Bigg) \int K^2(u)\,du. \end{aligned}$$

Proof Start by reproducing the arguments in proof of Theorem 4.19 to conclude that the random vector $\sqrt{nh_n}\Psi(x_1, \ldots, x_q)$ converges in distribution to a centred Gaussian random vector with covariance matrix

$$\Gamma = \begin{bmatrix} g_2(x_1,x_1) & & 0 & b_2^{**}(x_1,x_1) & & 0 \\ & \ddots & & & \ddots & \\ 0 & & g_2(x_q,x_q) & 0 & & b_2^{**}(x_q,x_q) \\ b_2^{**}(x_1,x_1) & & 0 & b_2^*(x_1,x_1) & & 0 \\ & \ddots & & & \ddots & \\ 0 & & b_2^{**}(x_q,x_q) & 0 & & b_2^*(x_q,x_q) \end{bmatrix}$$
$$\times \int K^2(u)\,du.$$

Finally, to obtain an asymptotic result for the regression estimator $\widehat{r}_n(x)$, apply the transformation $\Theta(v_1,\dots,v_q,u_1,\dots,u_q) = (\frac{u_1}{v_1},\dots,\frac{u_q}{v_q})$ to $\Psi(x_1,\dots,x_q)$ and use a Taylor expansion to find

$$
\begin{aligned}
&\sqrt{nh_n}\left(\widehat{r}_n(x_1) - \frac{\mathrm{E}\widehat{m}_n(x_1)}{\mathrm{E}\widehat{f}_n(x_1)},\dots,\widehat{r}_n(x_q) - \frac{\mathrm{E}\widehat{m}_n(x_q)}{\mathrm{E}\widehat{f}_n(x_q)}\right)\\
&\quad = \sqrt{nh_n}\,\Theta\big(\Psi(x_1,\dots,x_q)\big)\\
&\quad \approx \sqrt{nh_n}\sum_{s=1}^{q}\frac{\partial\Theta}{\partial y_s}(u_s)\big(\widehat{f}_n(x_s) - \mathrm{E}\widehat{f}_n(x_s)\big) + \sum_{s=q+1}^{2q}\frac{\partial\Theta}{\partial y_s}(u_s)\big(\widehat{m}_n(x_s) - \mathrm{E}\widehat{m}_n(x_s)\big),
\end{aligned}
$$

where $u_s = u_{q+s} = \frac{\mathrm{E}\widehat{m}_n(x_s)}{\mathrm{E}\widehat{f}_n(x_s)}$, for $s = 1,\dots,q$. The higher-order terms are negligible, so computing the partial derivatives and the covariance matrix of this new vector, we conclude the proof. $\square$

Chapter 5
Convergence in Distribution—Functional Results

Abstract This chapter addresses functional central limit theorems, that is invariance principles and the convergence of empirical processes. The importance of these processes come, of course, from the several statistical applications that are based on transformations of the random-sum process or of the empirical process. Both these sequences of processes are shown to converge in distribution to suitable Gaussian processes. Some transforms depend closely on the paths of processes, while others are only integral transformations, thus being less sensitive to the regularity of the observed path. These arguments justify that, depending on the functionals that we are interested in, we may require the convergence with respect to the usual Skorokhod space or with respect to some suitable L^p space. These, being weaker topological spaces, will be less demanding in order to have the convergence in distribution. The techniques are similar to those used in Chap. 4 but adapted to handle the technicalities that arise from the underlying functional space.

5.1 Introduction

Having studied the Central Limit Theorem, it is now time to interest ourselves with the functional versions of these results. Naturally, as for the Central Limit Theorem, the first functional results appeared assuming stationarity in Newman and Wright [72], and later without stationarity in Birkel [12], considering the convergence with respect to the supnorm in $C[0, 1]$, the space of continuous functions defined on $[0, 1]$, or the Skorokhod topology in $D[0, 1]$, the space of càdlàg functions (see Billingsley [10] for details). A large part of the effort relied on the extension to associated random variables of the classical inequalities on moments or tail probabilities. Of course, proofs became somewhat more intricate, but the essential of the results known for independent variables is extended. Much of the control obtained depends on some weak stationarity, as described for example, in Theorem 5.12, proved by Birkel [12], or on the decrease rate of the covariances, as in Theorem 5.14, proved later by Birkel [16]. As usual for this kind of results, much of the proving effort is spent with the tightness of the sequences. When dealing with integral functionals of the sample paths of random sums process or its continuous counterparts, we do not need a topology as strong as the one considered in the previous subsection. This means that it is reasonable to expect to prove the convergence

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8_5,

in distribution of these processes under weaker assumptions on the covariance structure of the underlying random variables. The first result in this direction was proved by Prokhorov [84], assuming the underlying variables to be independent, but afterwards these weaker spaces did not attract much attention. Having this in mind, Oliveira and Suquet [76, 78] rephrased the problem in some L^p space. These spaces have weaker topologies, so the tightness becomes simpler to characterize, but are still strong enough to allow for applying interesting functionals to the partial-sum process. In this weaker topological framework, the functional Central Limit Theorem is proved under some weak form of stationarity assumption and a Lindeberg condition. Analogous results and approaches for the asymptotics of empirical processes were studied by Yu [109], who proved the convergence towards a suitable Gaussian process assuming that $\mathrm{Cov}(X_1, X_n)$ decays fast enough, with a rate obtained by using inequality (2.25). This was later improved by Shao and Yu [94] and Louhichi [63], relaxing this decay rate for the covariances. Recasting the problem in a convenient L^p space, Oliveira and Suquet [77, 79] obtained the convergence of the empirical process under a still weaker decay rate on the covariances.

The theory of convergence in distribution with respect to the Skorokhod topology is well known and established. We refer the reader for the monographs by Billingsley [10, 11] for an account on this subject. In regards with convergence in distribution in L^p-spaces, the results seem to be spread throughout the literature, so we include some general characterizations of tightness and convergence in distribution on $L^p[0, 1]$ spaces.

5.2 General Results on Weak Convergence in $L^p[0, 1]$ Spaces

For underlying associated variables and these weaker spaces, the literature seems to have essentially concentrated on the study of empirical processes, more widely applied in statistical problems. Nevertheless, we can still prove convenient versions of the empirical process on these L^p spaces.

A general result implying the convergence in distribution in $L^p[0, 1]$ is the following, which is analogous to the well-known conditions in the Skorokhod space.

Theorem 5.1 *Let ζ_n, $n \in \mathbb{N}$, and ζ be random variables with values in $L^p[0, 1]$ for some $p > 1$. Assume that*:

(a) *for every $f \in L^q[0, 1]$ with $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\int_0^1 f(t)\zeta_n(t)\lambda(dt) \xrightarrow{d} \int_0^1 f(t)\zeta(t)\lambda(dt),$$

where λ is the Lebesgue measure on $[0, 1]$;

(b) *the sequence* ζ_n, $n \in \mathbb{N}$, *is relatively compact.*

Then ζ_n *converges in distribution to* ζ *in* $L^p[0, 1]$.

For the proof of this result and a more complete characterization of convergence in distribution on Banach spaces, we refer the interested reader to Vakhania, Tarieladze and Chobanyan [102] (see Sect. IV.3) or Ledoux and Talagrand [56] (see Sect. 2.1). In fact, condition (a) in the previous theorem is just a more convenient statement, adapted to $L^p[0, 1]$ spaces, of the a general condition for separable Banach spaces, where an analogue of a characteristic function should be considered. This remark is of particular interest in the case of a Hilbert space, as happens for $L^2[0, 1]$. When using Theorem 5.1 to prove convergence in distribution, verification of condition (a) reduces to a Central Limit Theorem for real random variables, so we are left with the need to prove the relative compactness of the sequence. Of course, as $L^p[0, 1]$ spaces are separable, this is equivalent to proving the tightness of the sequence, according to the well-known Prokhorov theorem (see, for example, Billingsley [10], Theorems 6.1 and 6.2). These tightness characterizations were first presented in Oliveira [73] and Oliveira and Suquet [77] for $L^2[0, 1]$ and in Suquet [100] for $L^p[0, 1]$ spaces using a special wavelet multiresolution analysis. Here we follow Oliveira and Suquet [79] for an approach using more classical $L^p[0, 1]$ arguments.

Theorem 5.2 *Let* ζ_n, $n \in \mathbb{N}$, *be a sequence of random elements in* $L^p[0, 1]$ *for some* $p \geq 1$ *verifying*:

(a) *For some* $\gamma > 1$, $\sup_{n\geq 1} \mathrm{E}\|\zeta_n\|_1^\gamma < \infty$,
(b) $\lim_{h\to 0} \sup_{n\geq 1} \mathrm{E}\|\zeta_n(\cdot + h) - \zeta_n(\cdot)\|_p^p = 0$.

Then ζ_n, $n \in \mathbb{N}$, *is tight in* $L^p[0, 1]$.

Proof To avoid notational complications, we extend the definition of the ζ_n outside the interval $[0, 1]$ by putting $\zeta_n(t) = 0$ for any $t \notin [0, 1]$. Consider now a probability density K with support $[-1, 1]$ that we assume to be Lipschitzian, that is,

$$\|K\|_{\mathrm{Lip}} := \sup_{s\neq t} \frac{|K(t) - K(s)|}{|t - s|} < \infty.$$

For any positive integer j, define $K_j(t) = jK(jt)$. The sequence K_j, $j \in \mathbb{N}$, is an approximate identity (see, for example, Stroock [97]), and we will use it to reproduce the convolution approach that is typical of the wavelet multiresolution analysis.

Writing

$$K_j * \zeta_n(x) - \zeta_n(x) = \int_{[-1/j,1/j]} \big(\zeta_n(x - t) - \zeta_n(x)\big) K_j(t)\lambda(dt)$$

and using Jensen's inequality with respect to the probability measure whose density function is $K_j(t)$, we easily obtain

$$\begin{aligned}
&\mathrm{E}\Big(\int_{\mathbb{R}}\big|K_j*\zeta_n(x)-\zeta_n(x)\big|^p\lambda(dx)\Big)\\
&\quad\le \mathrm{E}\Big(\int_{[-1/j,1/j]}K_j(t)\int_{\mathbb{R}}\big|\zeta_n(x-t)-\zeta_n(x)\big|^p\lambda(dx)\,dt\Big)\\
&\quad\le \sup_{t\in[-1/j,1/j]}\mathrm{E}\big\|\zeta_n(\cdot)-\zeta_n(\cdot+t)\big\|_p^p.
\end{aligned}$$

Hence by (b),

$$\lim_{j\to\infty}\sup_{n\ge 1}\mathrm{E}\|K_j*\zeta_n-\zeta_n\|_p^p=0. \tag{5.1}$$

Now, it is easily checked that

$$\mathrm{E}\|K_j*\zeta_n\|_\infty\le\|K_j\|_\infty\sup_{i\ge 1}\mathrm{E}\|\zeta_i\|_1$$

and, for any $0\le s<t\le 1$,

$$\mathrm{E}\big(\big|K_j*\zeta_n(t)-K_j*\zeta_n(s)\big|^\gamma\big)\le\|K_j\|_{\mathrm{Lip}}^\gamma\sup_{i\ge 1}\mathrm{E}\|\zeta_i\|_1^\gamma|t-s|^\gamma.$$

So, according to Billingsley [10], Theorem 12.3, it follows that, for each fixed $j\in\mathbb{N}$, the sequence $K_j*\zeta_n$, $n\in\mathbb{N}$, is tight in $C[0,1]$ and hence, also in $L^p[0,1]$.

Now we use the approximation to identity defined by the sequence K_j, $j\in\mathbb{N}$, to prove the tightness of the sequence ζ_n, $n\in\mathbb{N}$, itself. For any fixed $\eta>0$, define, for each $k\ge 1$, $\eta_k=2^{-k}\eta$ and choose a sequence of positive ε_k decreasing to 0. By (5.1) and the Markov inequality, there exists a subsequence j_k, $k\ge 1$, such that

$$P\big(\|\zeta_n-K_{j_k}*\zeta_n\|_p>\varepsilon_k\big)<\eta_k,\quad n\ge 1,k\ge 1. \tag{5.2}$$

By the $L^p[0,1]$-tightness of the sequence $K_{j_k}*\zeta_n$, $n\ge 1$, there is a compact C_k in $L^p[0,1]$ such that

$$P(K_{j_k}*\zeta_n\notin C_k)<\eta_k,\quad n\ge 1,k\ge 1. \tag{5.3}$$

Defining $A=\{f\in L^p[0,1]\colon K_{j_k}*f\in C_k \text{ and } \|K_{j_k}*f-f\|_p\le\varepsilon_k, k\ge 1\}$, it follows from (5.2) and (5.3) that

$$P(\zeta_n\in A)>1-2\eta,\quad n\ge 1.$$

Clearly, the set A is totally bounded, so, as $L^p[0,1]$ is complete, it follows that A is compact in $L^p[0,1]$, so the sequence ζ_n, $n\in\mathbb{N}$, is tight in $L^p[0,1]$. □

The following corollary gives alternative conditions for the tightness in $L^p[0,1]$ that are easier to verify.

Corollary 5.3 *Let ζ_n, $n\in\mathbb{N}$, be a sequence of random elements in $L^p[0,1]$ such that, for some $q\le p<r$:*

(a) *for some constant $c>0$, $\mathrm{E}|\xi_n(t)|^r\le c$ for all $t\in[0,1]$ and $n\in\mathbb{N}$,*
(b) *$\mathrm{E}|\xi_n(t+h)-\xi_n(t)|^q\le\varepsilon(h)$ for $0\le h<1$, $0\le t\le 1-h$ and $n\in\mathbb{N}$, for some function $\varepsilon(\cdot)$ such that $\varepsilon(h)\longrightarrow 0$ as $h\to 0$.*

Proof As (a) obviously implies Theorem 5.2(a), we only need to prove that Theorem 5.2(b) holds. If $q = p$, this is evident, so we are left with the case $q < p$. Denote by $\frac{1}{u}$ and $\frac{1}{v}$ the barycentric coordinates of p in the segment (q, r), that is,

$$p = \frac{1}{u}q + \frac{1}{v}r, \qquad \frac{1}{u} + \frac{1}{v} = 1, \quad u, v > 0.$$

The Hölder inequality applied to $|\xi_n(t+h) - \xi_n(t)|$ gives

$$\mathrm{E}\big|\xi_n(t+h) - \xi_n(t)\big|^p \le \mathrm{E}^{1/u}\big|\xi_n(t+h) - \xi_n(t)\big|^q \mathrm{E}^{1/v}\big|\xi_n(t+h) - \xi_n(t)\big|^r.$$

So, integrating with respect to t, we find

$$\begin{aligned}\mathrm{E}\big\|\xi_n(\cdot+h) - \xi_n(\cdot)\big\|_p^p &\le \int \mathrm{E}^{1/u}\big|\xi_n(t+h) - \xi_n(t)\big|^q \mathrm{E}^{1/v}\big|\xi_n(t+h) - \xi_n(t)\big|^r \lambda(dt)\\ &\le \big(2^{r-1}c\big)^{1/v}\varepsilon(h)^{1/u},\end{aligned}$$

using (a) and (b). Hence $\mathrm{E}\|\xi_n(\cdot+h) - \xi_n(\cdot)\|_p^p$ converges to zero, uniformly in n, as h goes to zero, that is, Theorem 5.2(b) is satisfied. □

The following result gives a characterization of relative compactness in separable Hilbert spaces, in terms of the coefficients representing the process with respect to some orthonormal basis. We state the result for $L^2[0,1]$, due to our particular interests. This is a restatement of Theorem 2.2 in Parthasarathy [80], taking into account a correction introduced by Suquet [99]. An immediate adaptation will give the corresponding result for a general separable Hilbert space.

Corollary 5.4 *Let e_n, $n \in \mathbb{N}$, is an orthonormal basis of $L^2[0,1]$, and ζ_n, $n \in \mathbb{N}$, be a sequence of random elements in $L^2[0,1]$ verifying*:

(a) $\sup_{n\ge1} \mathrm{E}\|\zeta_n\|_2^2 < +\infty$,
(b) $\lim_{N\to+\infty} \sup_{n\ge1} \mathrm{E}[\sum_{i=N}^{+\infty}(\int_0^1 e_i(t)\zeta_n(t)\lambda(dt))^2] = 0$.

Then ζ_n, $n \in \mathbb{N}$, is tight in $L^2[0,1]$.

Proof For simplicity, define, for each $f \in L^2[0,1]$,

$$r_N^2(f) = \sum_{i=N}^{+\infty}\left(\int_0^1 e_i(t)f(t)\lambda(dt)\right)^2.$$

Then, assumptions (a) and (b) rewrite as

$$\sup_{n\in\mathbb{N}} \mathrm{E}\big(r_1^2(\zeta_n)\big) < +\infty \quad \text{and} \quad \lim_{N\to+\infty}\sup_{n\ge1} \mathrm{E}\big(r_N^2(\zeta_n)\big) = 0,$$

respectively. Define, for each $N \in \mathbb{N}$,

$$\Psi(N) = \sup_{n\ge1} \mathrm{E}\big(r_N^2(\zeta_n)\big).$$

Now, given $\varepsilon > 0$, choose, taking into account (b), a sequence of nonnegative real numbers $\lambda_k \nearrow +\infty$ and a sequence of positive integers n_k, $k \ge 1$, such that

$\sum_{k=1}^{\infty} \lambda_k \Psi(n_k) < \varepsilon$. Define the sets $E_k = \{f \in L^2[0,1] : r_{N_k}^2(f) \le \lambda_k^{-1}\}$ and put $K = \bigcap_{k=1}^{\infty} E_k$. It follows from (a) that this set K is bounded. Thus, the projection over the first n_k coordinates is relatively compact, so it is possible to cover this projection using a finite number of balls with radii λ_k^{-1}. The set K is included in each E_k, which means that it is possible to cover K using the balls with the same centres as above but with radii $2\lambda_k^{-1}$. Thus, the set K is totally bounded, so it is compact, as $L^2[0,1]$ is complete. □

5.3 Invariance Principles

Having studied the Central Limit Theorem, it is now time to interest ourselves with the functional versions of these results. For this purpose, recalling that $S_n = X_1 + \cdots + X_n$ and $s_n^2 = \mathrm{E} S_n^2$, we define two versions of the partial-sum process with underlying variables $X_1, \ldots, X_n$:

$$\xi_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} \quad \text{and} \quad \xi_n^*(t) = \frac{1}{s_n} S_{[nt]}, \qquad t \in [0,1], \tag{5.4}$$

where $[x]$ represents the largest integer less than or equal to x, and σ needs to be defined later. The interest in characterizing the convergence in distribution of these sequences of random functions on appropriate function spaces is that this convergence gives a description of the global behaviour of the partial sums. Moreover, considering convenient continuous transformations, we will still have the convergence in distribution of the transformed random elements. This remark is quite useful as many statistical applications may be defined as transformations of the above-mentioned processes or of the empirical processes, to be studied in the next section. A simple example of such a transformation is $\max_{i \le n} |S_i|$, which is just the supnorm of the $\sigma\sqrt{n}\xi_n(t)$. All we need is to place ourselves in a space where this norm is a continuous transformation. Another statistically interesting family of transformations has the form $\int f(t)\xi_n(t)\,dt$ or $\int G(\xi_n(t))\,dt$, which require weaker topologies on the function space to be continuous. The natural choices of spaces just mentioned would be the Skorokhod space $D[0,1]$, extensively studied in Billingsley [10], for example, for the supnorm transformation, while the weaker $L^p[0,1]$ space for suitable $p \ge 1$ would be enough for the integral transformation of sample paths of the partial-sum process. The two cases mentioned are important choices for the function space and are essentially different in what concerns the topological characterizations, thus meaning that different treatments should be used in either case. We will study the convergence in distribution in the Skorokhod space in Sect. 5.3.1 and the L^2 case in Sect. 5.3.2.

The first suggestion to address functional versions of the Central Limit Theorem seems to be attributable to Erdos and Käc [35], which was promptly answered by Donsker [31] for the Skorokhod space and by Prokhorov [84] for the $L^2[0,1]$ space, in both cases considering independent underlying variables and obtaining as limits a version of suitably renormalized Brownian motion. As expected, this was followed by extensions to various dependence structures, in this latter case the limit becoming a Gaussian process with convenient covariance operator.

Definition 5.5 The random variables X_n, $n \in \mathbb{N}$, are said to fulfill the *invariance principle* if ξ_n or ξ_n^* converge in distribution to a random function W such that its distribution is the Wiener measure.

Remark 5.6 Notice that we consider the sequence ξ_n or ξ_n^*. In general, the convergence of each sequence is not equivalent to the convergence of the other. However, in the case of associated variables and the assumptions we will be considering, these will become, in fact, equivalent.

Remark 5.7 We also did not include any reference to the space were the convergence takes place. We will be interested in the convergence with respect to different topologies and spaces, so a full statement about an invariance principle must include the space to which this convergence refers.

The Skorokhod space attracted essentially all the attention in regards with invariance principles. Most of the approaches to results on dependent variables would consider some kind of approximation to independence, either by trying to prove that the underlying variables X_i and X_j become almost independent whenever $|i-j| \longrightarrow +\infty$, or by coupling the variables, or blocks of variables, by independent ones, in ways similar to what has been done in Sect. 2.6. The asymptotic independence was first controlled in terms of mixing conditions, as in Ibragimov [47] or Theorem 20.1 in Billingsley [10]. The first invariance principle for associated random variables was proved by Newman and Wright [71] and later improved by Birkel [12] dropping the stationarity assumption.

The weaker $L^2[0, 1]$ space received more attention in regards with the transformations of empirical processes that will be considered in more detail in Sect. 5.4.

5.3.1 Invariance Principle in $D[0, 1]$

The random functions ξ_n or ξ_n^* are both discontinuous, so the space $D[0, 1]$ is a more appropriate and natural choice to address the convergence if we are interested in continuity with respect to a strong topology. In some cases we may be interested in a continuous version of ξ_n or ξ_n^*, constructed by considering polygonal lines between the end points of each step, but, as it will be commented at the end of this subsection (see Remark 5.15), this is essentially the same problem. The noncontinuous random functions, considered in the Skorokhod space, even allow for a simpler proof.

We begin by considering the case were the variables are strictly stationary. In such a case, the invariance principle follows under the sole assumption that a limit covariance exists.

Theorem 5.8 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and strictly stationary associated random such that*

$$\sigma^2 = \mathrm{Var}(X_1) + 2\sum_{j=2}^{\infty} \mathrm{Cov}(X_1, X_j) < \infty. \tag{5.5}$$

Then the random variables X_n, $n \in \mathbb{N}$, fulfill the invariance principle in $D[0,1]$.

Proof We follow the usual steps based on Theorem 15.1 in Billingsley [10] to prove this theorem, that is, we need to prove the tightness of the sequence and the convergence in distribution of the finite-dimensional distributions. Tightness will be proved using a maximal inequality and standard arguments as for independent variables, once we have proved a Central Limit Theorem for the X_n random variables. Also, using the same Central Limit Theorem, the finite-dimensional distributions of ξ_n will be showed to converge to the corresponding ones of a Brownian process.

As our assumptions coincide with those of Theorem 4.1, we immediately have that

$$\frac{1}{\sigma\sqrt{n}} S_n \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Let now $m \in \mathbb{N}$ and $0 \le t_0 < t_1 < \cdots < t_m \le 1$ be fixed. From the previous convergence we will characterize the asymptotic distribution of the increments $\xi_n(t_{j+1}) - \xi_n(t_j)$, $j = 0, \ldots, m-1$. In fact, using the stationarity,

$$\begin{aligned}\xi_n(t_{j+1}) - \xi_n(t_j) &= \frac{1}{\sigma\sqrt{n}}(S_{[nt_{j+1}]} - S_{[nt_j]}) \\ &= \frac{1}{\sigma\sqrt{n}} S_{[nt_{j+1}]-[nt_j]} = \frac{1}{\sigma\sqrt{n}} S_{[n(t_{j+1}-t_j)]+z_j},\end{aligned}$$

where, for each $j = 0, \ldots, m-1$, z_j may be equal to -1, 0 or 1. Multiplying by $\sqrt{[n(t_{j+1}-t_j)] + z_j}$ and taking into account that, for each fixed $j = 0, \ldots, m-1$, $\frac{[n(t_{j+1}-t_j)]+z_j}{n} \to t_{j+1} - t_j$, it follows that

$$\xi_n(t_{j+1}) - \xi_n(t_j) \xrightarrow{d} Z_j \sim \mathcal{N}(0, t_{j+1} - t_j).$$

Let us now look at the covariance between different increments. Assume, without loss of generality, that $j < k$. Then

$$\begin{aligned}&\mathrm{Cov}\big(\xi_n(t_{j+1}) - \xi_n(t_j), \xi_n(t_{k+1}) - \xi_n(t_k)\big) \\ &\quad= \frac{1}{\sigma^2 n}\mathrm{Cov}(S_{[n(t_{j+1}-t_j)]+z_j}, S_{[n(t_{k+1}-t_k)]+z_k}) \\ &\quad= \frac{1}{\sigma^2 n}\mathrm{Cov}(X_{[nt_j]+1} + \cdots + X_{[nt_{j+1}]+1}, X_{[nt_k]+1} + \cdots + X_{[nt_{k+1}]+1}) \\ &\quad\le \frac{1}{\sigma^2} \sum_{\ell=[n(t_k-t_j)]+z_{jk}}^{\infty} \mathrm{Cov}(X_1, X_\ell),\end{aligned}$$

using the stationarity of the random variables and the nonnegativity of the covariances. It follows from (5.5) that this upper bound converges to 0 as $n \longrightarrow \infty$, so that $\operatorname{Cov}(\xi_n(t_{j+1}) - \xi_n(t_j), \xi_n(t_{k+1}) - \xi_n(t_k)) \longrightarrow 0$. It is also clear that the variables $\xi_n(t_1) - \xi_n(t_0), \ldots, \xi_n(t_m) - \xi_n(t_{m-1})$, being nondecreasing transformations of the original variables, are associated. Thus, taking into account Theorem 1.33, if $(\xi_n(t_1) - \xi_n(t_0), \ldots, \xi_n(t_m) - \xi_n(t_{m-1})) \xrightarrow{d} (W_1, \ldots, W_m)$, the coordinate variables of the limit are associated with null covariances, thus independent. That is, the finite-dimensional distributions converge weakly to those of a Brownian process.

Finally, let us prove the tightness of the sequence ξ_n, $n \in \mathbb{N}$. Recalling Theorem 2.28, we have

$$\mathbf{P}\big(\max\big(|S_1|, \ldots, |S_n|\big) \geq \lambda s_n\big) \leq 2\mathbf{P}\big(|S_n| \geq (\lambda - \sqrt{2})s_n\big),$$

so we may conclude the proof using standard techniques. In fact, if $\lambda > 2\sqrt{2}$, we have

$$\mathbf{P}\big(\max\big(|S_1|, \ldots, |S_n|\big) \geq \lambda s_n\big) \leq 2\mathbf{P}\left(|S_n| \geq \frac{\lambda s_n}{2}\right).$$

The stationarity of the variables obviously implies that

$$s_n^2 = n \operatorname{Var}(X_1) + 2\sum_{j=2}^{n}(n - i + 1)\operatorname{Cov}(X_1, X_j),$$

thus from (5.5) it follows $\frac{s_n^2}{n\sigma^2} \longrightarrow 1$. As already mentioned, the random variables X_n, $n \in \mathbb{N}$, satisfy the Central Limit Theorem, so $\frac{1}{s_n} S_n \xrightarrow{d} Z \sim \mathcal{N}(0, 1)$, and

$$\mathbf{P}\left(\frac{1}{s_n}|S_n| > \frac{\lambda}{2}\right) \longrightarrow \mathbf{P}\left(|Z| > \frac{\lambda}{2}\right).$$

Thus, for $n \in \mathbb{N}$ large enough,

$$\mathbf{P}\left(\frac{1}{s_n}|S_n| > \frac{\lambda}{2}\right) \leq \frac{16}{\lambda^3}\mathrm{E}|Z|^3,$$

so, choosing $\varepsilon = \frac{\mathrm{E}|Z^3|}{\lambda^2}$ and λ sufficiently large, the tightness follows from Theorem 8.3 in Billingsley [10] (see the proof of Theorem 16.1 in [10] for the reformulation on $D[0, 1]$). □

In order to relax the stationarity requirement, we will need a closer control on the covariance structure of the random variables. The control of the variance of S_n will be crucial. In particular, the assumption

$$\frac{1}{s_n^2} s_{nk}^2 \longrightarrow k, \quad k \in \mathbb{N}, \tag{5.6}$$

plays an important role.

Lemma 5.9 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Then (5.6) is satisfied if and only if $\frac{1}{s_n^2} s_{[nt]}^2 \longrightarrow t$ for every $t \in [0, +\infty)$.*

Proof Let us assume that $\frac{1}{s_n^2}s_{nk}^2 \longrightarrow k$ for every $k \in \mathbb{N}$. As the variables are associated, the sequence s_n^2 is nondecreasing, so, taking into account (5.6), we have that, for all $p \in \mathbb{N}$ and $0 \leq \ell \leq p-1$,

$$\limsup_{m\to+\infty} \frac{1}{s_{mp+\ell}^2} s_m^2 \leq \limsup_{m\to+\infty} \frac{1}{s_{mp}^2} s_m^2 = \frac{1}{p}. \tag{5.7}$$

Choose now an integer $k \geq 2$. Given $m \in \mathbb{N}$, define $r = [\frac{m}{k-1}]$, so we obviously have that $r(k-1) = [\frac{m}{k-1}](k-1) \leq m$. If $m \geq (k-1)^2$, then

$$rk - m > m + 1 - k + \left[\frac{m}{k-1}\right] - m \geq 0,$$

thus $\ell \leq p-1 \leq (rk-m)p$, that is, $mp+\ell \leq rkp$. Hence,

$$\liminf_{m\to\infty} \frac{1}{s_{mp+\ell}^2} s_m^2 \geq \liminf_{m\to\infty} \frac{1}{s_{rkp}^2} s_{r(k-1)}^2 = \left(1-\frac{1}{k}\right)\frac{1}{p}. \tag{5.8}$$

As k may be chosen arbitrarily large, it follows that

$$\frac{1}{s_n^2} s_{[nt]}^2 \to t \quad \text{for } t = \frac{1}{p} \text{ with } p \in \mathbb{N}.$$

For $t = \frac{q}{p}$, the conclusion now follows by writing

$$\frac{1}{s_n^2} s_{[n(q/p)]}^2 = \frac{1}{s_n^2} s_{[nq]}^2 \frac{1}{s_{[nq]}^2} s_{[n(q/p)]}^2 \longrightarrow \frac{q}{p}.$$

Finally, for nonrational $t \in [0, \infty)$, we can approximate from below and from above by sequences of rationals u_ℓ and v_ℓ, respectively, and use the association, which implies that s_m^2 is nondecreasing, to find

$$u_\ell = \lim \frac{1}{s_n^2} s_{nu_\ell}^2 \leq \liminf \frac{1}{s_n^2} s_{nt}^2 \leq \limsup \frac{1}{s_n^2} s_{nt}^2 \leq \lim \frac{1}{s_n^2} s_{nv_\ell}^2 = v_\ell,$$

from which the result follows. The other implication is obvious. □

We start by relating the fulfillment of the invariance principle and the Central Limit Theorem. To control the variances, we will need a somewhat strengthened version of (5.6).

Theorem 5.10 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. The following statements are equivalent*:

(a) *The random variables X_n, $n \in \mathbb{N}$, satisfy a Central Limit Theorem, and*

$$\frac{1}{s_n^2} \mathrm{E}(S_{nk} S_{n\ell}) \longrightarrow \min(k, \ell), \quad k, \ell \in \mathbb{N}. \tag{5.9}$$

(b) *The random variables X_n, $n \in \mathbb{N}$, satisfy the invariance principle in $D[0, 1]$.*

Proof We first prove that (a) implies (b). Taking into account Lemma 5.9, we get that, for each $t \in [0, 1]$, $\frac{1}{s_n} S_{[nt]} \xrightarrow{d} tZ$, where $Z \sim \mathcal{N}(0, 1)$. Consider now fixed $0 < s < t \le 1$. Then, as the marginal distributions are tight, the sequence of random vectors $(\frac{1}{s_n} S_{[ns]}, \frac{1}{s_n} S_{[nt]})$, $n \in \mathbb{N}$, is tight, so there exists a probability measure Q on $\mathbb{R}^2$ such that

$$\left(\frac{1}{s_n} S_{[ns]}, \frac{1}{s_n} S_{[nt]}\right) \xrightarrow{d} Q.$$

Denoting by π_1 and π_2 the coordinate projections from $\mathbb{R}^2$ onto $\mathbb{R}$, we have that

$$\left(\frac{1}{s_n} S_{[ns]}, \frac{1}{s_n} (S_{[nt]} - S_{[ns]})\right) \xrightarrow{d} Q(\pi_1, \pi_2 - \pi_1)^{-1}.$$

The random variables $S_{[ns]}$ and $S_{[nt]} - S_{[ns]}$ are nondecreasing transformations of the X_n's and so are associated. Thus, it follows from Theorem 1.33 that the projection π_1 and $\pi_2 - \pi_1$ are associated with respect to Q. From the convergence in distribution of $\frac{1}{s_n^2} S_{[nt]}$, using Theorem 5.4 in Billingsley [10], we derive the uniform integrability of each of the following sets of random variables: $\{\frac{1}{s_n} S_{[nu]}, n \in \mathbb{N}\}$ for each fixed $u \in [0, 1]$ and $\{\frac{1}{s_n^2}(S_{[nu]} - S_{[nv]}), n \in \mathbb{N}\}$ for each fixed $0 \le u < v \le 1$. Hence, using (5.9), it follows that, with respect to Q,

$$\begin{aligned}
&\operatorname{Cov}(\pi_1, \pi_2 - \pi_1) \\
&\quad = \lim_{n\to\infty} \operatorname{Cov}\left(\frac{1}{s_n} S_{[ns]}, \frac{1}{s_n}(S_{[nt]} - S_{[ns]})\right) \\
&\quad = \lim_{n\to\infty} \frac{1}{s_n^2} \mathrm{E}\big(S_{[ns]}(S_{[nt]} - S_{[ns]})\big) = 0.
\end{aligned}$$

That is, π_1 and $\pi_2 - \pi_1$ are, under Q, associated and uncorrelated, thus independent. Finally, we have that $Q\pi_1^{-1} = \mathcal{N}(0, s)$ and $Q(\pi_2 - \pi_1)^{-1} = \mathcal{N}(0, t - s)$, so that

$$\frac{1}{s_n}(S_{[nt]} - S_{[ns]}) \xrightarrow{d} (t - s)Z,$$

where $Z \sim \mathcal{N}(0, 1)$.

We still have to prove the tightness of the sequence ξ_n^*. For this, we will prove, arguing along the lines of the proof of Theorem 2.2 in Herrndorf [45], that, for all $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbf{P}\big(\omega\big(\xi_n^*, \delta\big) > \varepsilon\big) \le \eta,$$

where $\omega(\xi_n^*, \delta) = \sup_{|s-t|<\delta} |\xi_n^*(s) - \xi_n^*(t)|$. Using a standard decomposition (see, for example, the corollary to Theorem 8.3 in Billingsley [10]), we have, for $\delta \in (0, 1)$,

$$\mathbf{P}\big(\omega\big(\xi_n^*, \delta\big) > \varepsilon\big) \le \sum_{i=0}^{[1/\delta]} \mathbf{P}\left(\max_{[ni\delta]<r\le[n(i+1)\delta]} |S_r - S_{[ni\delta]}| > \frac{\varepsilon s_n}{3}\right).$$

Repeating the arguments leading to (2.20), we find that

$$\mathbf{P}\Big(\max_{[ni\delta]<r\le[n(i+1)\delta]}|S_r-S_{[ni\delta]}|>\frac{\varepsilon s_n}{3}\Big)$$
$$\le\Big(1-\frac{36}{s_n^2\varepsilon}\mathrm{E}(S_{[n(i+1)\delta]}-S_{[ni\delta]})\Big)^{-1}\mathbf{P}\Big(|S_{[n(i+1)\delta]}-S_{[ni\delta]}|\ge\frac{\varepsilon s_n}{6}\Big).$$

Using (5.9) and Lemma 5.9, it is easy to check that, for each fixed i,

$$\frac{1}{s_n^2}\mathrm{E}(S_{[n(i+1)\delta]}-S_{[ni\delta]})^2\longrightarrow\delta,$$

so, for n large enough (depending on i and δ), $\frac{1}{s_n^2}\mathrm{E}(S_{[n(i+1)\delta]}-S_{[ni\delta]})^2\le 2\delta$. Thus, if $\frac{72\delta}{\varepsilon}<1$,

$$\mathbf{P}\big(\omega(\xi_n,\delta)>\varepsilon\big)\le\sum_{i=0}^{[1/\delta]}\Big(1-\frac{72\delta}{\varepsilon^2}\Big)^{-1}\mathbf{P}\Big(|S_{[n(i+1)\delta]}-S_{[ni\delta]}|\ge\frac{\varepsilon s_n}{6}\Big).$$

Now, as $\frac{s_n^2}{s^2_{[n(i+1)\delta]-[ni\delta]}}\longrightarrow\delta$, for n large enough (depending now on i, δ and ε), we have

$$\begin{aligned}
&\mathbf{P}\big(\omega\big(\xi_n^*,\delta\big)>\varepsilon\big)\\
&\le\sum_{i=0}^{[1/\delta]}\Big(1-\frac{72\delta}{\varepsilon^2}\Big)^{-1}\mathbf{P}\Big(\frac{1}{s_{[n(i+1)\delta]-[ni\delta]}}|S_{[n(i+1)\delta]}-S_{[ni\delta]}|\ge\frac{\varepsilon}{12\delta^{1/2}}\Big)\\
&\le\Big(1-\frac{72\delta}{\varepsilon^2}\Big)^{-1}\Big(\Big[\frac{1}{\delta}\Big]+1\Big)\frac{144\delta}{\varepsilon^2}\sup_{i\le[1/\delta]}\frac{1}{s^2_{[n(i+1)\delta]-[ni\delta]}}\\
&\quad\times\mathrm{E}\big[(S_{[n(i+1)\delta]}-S_{[ni\delta]})^2\mathbb{I}_{|S_{[n(i+1)\delta]}-S_{[ni\delta]}|>\varepsilon(s_{[n(i+1)\delta]-[ni\delta]})/(12\delta^{1/2})}\big]\\
&=\Big(1-\frac{72\delta}{\varepsilon^2}\Big)^{-1}\Big(\Big[\frac{1}{\delta}\Big]+1\Big)\frac{144\delta}{\varepsilon^2}\sup_{m,n\in\mathbb{N}}\frac{1}{s_n^2}\\
&\quad\times\mathrm{E}\big[(S_{m+n}-S_m)^2\mathbb{I}_{|S_{m+n}-S_n|>\varepsilon s_n/(12\delta^{1/2})}\big].
\end{aligned}$$

Finally, as the variables appearing in this last mathematical expectation are uniformly integrable, taking into account Theorem 5.4 in Billingsley [10], we may choose $\delta>0$ small enough so that this upper bound is smaller than the given $\eta>0$, which concludes the proof of the tightness.

We now prove that (b) implies (a). It is enough to prove that (5.9) holds. Let us first prove that $s_n^2=nh(n)$, where the function h is slowly varying. Obviously, for u large enough, $h(u)=\frac{s^2_{[u]}}{u}$. As the Central Limit Theorem is satisfied, we have both

$$\frac{1}{s_{[nt]}}S_{[nt]}\xrightarrow{d}Z\quad\text{and}\quad\frac{1}{s_n}S_{[nt]}\xrightarrow{d}tZ,\quad\text{where }Z\sim\mathcal{N}(0,1).$$

Therefore, $\frac{s^2_{[nt]}}{s^2_n} \longrightarrow t$. As the variables fulfill the invariance principle, the sequence ξ^*_n, $n \in \mathbb{N}$, is tight, so, for every $\delta > 0$,

$$\limsup_{n\to\infty} \mathbf{P}\Big(\max_{1\le i\le n} |X_i| \ge \varepsilon\sigma_n\Big) \le \limsup_{n\to\infty} \mathbf{P}\big(\omega\big(\xi^*_n, \delta\big) \ge \varepsilon\big).$$

This later upper bound converges to 0 as $\delta \longrightarrow 0$, which implies that $\frac{X_n}{s_n} \longrightarrow 0$ in probability. Thus, as $\frac{1}{s_{n+1}} S_{n+1} = \frac{1}{s_{n+1}} S_n + \frac{X_{n+1}}{s_{n+1}}$, the limits in distribution of $\frac{1}{s_n} S_n$ and $\frac{1}{s_{n+1}} S_n$ coincide, hence $\frac{s_n}{s_{n+1}} \longrightarrow 1$. Consequently, for each $t \in (0, 1]$, $\lim_{s\to\infty} \frac{s_{[ts]}}{s_{t[s]}} = \lim_{s\to\infty} \frac{h(ts)}{h(s)} = 1$, that is, h is slowly varying. Thus,

$$\frac{1}{s^2_n} s^2_{nk} \longrightarrow k, \quad k \in \mathbb{N}.$$

Let now $0 \le s \le t \le u \le v \le 1$ be fixed. The random variables $\frac{1}{s^2_n} S^2_n$, $n \in \mathbb{N}$, are uniformly integrable, hence, it follows from Lemma 5.9 that the variables $\frac{1}{s^2_n}(S_{[nt]} - S_{[ns]})(S_{[nv]} - S_{[nu]})$, $n \in \mathbb{N}$, are also uniformly integrable. As the invariance principle is verified, we have that

$$\frac{1}{s^2_n}(S_{[nt]} - S_{[ns]})(S_{[nv]} - S_{[nu]}) \xrightarrow{d} \big(W(t) - W(s)\big)\big(W(v) - W(u)\big),$$

where W is a Brownian motion. Then, taking into account Theorem 5.4 in Billingsley [10], we have

$$\frac{1}{s^2_n}\mathrm{E}\big[(S_{[nt]} - S_{[ns]})(S_{[nv]} - S_{[nu]})\big] \longrightarrow \mathrm{E}\big[\big(W(t) - W(s)\big)\big(W(v) - W(u)\big)\big] = 0,$$

using the independence of the increments of W. From this it immediately follows that

$$\frac{1}{s^2_n}\mathrm{E}\big[(S_{nj} - S_{ni})(S_{n\ell} - S_{nk})\big] \longrightarrow 0, \quad i \le j \le k \le \ell,$$

and, finally, (5.9) by choosing $i = 0$ and $j = k$. □

Remark 5.11 Notice that the association of the random variables is only used in the first part of the proof of the previous result.

We now extend the results from Sect. 4.2. We will look for assumptions on the covariance structure instead of (5.9), as, in view of Theorem 2.37, this is a much more natural way of controlling the asymptotics for associated sequences. The control over the covariance structure is achieved through the usual coefficient $u(n)$ introduced earlier (see page 41),

$$u(n) = \sup_{k\in\mathbb{N}} \sum_{j:|j-k|\ge n} \mathrm{Cov}(X_j, X_k).$$

Theorem 5.12 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. If* (4.12), (4.13), (4.14) *and* (5.6) *are satisfied, then the random variables X_n, $n \in \mathbb{N}$, verify the invariance principle in $D[0,1]$.*

Proof Taking into account Theorems 4.8 and 5.10, it is enough to prove that (5.9) is satisfied. For this, we shall verify that

$$\frac{1}{s_n^2}\mathrm{E}\big[(S_{nj}-S_{ni})(S_{n\ell}-S_{nk})\big] \longrightarrow 0, \quad i \le j \le k \le \ell,$$

as (5.9) then follows by choosing $i=0$ and $j=k$. Now, due to the nonnegativity of the covariances,

$$\begin{aligned}&\frac{1}{s_n^2}\mathrm{E}\big[(S_{nj}-S_{ni})(S_{n\ell}-S_{nk})\big]\\ &\quad= \frac{1}{s_n^2}\sum_{p=ni+1}^{nj}\sum_{q=nk+1}^{n\ell}\mathrm{Cov}(X_p,X_q) \le \frac{1}{s_n^2}u\big(n(k-j)\big) \longrightarrow 0,\end{aligned}$$

as follows from (4.12) and (4.13). □

In the proof of Theorem 5.10 we have verified that $\frac{1}{n}s_n^2$ is a slowly varying function. A common assumption in Central Limit Theorems or invariance principles requires the convergence of this sequence. In such a case, the following result is an immediate consequence of Theorem 5.12.

Corollary 5.13 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Assume that* (4.12), (4.14) *and*

$$\frac{1}{n}s_n^2 \longrightarrow \sigma^2 \in (0,\infty) \tag{5.10}$$

are satisfied. Then the random variables X_n, $n \in \mathbb{N}$, verify the invariance principle in $D[0,1]$.

Proof It suffices to remark that (4.13) and (5.6) are both consequences of (5.10). □

The Lindeberg assumption (4.14) above may be replaced by the existence of higher-order moments. But this must be accompanied by a convenient behaviour on the covariance structure. In some sense, this is alike what happens when replacing the Lindeberg assumption by the Lyapunov condition for the classical Central Limit Theorem. In our framework, dealing with dependent variables, we need a closer control on the covariances expressed through the decrease rate of $u(n)$.

Theorem 5.14 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables. Assume that* (4.13), (5.6),

$$\begin{aligned}&u(n) \sim n^{-\theta} \quad \textit{for some } \theta > 0,\\ &\sup_{n\in\mathbb{N}} \mathrm{E}|X_n|^{r+\eta} < \infty \quad \textit{for some } r>2, \eta>0,\end{aligned}$$

are satisfied. Then the random variables X_n, $n \in \mathbb{N}$, verify the invariance principle in $D[0,1]$.

Proof We start by verifying that the finite-dimensional distributions of ξ_n^* converge in distribution to those of a stochastic process W that has finite-dimensional projections behaving like the ones from a Brownian motion. It follows from (5.6) and, as the random variables are associated, Lemma 5.9 that, for each $t \in [0,1]$,

$$\frac{1}{s_n} S_{[nt]} \xrightarrow{d} tZ, \quad Z \sim \mathcal{N}(0,1).$$

Hence, each of the two families of variables $\frac{1}{s_n} S_{[nt]}$, $n \in \mathbb{N}$, and $\frac{1}{s_n^2} S_{[nt]}^2$, $n \in \mathbb{N}$, is uniformly integrable. Thus, taking into account Theorem 5.4 in Billingsley [10], we have

$$\mathrm{E}\big(W(t)\big) = \lim_{n\to\infty} \mathrm{E}\left(\frac{1}{s_n} S_{[nt]}\right) = 0 \quad \text{and} \quad \mathrm{E}\big(W^2(t)\big) = \lim_{n\to\infty} \mathrm{E}\left(\frac{1}{s_n^2} S_{[nt]^2}\right) = t.$$

Consider now fixed $k \in \mathbb{N}$ and $0 \le t_1 \le t_2 \le \cdots \le t_k \le 1$. The random variables $\xi_n^*(t_{j+1}) - \xi_n^*(t_j)$, $j = 0,\ldots,k-1$, are nondecreasing functions of the X_n's and thus are associated. So, repeating the arguments of the proof of Theorem 5.8, it follows that

$$\xi_n^*(t_{j+1}) - \xi_n^*(t_j) \xrightarrow{d} Z_j \sim \mathcal{N}(0, t_{j+1} - t_j), \quad j = 0,\ldots,k-1,$$

and

$$\begin{aligned}
&\big(\xi_n^*(t_1) - \xi_n^*(t_0), \ldots, \xi_n^*(t_k) - \xi_n^*(t_{k-1})\big) \\
&\quad \xrightarrow{d} \big(W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})\big).
\end{aligned}$$

Moreover, from Theorem 1.33 it follows that $W(t_1) - W(t_0), \ldots, W(t_k) - W(t_{k-1})$ are associated random variables. Proceeding as in the proof of Theorem 5.12, we get from (4.13) that, for $j \neq \ell$,

$$\begin{aligned}
&\mathrm{Cov}\big(W(t_{j+1}) - W(t_j), W(t_{\ell+1}) - W(t_\ell)\big) \\
&\quad = \lim_{n\to\infty} \mathrm{Cov}\big(\xi_n^*(t_{j+1}) - \xi_n^*(t_j), \xi_n^*(t_{\ell+1}) - \xi_n^*(t_\ell)\big) = 0.
\end{aligned}$$

Hence, it follows that the variables $W(t_{j+1}) - W(t_j)$, $j = 0,\ldots,k-1$, are uncorrelated and thus independent, taking into account Theorem 1.17. This means that the limit stochastic process has finite-dimensional distributions as those of a Brownian motion.

To complete the proof, we need to verify the tightness of the sequence ξ_n^*, $n \in \mathbb{N}$. Recall that, as used before, we have, for each $\varepsilon > 0$,

$$\mathbf{P}\big(\omega(\xi_n^*, \delta) > \varepsilon\big) \le \sum_{i=0}^{[1/\delta]} \mathbf{P}\left(\max_{[ni\delta] < p \le [n(i+1)\delta]} |S_p - S_{[ni\delta]}| > \frac{\varepsilon s_n}{3}\right).$$

From Corollary 2.21 it follows that $\sup_{m\in\mathbb{N}} \mathrm{E}|S_{m+n} - S_m|^r \le Bn^{r/2}$, so, using Theorem A.5, we have

$$\sup_{m\in\mathbb{N}} \mathrm{E}\Big(\max_{1\le p\le n} |S_{m+p} - S_m|^r\Big) \le B' n^{r/2},$$

where $B' > 0$ is independent of n. Using this upper bound together with Markov's inequality, we find

$$\begin{aligned}\mathbf{P}\big(\omega\big(\xi_n^*, \delta\big) > \varepsilon\big) &\le \left(\frac{3}{\varepsilon s_n}\right)^r \sum_{i=0}^{[1/\delta]} \mathrm{E}\Big(\max_{[ni\delta]<p\le[n(i+1)\delta]} |S_p - S_{[ni\delta]}|^r\Big)\\ &\le B' \frac{1}{n^{r/2}} \left(\frac{3n^{1/2}}{\varepsilon s_n}\right)^r \sum_{i=0}^{[1/\delta]} \big(\big[n(i+1)\delta\big] - [ni\delta]\big)^{r/2}.\end{aligned}$$

Finally, taking into account (4.13), we have

$$\mathbf{P}\big(\omega\big(\xi_n^*, \delta\big) > \varepsilon\big) \le B' \left(\left[\frac{1}{\delta}\right] + 1\right) \left(\delta + \frac{1}{n}\right)^{r/2} \left(\frac{3}{\varepsilon}\right)^r \inf_{n\in\mathbb{N}} \left(\frac{n}{s_n^2}\right)^{r/2}.$$

As $r > 2$, this upper bound approaches zero as $\delta \searrow 0$, and thus the tightness of ξ_n^*, $n \in \mathbb{N}$, in $D[0, 1]$ follows recalling Theorem 15.5 in Billingsley [10]. □

Remark 5.15 We have been looking at the noncontinuous random functions ξ_n or ξ_n^*. As mentioned previously, we might be interested in the continuous versions obtained by considering polygonal lines between the end points of each step,

$$\xi_n^c(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + \frac{nt - [nt]}{\sigma\sqrt{n}} X_{[nt]+1}$$

for the continuous version of ξ_n, and analogously constructed ξ_n^{*c}, the continuous version of ξ_n^*. It is obvious that, for each $\varepsilon > 0$,

$$\mathbf{P}\Big(\sup_{t\in[0,1]} \big|\xi_n(t) - \xi_n^c(t)\big| \ge \varepsilon\Big) \le \mathbf{P}\Big(\max_{1\le i\le n} |X_i| \ge \varepsilon\sigma\sqrt{n}\Big).$$

Now, if either ξ_n or ξ_n^c fulfills the invariance principle, repeating the arguments used on the proof of Theorem 5.10 (see the end of page 141), it follows that the upper bound above converges to 0. So, finally, Theorem 4.1 in Billingsley [10] implies that ξ_n^c or ξ_n, respectively, also fulfills the invariance principle. Of course, these arguments may applied with respect to ξ_n^* and the corresponding continuous version.

5.3.2 *Invariance Principle in $L^2[0, 1]$*

We shall now look at how the general results of the previous subsection translate into characterizations concerning the invariance principle. First we present a simple condition for the tightness in $L^2[0, 1]$. This result has appeared in several versions and with different proofs in Oliveira [73] and Oliveira and Suquet [76, 78].

Theorem 5.16 *Let X_n, $n \in \mathbb{N}$, be centred, square-integrable and associated random variables. Assume that there exists a constant $C > 0$ such that*

$$\frac{1}{n} \sum_{j,k=1}^{n} \mathrm{E}(X_j X_k) \leq C. \tag{5.11}$$

Then the sequence ξ_n, $n \in \mathbb{N}$, is weakly relatively compact in $L^2[0, 1]$.

Proof Let e_n, $n \in \mathbb{N}$, be an orthonormal basis of $L^2[0, 1]$ and define, for each $n \in \mathbb{N}$, $f_n(s) = \int_s^1 e_n(t)\lambda(dt)$. According to Corollary 5.4, it is enough to prove

$$\lim_{N\to+\infty} \sup_{n\in\mathbb{N}} \int \sum_{i=N}^{\infty} \left(\int_0^1 \xi_n(t) e_i(t) \lambda(dt) \right)^2 d\mathbf{P} = 0$$

and

$$\sup_{n\in\mathbb{N}} \int \sum_{i=0}^{\infty} \left(\int_0^1 \xi_n(t) e_i(t) \lambda(dt) \right)^2 d\mathbf{P} < \infty.$$

It is easily checked that

$$\int_0^1 \xi_n(t) e_i(t) \lambda(dt) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \int_{j/n}^1 e_i(t) \lambda(dt),$$

so that we have

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} \int \sum_{i=N}^{\infty} \left(\int_0^1 \xi_n(t) e_i(t) \lambda(dt) \right)^2 d\mathbf{P} \\
&\quad = \sup_{n\in\mathbb{N}} \sum_{i=N}^{\infty} \frac{1}{n} \sum_{j,k=1}^{n} f_i\left(\frac{j}{n}\right) f_i\left(\frac{k}{n}\right) \mathrm{E}(X_j X_k) \\
&\quad \leq \sup_{n\in\mathbb{N}} \left(\sup_{x\in[0,1]} \sum_{i=N}^{\infty} f_i^2(x) \right) \frac{1}{n} \sum_{j,k=1}^{n} \mathrm{E}(X_j X_k),
\end{aligned}$$

as $\mathrm{E}(X_j X_k) \geq 0$, due to the association. Now, taking into account (5.11) and Dini's theorem, we get that this converges to zero. The second condition is trivially verified by choosing $N = 1$ in the previous calculation. □

Remark 5.17 A common assumption when proving central limit theorems, invariance principles or the convergence of empirical processes is $\frac{1}{n}\mathrm{E}(S_n^2) \longrightarrow \sigma^2 < \infty$, as in Corollary 5.13. If we assume that the variables X_n, $n \in \mathbb{N}$, are associated, this obviously implies (5.11), so the relative compactness of ξ_n, $n \in \mathbb{N}$, in $L^2[0, 1]$ follows.

Remark 5.18 The statement of Theorem 5.16 only uses the association in the final step of the majorization of the series that we need to control. A general condition for the relative compactness in $L^2[0,1]$, without any kind of positive dependence, obviously is

$$\frac{1}{n}\sum_{j,k=1}^{n}\left|\mathrm{E}(X_jX_k)\right|\leq C.$$

It is an easy consequence of the computations above that we always have relative compactness in $L^2[0,1]$ if we consider the normalization by s_n instead of $\sqrt{n}$.

Corollary 5.19 *Let X_n, $n\in\mathbb{N}$, be centred, square-integrable and associated random variables. Then the sequence ξ_n^*, $n\in\mathbb{N}$, is weakly relatively compact in $L^2[0,1]$.*

Proof Using the calculation as in the previous theorem, we would find the upper bound

$$\sup_{n\in\mathbb{N}}\left(\sup_{x\in[0,1]}\sum_{i=N}^{\infty}f_i^2(x)\right)\frac{1}{s_n^2}\sum_{j,k=1}^{n}\mathrm{E}(X_jX_k)=\sup_{n\in\mathbb{N}}\left(\sup_{x\in[0,1]}\sum_{i=N}^{\infty}f_i^2(x)\right).$$

Finally, Dini's theorem gives the convergence to zero we sought. □

We may prove the tightness in $L^p[0,1]$ spaces for $p>2$, using Corollary 2.21, by strengthening somewhat our assumptions requiring, in particular, a faster convergence decay for the covariances.

Theorem 5.20 *Let X_n, $n\in\mathbb{N}$, be centred and associated random variables.*

(a) *Assume that*

$$\begin{aligned}&u(n)\sim n^{-\theta}\quad \text{for some } \theta>0,\\&\sup_{n\in\mathbb{N}}\mathrm{E}|X_n|^{p+\eta}<\infty\quad\text{for some } p>2,\eta>0.\end{aligned}\tag{5.12}$$

Then the sequence ξ_n, $n\in\mathbb{N}$, is tight in $L^p[0,1]$.

(b) *If, in addition, we assume that $\frac{1}{n}\mathrm{E}(S_n^2)\longrightarrow\sigma^2<\infty$, then the sequence ξ_n^*, $n\in\mathbb{N}$, is tight in $L^p[0,1]$.*

Proof We will verify that the conditions of Theorem 5.2 are satisfied. In what concerns condition (b) of Theorem 5.2, notice first that, as the random variables are associated, according to Corollary 2.21,

$$\begin{aligned}\mathrm{E}\left\|\xi_n(\cdot+h)-\xi_n(\cdot)\right\|_p^p&=\frac{1}{\sigma^pn^{p/2}}\mathrm{E}(X_{[nt]+1}+\cdots+X_{[n(t+h)]})^p\\&\leq\frac{B[nh]^{p/2}}{\sigma^pn^{p/2}}\end{aligned}\tag{5.13}$$

for some $B>0$ (independent of $n\in\mathbb{N}$ and $t\in[0,1]$). So,

$$\lim_{h\to0}\sup_{n\geq1}\mathrm{E}\left\|\xi_n(\cdot+h)-\xi_n(\cdot)\right\|_p^p\leq\lim_{h\to0}\frac{Bh^{p/2}}{\sigma^p}=0.$$

Condition (a) of Theorem 5.2 follows choosing $\gamma = p$ and the same arguments as above, thus proving (a).

As for ξ_n^*, $n \in \mathbb{N}$, the upper bound in (5.13) becomes $\frac{B[nh]^{p/2}}{s_n^{p/2}}$, where, as usual, $s_n^2 = \mathrm{E}(S_n^2)$. If $\frac{s_n^2}{n} \longrightarrow \sigma^2$, the previous is, for n large enough, bounded above by $2Bh^{p/2} \longrightarrow 0$ as $h \longrightarrow 0$. □

To complete the proof of an invariance principle, according to Theorem 5.1, we must now have a control on the convergence in distribution of $\int \xi_n(t) f(t)\lambda(dt)$ for every f chosen in the appropriate space. Let us recall that we may rewrite the integral above as

$$\int_0^1 \xi_n(t) f(t)\lambda(dt) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n X_i F\left(\frac{i}{n}\right), \tag{5.14}$$

where $F(s) = \int_s^1 f(t)\lambda(dt)$, so we need to prove a Central Limit Theorem for these triangular arrays of random variables. As before, when proving convergence in distribution, the main problem is to have some control on the covariances placed outside of the principal diagonal of the covariance matrix. The essence is to impose conditions that imply that the sum of those covariances became negligible. We achieve this in a somewhat different way than that used for the invariance principles in $D[0, 1]$, saving us from the need to assume some convergence rate to zero of the $u(n)$, as done, for example in Theorem 5.14. Besides, we will need only the existence of moments of order 2, thus also relaxing somewhat the assumptions on the moments.

Before the result on the convergence of these integrals, we need to prove an additional technical lemma.

Lemma 5.21 *Let u_n, $n \in \mathbb{N}$, be a sequence of real numbers such that*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^n u_k = \tau.$$

Then, for each absolutely continuous function h defined on $[0, 1]$,

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^n h^2\left(\frac{k}{n}\right) u_k = \tau \|h\|_2^2.$$

Proof It is enough to verify the convergence for h Lipschitzian, as this class of functions is dense in the space of absolutely continuous functions. Denote $v_n = \sum_{k=1}^n u_k$. We may write $\frac{1}{n} v_n = \tau + \varepsilon_n$ where, $\varepsilon_n \longrightarrow 0$. Then, it follows that

$$u_k = k(\tau + \varepsilon_k) - (k-1)(\tau + \varepsilon_{k-1}) = \tau + k\varepsilon_k - (k-1)\varepsilon_{k-1},$$

so to prove the lemma, it suffices to verify

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^n h^2\left(\frac{k}{n}\right) \big(k\varepsilon_k - (k-1)\varepsilon_{k-1}\big) = 0.$$

As h is Lipschitzian, there exists a constant $\alpha > 0$ such that $|h(x)-h(y)| \leq \alpha|x-y|$, so

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{k=1}^{n} h^2\left(\frac{k}{n}\right)\left(k\varepsilon_k - (k-1)\varepsilon_{k-1}\right)\right| \\
&\quad = \left|\frac{1}{n}\sum_{k=1}^{n-1} k\varepsilon_k\left(h^2\left(\frac{k}{n}\right) - h^2\left(\frac{k+1}{n}\right)\right) + \varepsilon_n h^2(1)\right| \\
&\quad \leq \frac{2\alpha\|h\|_\infty}{n}\sum_{k=1}^{n-1}|\varepsilon_k| + |\varepsilon_n| h^2(1) \longrightarrow 0,
\end{aligned}$$

by Cesàro convergence (see Lemma C.6). For the general case, remark that $\|h\|_2 \leq \|h\|_\infty \leq C\|h'\|_2$ (where h' denotes the almost everywhere derivative of the absolutely continuous function h) and use standard density arguments. □

The next result states a Central Limit Theorem for triangular arrays, as needed to control (5.14). The proof follows a method similar to the proofs of Theorems 4.1, 4.4 or 4.8, which consists in approximating the characteristic function of sums by the characteristic functions of sums of blocks and treat these as if they were independent.

Theorem 5.22 *Let X_n, $n \in \mathbb{N}$, be centred and associated random variables. For each $\ell \in \mathbb{N}$, put $m = [\frac{n}{\ell}]$, $Y_{j,\ell} = \sum_{i=(j-1)\ell+1}^{j\ell} X_i$, $j = 1, \dots, m$, and $Y_{m+1,\ell} = \sum_{k\ell+1}^{n} X_i$. Assume that the following conditions are verified*:

$$\lim_{n\to+\infty}\frac{1}{n}\mathrm{E}\left(S_n^2\right) = \sigma^2 > 0, \tag{5.15}$$

$$\lim_{m\to+\infty}\frac{1}{m}\sum_{j=1}^{m}\mathrm{E}\left(Y_{j,\ell}^2\right) = a_\ell \quad \text{and} \quad \lim_{\ell\to+\infty}\frac{a_\ell}{\ell} = \sigma^2, \tag{5.16}$$

$$\forall\delta > 0, \quad \frac{1}{n}\sum_{i=1}^{n}\int_{\{|X_i|>\delta\sqrt{n}\}} X_i^2\, d\mathbf{P} \longrightarrow 0. \tag{5.17}$$

Then, for all $p > 1$ and $f \in L^p[0,1]$,

$$\int_0^1 \xi_n(t) f(t)\lambda(dt) \xrightarrow{d} Z_f \sim \mathcal{N}\left(0, \|F\|_2\right),$$

where $F(s) = \int_s^1 f(t)\lambda(dt)$.

Proof As mentioned above, we are interested in the random variables

$$S_n(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} F\left(\frac{i}{n}\right) X_i, \quad \text{where } F(s) = \int_s^1 f(t)\lambda(dt).$$

We may consider only functions F which are Lipschitzian, as these are a dense subset of the space $\{h : h(s) = \int_s^1 g(t)\lambda(dt),\, g \in L^p[0,1]\}$. To prove the Central Limit Theorem for the triangular array, we will approximate the characteristic function of $S_n(f)$, $\varphi_{S_n(f)}$, by the product of the characteristic functions of the blocks $Y_{j,\ell}$, $j = 1, \ldots, m$. This will be accomplished in five steps, where we take ℓ fixed, except in the final one.

Step 1. We first prove that, as the convergence in distribution is regarded, we may replace the sum with n terms by a sum considering only a multiple of ℓ terms. Using $|e^{it} - e^{is}| \le |t - s|$ for $t, s \in \mathbb{R}$, and the Cauchy–Schwarz inequality, it follows that

$$\left|\varphi_{S_n(f)}(t) - \varphi_{S_{m\ell}(f)}(t)\right| \le |t| \operatorname{Var}^{1/2}\bigl(S_n(f) - S_{m\ell}(f)\bigr). \tag{5.18}$$

Recalling that this upper bound is just the $L^2[0,1]$-norm of $S_n(f) - S_{m\ell}(f)$, we still have that

$$\begin{aligned}
&\left|\varphi_{S_n(f)}(t) - \varphi_{S_{m\ell}(f)}(t)\right|^2 \\
&\quad \le \frac{1}{n}\mathrm{E}\left(\sum_{j=1}^{m\ell}\left(F\left(\frac{j}{n}\right) - F\left(\frac{j}{m\ell}\right)\right)X_j\right)^2 \\
&\qquad + \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{m\ell}}\right)^2 \mathrm{E}\left(\sum_{j=1}^{m\ell} F\left(\frac{j}{m\ell}\right)X_j\right)^2 \\
&\qquad + \frac{1}{n}\mathrm{E}\left(\sum_{j=m\ell+1}^{n} F\left(\frac{j}{n}\right)X_j\right)^2.
\end{aligned} \tag{5.19}$$

Using the fact that F is Lipschitzian and the association of the random variables, the first term in this upper bound is, up to the multiplication by a constant depending only on F, less than or equal to

$$\frac{1}{n}\left(\frac{1}{n} - \frac{1}{m\ell}\right)^2 \sum_{j,k=1}^{m\ell} jk\mathrm{E}(X_j X_k) \le \left(\frac{m\ell}{n} - 1\right)^2 \frac{1}{n}\mathrm{E}\bigl(S_{m\ell}^2\bigr) \longrightarrow 0$$

as $n \longrightarrow +\infty$, taking into account (5.15), as, ℓ being fixed, we also have that $m \longrightarrow +\infty$. From the association of the X_n, $n \in \mathbb{N}$, it follows that the second term in (5.19) is bounded above by

$$\|F\|_\infty^2 \left(\frac{m\ell}{n} - 1\right)\frac{1}{m\ell}\mathrm{E}\bigl(S_{m\ell}^2\bigr) \longrightarrow 0$$

as $n \longrightarrow +\infty$, again taking into account (5.15). Finally, the third term in the upper bound in (5.19) is, due to the association of the underlying variables, less than or equal to

$$\frac{\|F\|_\infty^2}{n} \sum_{j,k=m\ell+1}^{n} \operatorname{Cov}(X_j, X_k) = \frac{\|F\|_\infty^2}{n}\mathrm{E}\left(S_n^2 - S_{m\ell}^2 - 2\sum_{j=1}^{m\ell}\sum_{k=m\ell+1}^{n} X_j X_k\right).$$

Obviously, form (5.15) it follows that $\frac{1}{n}\mathrm{E}(S_n^2)$ and $\frac{1}{n}\mathrm{E}(S_{m\ell}^2)$ both converge to σ^2 as $n \longrightarrow +\infty$. Using again the association of the random variables X_n, $n \in \mathbb{N}$, this implies that all the covariances are nonnegative, and from (5.15) it also follows that $\frac{1}{n}\mathrm{E}(\sum_{j=1}^{m\ell}\sum_{k=m\ell+1}^{n} X_j X_k) \longrightarrow 0$ as $n \longrightarrow +\infty$. Thus, collecting these upper bounds and inserting into (5.19), it follows that

$$\lim_{k\to+\infty}\left|\varphi_{S_n(f)}(t) - \varphi_{S_{m\ell}(f)}(t)\right| = 0.$$

Step 2. We now verify that we may approach F by a simple function keeping an approximation of the corresponding characteristic functions. Using the same arguments as in the first upper bounds in Step 1, we have

$$\begin{aligned}&\left|\varphi_{S_{m\ell}(f)}(t) - \mathrm{E}\exp\left(\frac{it}{\sqrt{m\ell}}\sum_{j=1}^{m} F\left(\frac{j}{m}\right)Y_{j,\ell}\right)\right| \\ &\qquad \le |t|\,\mathrm{Var}^{1/2}\left(S_{m\ell}(f) - \frac{1}{\sqrt{m\ell}}\sum_{j=1}^{k} F\left(\frac{j}{m}\right)Y_{j,\ell}\right).\end{aligned}$$

Expanding the variance above and using the Lipschitz property of F, that is, $|F(x) - F(y)| \le \alpha|x-y|$ for some $\alpha > 0$, we easily find that

$$\mathrm{Var}\left(S_{m\ell}(f) - \frac{1}{\sqrt{m\ell}}\sum_{j=1}^{m} F\left(\frac{j}{m}\right)Y_{j,\ell}\right) \le \frac{\alpha^2}{m^3\ell}\sum_{j,k=1}^{m\ell}\mathrm{E}(X_jX_k),$$

hence, from (5.15) we have, for m large enough,

$$\left|\mathrm{E}\exp\bigl(itS_{m\ell}(f)\bigr) - \mathrm{E}\exp\left(it\frac{1}{\sqrt{m\ell}}\sum_{j=1}^{m} F\left(\frac{j}{m}\right)Y_{j,\ell}\right)\right| \le \frac{\sqrt{2}\sigma\alpha|t|}{m} \tag{5.20}$$

which converges to zero as $n \longrightarrow +\infty$.

Step 3. Now we approximate the characteristic function of $(m\ell)^{-1/2} \times \sum_{j=1}^{m} F(j/m)Y_{j,\ell}$ by what we would find if the blocks were independent. Using Theorem 2.37, it follows, taking into account (5.15) and (5.16), that for m large enough,

$$\begin{aligned}&\left|\mathrm{E}\exp\left(it\frac{1}{\sqrt{m\ell}}\sum_{j=1}^{m} F\left(\frac{j}{m}\right)Y_{j,\ell}\right) - \prod_{j=1}^{k}\mathrm{E}\exp\left(\frac{it}{\sqrt{m\ell}}F\left(\frac{j}{m}\right)Y_{j,\ell}\right)\right| \\ &\quad \le \frac{1}{2}\sum_{\substack{j,k=1\\ j\neq k}}^{m}\frac{t^2}{m\ell}\left|F\left(\frac{j}{m}\right)F\left(\frac{m}{m}\right)\right|\mathrm{E}(Y_{j,\ell}Y_{k,\ell}) \\ &\quad \le \frac{t^2\|F\|_\infty^2}{2m\ell}\sum_{\substack{j,k=1\\ j\neq k}}^{m}\mathrm{E}(Y_{j,\ell}Y_{k,\ell}) \le t^2\|F\|_\infty^2\left(\sigma^2 - \frac{a_\ell}{\ell}\right).\end{aligned} \tag{5.21}$$

Step 4. We prove that the product $\prod_{j=1}^{m}\mathrm{E}\exp(it(m\ell)^{-1/2}F(j/m)Y_{j,\ell})$ converges to the characteristic function of a Gaussian distribution where the $Y_{j,\ell}$,

$j = 1, \dots, m$, may be assumed to be independent. Using Lemma 5.21, from (5.16) we have that

$$s_n^2(f) = \frac{1}{m\ell} \sum_{j=1}^{m} F^2\left(\frac{j}{m}\right) \mathrm{E}\left(Y_{j,\ell}^2\right) \longrightarrow \frac{a_\ell}{\ell} \|F\|_2^2.$$

So, to prove the Lindeberg condition for the triangular array $(m\ell)^{-1/2} F(j/m) Y_{j,\ell}$, $j = 1, \dots, m$, $m \in \mathbb{N}$, it is enough to prove that, for every $\varepsilon > 0$,

$$\sum_{j=1}^{m} \int_{\{|F(j/m)||Y_{j,\ell}|>\varepsilon s_n(f)\sqrt{m\ell}\}} \frac{1}{m\ell} F^2\left(\frac{j}{m}\right) Y_{j,\ell}^2 \, d\mathbf{P} \longrightarrow 0. \tag{5.22}$$

For m large enough and applying (4.33), an upper bound for this integral is

$$\begin{aligned} &\frac{\|F\|_\infty^2}{m} \sum_{j=1}^{m} \sum_{m=(j-1)\ell+1}^{j\ell} \int_{\{|X_k|>(\varepsilon/2)\sqrt{a_\ell/\ell}(\|F\|_2/\|F\|_\infty)\sqrt{m/\ell}\}} X_k^2 \, d\mathbf{P} \\ &\quad \leq \frac{\|F\|_\infty^2}{m} \sum_{j=1}^{m\ell} \int_{\{|X_j|>(\varepsilon/(2\ell))\sqrt{a_\ell/\ell}(\|F\|_2/\|F\|_\infty)\sqrt{m\ell}\}} X_j^2 \, d\mathbf{P} \longrightarrow 0, \end{aligned} \tag{5.23}$$

taking into account (5.17).

Step 5. Now summing up inequalities (5.18), (5.20), (5.21) and (5.23), we get, for fixed $\ell \in \mathbb{N}$,

$$\limsup_{n\to+\infty} \left| \mathrm{E} \exp\left(itS_n(f)\right) - \exp\left(-\frac{\sigma^2}{2} t^2 \|F\|_2^2\right) \right| \leq Ct^2 \|F\|_\infty^2 \left(\sigma^2 - \frac{a_\ell}{\ell}\right),$$

and now letting $\ell \longrightarrow +\infty$, we have the Central Limit Theorem that concludes the proof. □

Remark 5.23 The association of the random variables is crucial in the first two steps of the proof, which take care of the control of the deviance from independence, just as in the proof of Theorem 4.1.

Example 5.24 Assumption (5.16) is a sort of relaxed weak stationarity, allowing for some perturbation of a few random variables. Moreover, the assumption puts some control on the covariance structure which is more flexible than the conditions used for the treatment of the $D[0, 1]$ case as, for instance, Theorem 5.14. An example that verifies Theorem 5.22 but not Theorem 5.14 can be constructed as follows: let X_n, $n \in \mathbb{N}$, be stationary and associated random variables such that $\mathrm{E}(X_n) = 0$, $\mathrm{Var}(X_n) = 1$ and $\mathrm{Cov}(X_j, X_\ell) = \gamma(|j - \ell|)$, where $\sum_{n=1}^{\infty} \gamma(n) < \infty$; for each $n \in \mathbb{N}$, define $X_n' = c_n X_n$, where

$$c_n = \begin{cases} q^{1/2} & \text{if } n = 2^q \text{ for some } q \in \mathbb{N}, \\ 1 & \text{otherwise.} \end{cases}$$

The perturbed sequence remains associated. Its coefficient $u'(n)$ is now

$$u'(n) = \sup_{k\geq 1} u_k'(n) \quad \text{with } u_k'(n) = \sum_{j:|j-k|\geq n} c_j c_k \mathrm{Cov}(X_j, X_k).$$

For each $n \in \mathbb{N}$, we can find k large enough such that $k + n = 2^q$ and $q^{1/2}\gamma(n) \geq 1$, so, $\sup_{k\geq 1} u'_k(n) \geq 1$, and $u'(n)$ does not converge to zero.

Next we check conditions (5.15), (5.16) and (5.17) for X'_n, $n \in \mathbb{N}$. Write $S'_n = \sum_{i=1}^n X'_i$ and $Y'_{j,\ell}$ for the blocks relative to the X'_i.

To (5.15). By the stationarity of X_n, $n \in \mathbb{N}$, we have

$$\frac{1}{n}\mathrm{E}\big(S_n^2\big) \longrightarrow \sigma^2 = \mathrm{Var}(X_1) + 2\sum_{n=1}^{\infty}\gamma(n) < \infty.$$

As what regards S'_n, we have

$$\mathrm{E}\big(S_n'^2\big) = \mathrm{E}\big(S_n^2\big) + \sum_{\substack{1\leq i,j\leq n\\ c_i c_j > 1}} (c_i c_j - 1)\mathrm{E}(X_i X_j).$$

The second term is bounded above by

$$\begin{aligned}\sum_{\substack{1\leq i,j\leq n\\ c_i c_j > 1}} c_i c_j \mathrm{E}(X_i X_j) &= \sum_{\substack{1\leq i,j\leq n\\ c_i > 1, c_j > 1}} c_i c_j \mathrm{E}(X_i X_j) + 2\sum_{\substack{1\leq i\leq n\\ c_i = 1}}\sum_{\substack{1\leq j\leq n\\ c_j > 1}} c_j \mathrm{E}(X_i X_j)\\ &= T_1 + T_2,\end{aligned}$$

where

$$T_1 \leq \sum_{\substack{1\leq i,j\leq n\\ c_i > 1, c_j > 1}} c_i c_j = \left(\sum_{i=1, c_i > 1}^{n} c_i\right)^2 = O\big(\ln^3 n\big)$$

and

$$T_2 \leq 4\sigma^2 \sum_{j=1, c_j > 1}^{n} c_j = O\big(\ln^{3/2} n\big).$$

Hence $\frac{1}{n}(T_1 + T_2) \longrightarrow 0$ and $\frac{1}{n}\mathrm{E}(S_n'^2) \longrightarrow \sigma^2$.

To (5.16). Observing that the number of blocs $Y'_{j,\ell}$ having at least one perturbed term is dominated by $\log_2(m\ell)$ and that the variance of such a perturbed block is bounded by $\ell^2 \log_2(m\ell)$, we have the estimate

$$0 \leq \sum_{j=1}^{k}\mathrm{E}\big(Y_{j,\ell}'^2\big) - \sum_{j=1}^{m}\mathrm{E}\big(Y_{j,\ell}'^2\big) \leq \ell^2(\log_2 n)^2,$$

so $\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}\mathrm{E}(Y_{j,\ell}'^2) = a_\ell = \mathrm{E}(S_\ell^2)$.

To (5.17). Obviously,

$$\sum_{i=1, c_i > 1}^{n}\int_{\{c_i|X_i| > \delta\sqrt{n}\}} c_i^2 X_i^2\, d\mathbf{P} \leq \sum_{i=1, c_i > 1}^{n} c_i^2 \mathrm{E}X_i^2 = \sum_{i=1, c_i > 1}^{n} c_i^2 = O\big(\ln^2 n\big),$$

which is enough to prove (5.17) for the perturbed sequence of random variables.

It is easily verified that, assuming the wide sense stationary, condition (5.16) is superfluous, thus we immediately have the following result.

Corollary 5.25 *Let X_n, $n \in \mathbb{N}$, be centred, weakly stationary and associated random variables. Assume that* (5.15) *and* (5.17) *are verified. Then, for every $p > 1$ and $f \in L^p[0,1]$,*

$$\int_0^1 \xi_n(t) f(t) \lambda(dt) \xrightarrow{d} Z_f \sim \mathcal{N}\big(0, \|F\|_2\big),$$

where $F(s) = \int_s^1 f(t)\lambda(dt)$.

The previous results concern only the convergence in distribution of the analogous of the finite-dimensional distributions for $L^2[0,1]$. The corresponding statement about the invariance principle is now obvious.

Theorem 5.26 *Let X_n, $n \in \mathbb{N}$, be random variables fulfilling the assumptions of Theorem* 5.22. *Then, the sequence ξ_n, $n \in \mathbb{N}$, verifies the invariance principle in $L^2[0,1]$.*

Proof It is enough to remark that, due to the association of the underlying variables, (5.15) implies (5.11), so the relative compactness of ξ_n, $n \in \mathbb{N}$, follows. Finally, Theorem 5.22 translates into

$$\int_0^1 \xi_n(t) f(t) \lambda(dt) \xrightarrow{d} \int_0^1 \sigma W(t) f(t) \lambda(dt), \qquad f \in L^2[0,1],$$

so the invariance principle follows. □

Remark 5.27 Notice that one could think of using Theorem 4.8 to try to prove an $L^p[0,1]$ invariance principle. This is not possible, as Theorem 4.8 does not state a Central Limit Theorem for triangular arrays of associated variables, and this is required to deal with the sums

$$S_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n F\left(\frac{i}{n}\right) X_i.$$

5.4 Empirical Processes

Given a sample $X_1, \ldots, X_n$ from a random variable X with distribution function F, there is an obvious interest in describing the probabilistic behaviour of the distance between F and the empirical distribution function $F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{(-\infty,t]}(X_i)$. This is achieved through the process

$$Z_n(t) = \sqrt{n}\big(F_n(t) - F(t)\big), \qquad t \in \mathbb{R}.$$

This random function Z_n is naturally of importance in statistics, thus it has attracted a lot of interest in the literature. As is well known, for theoretical purposes and on what convergence in distribution is regarded, transforming the random variables by their quantile function Q allows us to be more precise about the common distribution of the variables: $Z_n(Q(t))$ has the same distribution as if we do the construction assuming the random variables X_n, $n \in \mathbb{N}$, to be uniformly distributed on $[0, 1]$. So, throughout the section we will assume the random variables to be associated and uniformly distributed on the interval $[0, 1]$. We adapt accordingly the definition of the empirical process.

Definition 5.28 Given the random variables X_n, $n \in \mathbb{N}$, uniformly distributed on $[0, 1]$, the *empirical process* is defined as the sequence of random functions

$$Z_n(t) = \sqrt{n}\big(F_n(t) - t\big), \quad t \in [0, 1], \tag{5.24}$$

where F_n is the empirical distribution function.

If the underlying variables X_n, $n \in \mathbb{N}$, are independent, it was mentioned by Doob [33] and later proved by Donsker [32] that Z_n converges in distribution to a Brownian bridge in the Skorokhod space $D[0, 1]$. As what concerns associated random variables, Newman [70] proved the weak convergence of the finite-dimensional distributions of the empirical process assuming strict stationarity of the X_n and that $\mathrm{Cov}(\mathbb{I}_{\{X_1>x\}}, \mathbb{I}_{\{X_n>y\}}) = \mathbf{P}(X_1 > x, X_n > y) - \mathbf{P}(X_1 > x)\mathbf{P}(X_n > y)$ defines a uniformly convergent series. As $\mathrm{Cov}(X_1, X_n) = \int\int \mathrm{Cov}(\mathbb{I}_{\{X_1>x\}}, \mathbb{I}_{\{X_n>y\}})\,dx\,dy$, as described by Hoeffding's formula (1.2), it is natural to seek for assumptions on the covariances inside the integral. Now, taking into account Corollary 2.36, the uniform convergence of the series defined by $\mathrm{Cov}(\mathbb{I}_{\{X_1>x\}}, \mathbb{I}_{\{X_n>y\}})$ follows from $\sum_{n=1}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_n) < \infty$. Hence, to have an asymptotic result on the empirical process, we should expect that, at least, $\mathrm{Cov}(X_1, X_n) = O(n^{-3})$. The result by Newman [70] only deals with finite-dimensional distributions, thus leaving tightness untreated. This is usually, at least on what the Skorokhod topology is concerned, the hardest step in the proof of asymptotic results. The first complete proof about the limit behaviour of empirical processes on $D[0, 1]$ assuming that the X_n variables are associated was obtained by Yu [109] requiring that $\mathrm{Cov}(X_1, X_n) = O(n^{-b})$ for some $b > \frac{15}{2}$. The approach relied on an extension of the above-mentioned integral representation of the covariances and a subsequent control over moments of the form $\mathrm{E}(Z_n(t) - Z_n(s))^4$, so the classical conditions for tightness, as in Theorem 15.5 in Billingsley [10], could be verified. Improving on moment bounds, relying now on inequalities of the form stated in Theorem 2.18, Shao and Yu [94] weakened the assumption on the covariance, requiring that $b > \frac{3+\sqrt{33}}{2} \approx 4.373$, later improved by Louhichi [63], requiring only that $b > 4$.

Having in mind integral transformations of the empirical process as, for example, the Cramér–von Mises ω^2 test statistic, the Watson statistic [103], the Shepp statistic [95], the Anderson Darling statistics [1], some von Mises functionals and some functionals of the form $\int_0^1 G(t, Z_n(t))\mu(dt)$, it is enough to require only the $L^2[0, 1]$ or $L^p[0, 1]$ topology on the space of paths (for more extensions and examples we refer the reader to Cremers and Kadelka [26]). These spaces have weaker topologies, so we might expect to find convergence under weaker assumptions. In fact, using a special wavelet multiresolution approach Oliveira and Suquet [77] proved the $L^2[0,1]$ convergence assuming that $\mathrm{Cov}(X_1, X_n) = O(n^{-b})$ with $b > 3$, thus achieving the best rate that could be expected. This was later extended to $L^p[0,1]$ by Oliveira and Suquet [79], assuming that $b > \frac{3p}{2}$. Of course, the interest of this extension is limited by the available result on convergence in the Skorokhod space to the case where $p < \frac{8}{3}$, thus leaving essentially the $L^2[0,1]$ space with a weaker assumption.

5.4.1 Convergence in $D[0,1]$

The main step for the proof of the convergence in distribution of the empirical process in $D[0,1]$ is the control of moments of the form $\mathrm{E}(Z_n(t) - Z_n(s))^2$. To achieve this, we will need to be able to control several covariance terms and some approximation properties. We start with a few technical lemmas on the control of covariances of functions of associated and uniformly distributed random variables. Notice that the definition of Z_n involves transforming the random variables through indicator functions of convenient sets. Results like Theorem 2.10 give control on covariances of transformations of the variables through differentiable functions. The lack of differentiability at, at most, countably many points, might be overcome by the usual approximation procedures using smooth functions. But we must have control over continuous transformations. This is the object of the following lemmas, where indicator functions of real intervals are approximated using continuous transformations. We need some notation to proceed. Let $a, b, \delta \in [0,1]$ be given, and define the real valued function

$$f_{a,b,\delta}(x) = \begin{cases} 0 & \text{if } x \le a - \delta, \\ 1 + \frac{x-a}{\delta} & \text{if } a - \delta \le x \le a, \\ 1 & \text{if } a \le x \le b, \\ 1 + \frac{b-x}{\delta} & \text{if } b \le x \le b + \delta, \\ 0 & \text{if } x \ge b + \delta, \end{cases} \tag{5.25}$$

which is, of course, a continuous approximation of $\mathbb{I}_{[a,b]}(x)$.

Lemma 5.29 *Let X_1 and X_2 be associated random variables uniformly distributed on $[0,1]$. Assume that $a, b \in [0,1]$, $b > a$ and $0 \le \delta \le b - a$. Then, the following inequalities hold*:

(a) *If* $\delta \ge \min(\mathrm{Cov}^{1/4}(X_1, X_2), (b-a)^{-1/3}\,\mathrm{Cov}^{1/3}(X_1, X_2))$,
$$\left|\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\right| \le \sqrt{6(b-a)}\,\mathrm{Cov}^{1/4}(X_1, X_2).$$

(b) *Assume that* $\delta < \min(\mathrm{Cov}^{1/4}(X_1, X_2), (b-a)^{-1/3}\,\mathrm{Cov}^{1/3}(X_1, X_2))$.
   (i) *If* $\mathrm{Cov}(X_1, X_2) \le (b-a)^4$,
$$\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \le 9(b-a)\,\mathrm{Cov}^{1/4}(X_1, X_2).$$

(ii) *If* $(b-a)^4 \le \mathrm{Cov}(X_1, X_2) \le (b-a)^2$,

$$\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \le 9\,\mathrm{Cov}^{1/2}(X_1, X_2).$$

(iii) *If* $(b-a)^2 \le \mathrm{Cov}(X_1, X_2)$,

$$\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \le 5(b-a).$$

(iv) *For all the previous three cases*,

$$-\,\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \le 18(b-a)^{2/3}\,\mathrm{Cov}^{1/3}(X_1, X_2).$$

(c) *Moreover, when* $b = a$,

$$\big|\mathrm{Cov}\big(f_{a,a,\delta}(X_1), f_{a,a,\delta}(X_2)\big)\big| \le 2\big(\mathrm{E} f^2_{a,a,\delta}(X_1)\big)^{1/2}\,\mathrm{Cov}^{1/4}(X_1, X_2).$$

Proof As the random variables are associated, taking into account Theorem 2.10, the following inequality is obvious:

$$\big|\mathrm{Cov}\big(f(X_1), f(X_2)\big)\big| \le \min\big(2\mathrm{E} f^2(X_1), \|f'\|^2_\infty\,\mathrm{Cov}(X_1, X_2)\big). \tag{5.26}$$

Proof of (a). As $\|f'_{a,b,\delta}\|_\infty = \delta^{-1}$, it follows that:

(i) If $\delta \ge \mathrm{Cov}^{1/4}(X_1, X_2)$ then, from (5.26) it follows that

$$\begin{aligned}\big|\mathrm{Cov}&\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\big| \\ &\le \min\big(2\mathrm{E} f^2_{a,b,\delta}(X_1), \mathrm{Cov}^{1/2}(X_1, X_2)\big) \\ &\le \sqrt{2}\big(\mathrm{E} f^2_{a,b,\delta}(X_1)\big)^{1/2}\,\mathrm{Cov}^{1/4}(X_1, X_2).\end{aligned}$$

Now, recalling that the variables are uniformly distributed on [0, 1], it is easily checked that $\mathrm{E} f^2_{a,b,\delta}(X_1) \le 2\delta + b - a \le 3(b-a)$.

(ii) If $\delta \ge (b-a)^{-1/3}\,\mathrm{Cov}^{1/3}(X_1, X_2)$, then, again from (5.26) it follows directly that

$$\big|\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\big| \le (b-a)^{2/3}\,\mathrm{Cov}^{1/3}(X_1, X_2).$$

As $b - a \ge \delta$, it follows that $\mathrm{Cov}^{1/3}(X_1, X_2) \le (b-a)^{4/3}$, so

$$\begin{aligned}\big|\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\big| &\le (b-a)\,\mathrm{Cov}^{1/4}(X_1, X_2) \\ &\le (b-a)^{1/2}\,\mathrm{Cov}(X_1, X_2),\end{aligned}$$

as $b - a < 1$.

Thus, (a) is proved.

Proof of (b). Denote, for simplicity, $\eta = \mathrm{Cov}^{1/4}(X_1, X_2)$, so $\delta < \eta$. Hence $f_{a,b,\delta} \le f_{a,b,\eta}$, and, taking into account that the variables are associated, we have

$$\begin{aligned}\mathrm{Cov}&\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \\ &\le \mathrm{Cov}\big(f_{a,b,\eta}(X_1), f_{a,b,\eta}(X_2)\big) \\ &\quad + \mathrm{E} f_{a,b,\eta}(X_1)\mathrm{E} f_{a,b,\eta}(X_2) - \mathrm{E} f_{a,b,\delta}(X_1)\mathrm{E} f_{a,b,\delta}(X_2).\end{aligned}$$

As the variables are uniformly distributed on [0, 1], $\mathrm{E} f_{a,b,\eta}(X_1) \le 2\eta + b - a$ and $\mathrm{E} f_{a,b,\delta}(X_1) \ge b - a$, and likewise for X_2. So, recalling (5.26), we have

$$\begin{aligned}
&\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\\
&\quad \le \frac{1}{\eta^2}\mathrm{Cov}(X_1, X_2) + (2\eta + b - a)^2 - (b-a)^2\\
&\quad = 5\,\mathrm{Cov}^{1/2}(X_1, X_2) + 4(b-a)\,\mathrm{Cov}^{1/4}(X_1, X_2).
\end{aligned} \tag{5.27}$$

Proof of (b)(i). Assume that $\mathrm{Cov}(X_1, X_2) \le (b-a)^4$. Then, the upper bound in (5.27) is less than or equal to $9(b-a)\,\mathrm{Cov}^{1/4}(X_1, X_2)$.

Proof of (b)(ii). If we assume that $(b-a)^4 \le \mathrm{Cov}(X_1, X_2) \le (b-a)^2$, then the upper bound in (5.27) is less than or equal to $9\,\mathrm{Cov}^{1/2}(X_1, X_2)$.

Proof of (b)(iii). Remark first that, taking into account that $0 \le f_{a,b,\delta} \le 1$, we have

$$\begin{aligned}
\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) &\le \mathrm{E}\big(f_{a,b,\delta}(X_1) f_{a,b,\delta}(X_2)\big) \le \mathrm{E} f_{a,b,\delta}(X_1)\\
&\le b - a + 2\delta \le 3(b-a),
\end{aligned}$$

as the variables are both uniformly distributed on [0, 1]. So, we still may write

$$\begin{aligned}
&\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\\
&\quad \le 5\min\big(\mathrm{Cov}^{1/2}(X_1, X_2) + (b-a)\,\mathrm{Cov}^{1/4}(X_1, X_2), b-a\big)\\
&\quad \le 5(b-a).
\end{aligned}$$

Proof of (b)(iv). We now verify the bound for $-\mathrm{Cov}(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2))$. The easy case is where $(b-a)^4 \le 8\,\mathrm{Cov}(X_1, X_2)$:

$$\begin{aligned}
&-\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\\
&\quad \le \big(\mathrm{E} f_{a,b,\delta}(X_1)\big)^2 \le 9(b-a)^2 \le 18(b-a)^{2/3}\,\mathrm{Cov}^{1/3}(X_1, X_2).
\end{aligned}$$

If $(b-a)^4 > 8\,\mathrm{Cov}(X_1, X_2)$, choose $\eta \in [\delta, \frac{b-a}{2})$. Then, obviously,

$$f_{a+\eta,b-\eta,\eta} \le f_{a,b,\delta}, \qquad \mathrm{E} f_{a+\eta,b-\eta,\eta}(X_1) \ge (b-a) - 2\eta,$$

and

$$\big\| f'_{a+\eta,b-\eta,\eta} \big\|_\infty = \eta^{-1},$$

so, due to the association of the variables,

$$\begin{aligned}
&-\mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big)\\
&\quad \le -\mathrm{Cov}\big(f_{a+\eta,b-\eta,\eta}(X_1), f_{a+\eta,b-\eta,\eta}(X_2)\big)\\
&\qquad - \mathrm{E} f_{a+\eta,b-\eta,\eta}(X_1)\mathrm{E} f_{a+\eta,b-\eta,\eta}(X_2) + (b - a + 2\delta)^2\\
&\quad \le \frac{1}{b^2}\mathrm{Cov}(X_1, X_2) - \big((b-a) - 2\eta\big)^2 + (b - a + 2\delta)^2\\
&\quad \le \frac{1}{b^2}\mathrm{Cov}(X_1, X_2) + 8(b-a)\eta,
\end{aligned}$$

by the choice of η. If we now take $\eta = (b-a)^{-1/3}\operatorname{Cov}^{1/3}(X_1, X_2)$, which verifies the assumption made above, it follows that

$$-\operatorname{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_2)\big) \leq 18(b-a)^{2/3}\operatorname{Cov}^{1/3}(X_1, X_2),$$

so the proof of (b)(iv) is complete.

Proof of (c). As before, the proof is divided in two cases.

(i) If $\delta > \operatorname{Cov}^{1/4}(X_1, X_2)$ then, proceeding as in (a) for the corresponding case,

$$\big|\operatorname{Cov}\big(f_{a,a,\delta}(X_1), f_{a,a,\delta}(X_2)\big)\big| \leq \sqrt{2}\big(\mathrm{E} f_{a,a,\delta}^2(X_1)\big)^{1/2}\operatorname{Cov}^{1/4}(X_1, X_2).$$

(ii) If $\delta \leq \operatorname{Cov}^{1/4}(X_1, X_2)$, then

$$\begin{aligned}\operatorname{Cov}\big(f_{a,a,\delta}(X_1), f_{a,a,\delta}(X_2)\big) &\leq \mathrm{E} f_{a,a,\delta}^2(X_1) = \frac{2\delta}{3} \leq \sqrt{\frac{2\delta}{3}}\operatorname{Cov}^{1/4}(X_1, X_2)\\ &= \big(\mathrm{E} f_{a,a,\delta}^2(X_1)\big)^{1/2}\operatorname{Cov}^{1/4}(X_1, X_2)\end{aligned}$$

and

$$\begin{aligned}-\operatorname{Cov}\big(f_{a,a,\delta}(X_1), f_{a,a,\delta}(X_2)\big) &\leq \big(\mathrm{E} f_{a,a,\delta}(X_1)\big)^2 \leq 2\delta\big(\mathrm{E} f_{a,a,\delta}^2(X_1)\big)^{1/2}\\ &\leq 2\big(\mathrm{E} f_{a,a,\delta}^2(X_1)\big)^{1/2}\operatorname{Cov}^{1/4}(X_1, X_2).\end{aligned}$$ □

We may now obtain an upper bound for covariances of the variables transformed by the functions $f_{a,b,\delta}$, assuming a decrease rate of the covariances of the original variables.

Lemma 5.30 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables uniformly distributed on* $[0, 1]$, *and $a, b, \delta \in [0, 1]$. Assume that, for some constant $C > 0$,*

$$\operatorname{Cov}(X_1, X_n) \leq Cn^{-b}, \quad b > 4. \tag{5.28}$$

Then, there exists a constant $C^* > 0$, independent of a, b and δ, such that

$$\sum_{n=1}^{\infty}\big|\operatorname{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_n)\big)\big| \leq C^*\big(\mathrm{E} f_{a,b,\delta}(X_1)\big)^{1/2}, \tag{5.29}$$

where either $0 \leq \delta \leq b-a$, or $b = a$ and $\delta > 0$.

Proof We will divide the summation in (5.29) according to the cases considered in Lemma 5.29. Define

$$I = \big\{i \leq n : \delta < \min\big(\operatorname{Cov}^{1/4}(X_1, X_i), (b-a)^{-1/3}\operatorname{Cov}^{1/3}(X_1, X_i)\big)\big\}.$$

Taking into account Lemma 5.29(a), we have

$$\sum_{i \notin I}\operatorname{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_n)\big) \leq \sqrt{6(b-a)}\sum_{i \notin I}\operatorname{Cov}^{1/4}(X_1, X_i).$$

For the control of the summation when $i \in I$, we further divide this set into I_1, I_2 and I_3 according to which of the conditions mentioned in (b)(i), (b)(ii) or (b)(iii) from Lemma 5.29 is verified. Then

$$\begin{aligned}
&\sum_{i\in I} \mathrm{Cov}\big(f_{a,b,\delta}(X_1), f_{a,b,\delta}(X_n)\big) \\
&\quad\le 18\Bigg[(b-a)^{2/3}\sum_{i\in I}\mathrm{Cov}^{1/3}(X_1,X_i)+\sum_{i\in I_1}(b-a)\,\mathrm{Cov}^{1/4}(X_1,X_i) \\
&\qquad + \sum_{i\in I_2}\mathrm{Cov}^{1/2}(X_1,X_i)+\sum_{i\in I_3}(b-a)\Bigg] \\
&\quad\le 18(b-a)^{1/2}\Bigg[\sum_{i\in I}\mathrm{Cov}^{1/3}(X_1,X_i)+\sum_{i\in I_1}\mathrm{Cov}^{1/4}(X_1,X_i) \\
&\qquad + \sum_{i\in I_2\cup I_3}\mathrm{Cov}^{1/2}(X_1,X_i)\Bigg] \\
&\quad\le C^*\big(\mathrm{E}f^2_{a,b,\delta}(X_1)\big)^{1/2},
\end{aligned}$$

where $C^* = 18\sum_{n=1}^{\infty}[\mathrm{Cov}^{1/3}(X_1,X_i)+\mathrm{Cov}^{1/4}(X_1,X_i)+\mathrm{Cov}^{1/2}(X_1,X_i)] < \infty$, taking into account that $b-a \le \mathrm{E}f^2_{a,b,\delta}(X_1)$. □

We have obtained a way to control covariances between transformations through functions $f_{a,b,\delta}$. We still have to control how the approximations are affected through covariances.

Definition 5.31 Given $\delta > 0$, define $\mathcal{C}(\delta)$ as the smallest nonnegative integer n such that there exist $f_1,\dots,f_n$ bounded by 1 with first derivatives bounded by δ^{-1} so that given any $t\in[0,1]$, there exist $i,j\in\{1,\dots,n\}$ verifying $f_i \le \mathbb{I}_{(-\infty,t]} \le f_j$ and $\mathrm{E}(f_j(X_1)-f_i(X_1)) \le C\delta$, where $C>0$ depends only on the distribution of X_1. Denote by $\mathfrak{I}_\delta$ a set of functions $f_1,\dots,f_{\mathcal{C}(\delta)}$ that fulfills the previous condition.

Before the main result of this section, we still need a maximal inequality, which is an extension of the approximation given in Theorem 2.2 in Andrews and Pollard [2]. In order to simplify somewhat the notation, define, for a real function f and $S_n(f) = f(X_1)+\cdots+f(X_n)$, $\rho(f) = \|f(X_1)\|_2$. Notice that, with this notation, we may rewrite $Z_n(t) = \frac{1}{\sqrt{n}}(S_n(\mathbb{I}_{(-\infty,t]}) - \mathrm{E}S_n(\mathbb{I}_{(-\infty,t]}))$. That is, we may think of the empirical process as a functional defined on the family of indicator functions. It is then natural to extend the definition of the empirical process to more general functions than just indicators: given a real function f, define

$$Z_n(f) = \frac{1}{\sqrt{n}}\big(S_n(f) - \mathrm{E}S_n(f)\big).$$

We first prove two general inequalities. The first one is a reduced version, adapted to our purposes, of a more general inequality proved by Pisier [81].

Lemma 5.32 *Let $X_1, \ldots, X_n$ be random variables with finite rth moments. Then*

$$\Big\| \max_{1 \le i \le n} |X_i| \Big\|_r \le n^{1/r} \max_{1 \le i \le n} \|X_i\|_r. \tag{5.30}$$

Proof As $\psi(x) = x^r$ is convex and increasing, using Jensen's inequality, we have that

$$\begin{aligned} \Big(\mathrm{E}\Big(\max_{1 \le i \le n} |X_i| \Big) \Big)^r &\le \mathrm{E}\Big(\max_{1 \le i \le n} |X_i| \Big)^r \\ &= \mathrm{E}\Big(\max_{1 \le i \le n} |X_i|^r \Big) \le \mathrm{E}\left(\sum_{i=1}^n |X_i|^r \right) \le n \max_{1 \le i \le n} |X_i|^r, \end{aligned}$$

which concludes the proof. □

Lemma 5.33 *Let X_n, $n \in \mathbb{N}$, be stationary random variables such that for some $p > 2$ and $\mu > 0$, there exists a constant $C_{p,\mu}$ such that, for all $h, g \in \mathfrak{I}_{2^{-k}}$,*

$$\mathrm{E}\big| S_n(h-g) - \mathrm{E} S_n(h-g) \big|^p \le C_{p,\mu} \big(2^{2k} n^{1+\mu} + n^{p/2} \rho^{p/2}(h-g) \big). \tag{5.31}$$

Assume that, for some sequence $a_n \searrow 0$,

$$\lim_{n \to +\infty} \frac{n^{2-p/2+\mu}}{a_n^2} \mathcal{C}\left(\frac{a_n}{\sqrt{n}} \right) = 0 \quad \textit{and} \quad \int_0^1 x^{-3/4} \mathcal{C}^{1/p}(x)\, dx < \infty. \tag{5.32}$$

Then, for all $\varepsilon > 0$ and $\delta > 0$, there exists $m \in \mathbb{N}$, depending only on ε, p and $\mathcal{C}$, such that

$$\Big\| \sup_{|t-s| \le \delta} |Z_n(t) - Z_n(s)| \Big\|_p \le \varepsilon + \mathcal{C}^{2/p}\big(2^{-m}\big) \sup_{|t-s| \le 2\delta} \big\| Z_n(t) - Z_n(s) \big\|_p. \tag{5.33}$$

Proof We first prove that, for some suitably chosen sequence $k_n \longrightarrow +\infty$, there exist functions $f_{k_n} \in \mathfrak{I}_{2^{-k_n}}$ such that

$$\begin{aligned} \lim_{n \to +\infty} \Big\| \sup_{t \in [0,1]} \big| Z_n(t) - Z_n(f_{k_n}) \big| \Big\|_p &= \lim_{n \to +\infty} \Big\| \sup_{t \in [0,1]} \big| Z_n(\mathbb{I}_{(-\infty,t]}) - Z_n(f_{k_n}) \big| \Big\|_p \\ &= 0. \end{aligned} \tag{5.34}$$

Given $t \in [0,1]$ and $k \in \mathbb{N}$, choose functions $f_k, g_k \in \mathfrak{I}_{2^{-k}}$ such that $f_k \le \mathbb{I}_{(-\infty,t]} \le g_k$ and $\mathrm{E}(f_k(X_1) - g_k(X_1))^2 \le C 2^{-k}$, where C is the constant from Definition 5.31. It is easy to verify that

$$Z_n(t) - Z_n(f_k) \le \big| Z_n(g_k) - Z_n(f_k) \big| + C 2^{-k} \sqrt{n}$$

and, since $f_k \le \mathbb{I}_{(-\infty,t]}$,

$$Z_n(f_k) - Z_n(t) = \frac{1}{\sqrt{n}} \big(S_n(f_k - \mathbb{I}_{(-\infty,t]}) - \mathrm{E} S_n(f_k - \mathbb{I}_{(-\infty,t]}) \big) \le C 2^{-k} \sqrt{n}.$$

Thus, it follows that

$$\sup_{t \in [0,1]} \big| Z_n(t) - Z_n(f_k) \big| \le \max_{f_k \in \mathfrak{I}_{2^{-k}}} \big| Z_n(g_k) - Z_n(f_k) \big| + C 2^{-k} \sqrt{n}.$$

Applying now the maximal inequality (5.30), we have

$$\Big\| \sup_{t\in[0,1]} \big|Z_n(t)-Z_n(f_k)\big| \Big\|_p \le \mathcal{C}^{1/p}\big(2^{-k}\big) \max_{f_k\in\mathfrak{I}_{2^{-k}}} \big\|Z_n(g_k)-Z_n(f_k)\big\|_p + C2^{-k}\sqrt{n},$$

so (5.31) finally implies that

$$\begin{aligned}\Big\| \sup_{t\in[0,1]} \big|Z_n(t)-Z_n(f_k)\big| \Big\|_p \\ \le C_{p,\mu}\mathcal{C}^{1/p}\big(2^{-k}\big)\big(2^{2k/p}n^{(1+\mu)/p-1/2}+2^{-k/4}\big)+C2^{-k}\sqrt{n}.\end{aligned}$$

Choose a sequence k_n such that $2^{k_n}=\frac{\sqrt{n}}{a_n}$. As $a_n \searrow 0$, we have $k_n \longrightarrow +\infty$. It now follows from (5.32) that

$$\mathcal{C}^{1/p}\big(2^{-k}\big)\big(2^{2k/p}n^{(1+\mu)/p-1/2}+2^{-k/4}\big)+C2^{-k}\sqrt{n}\longrightarrow 0,$$

thus proving (5.34).

Let $m\in\mathbb{N}$ be fixed. Then, for n large enough, we will have $k_n\ge m$, and we want to control $\|\sup_{t\in[0,1]}|Z_n(f_{k_n})-Z_n(f_m)|\|_p$. Notice first that the sup is really attained as we are approximating within a finite range of functions. Consider intermediate functions $h_i\in\mathfrak{I}_{2^{-i}}$, $i=k_n-1,\ldots,m+1$, $h_m=f_m$, $h_{k_n}=f_{k_n}$ such that $\mathrm{E}(h_i-h_{i-1})^2\le C2^{-i+1}$ (such functions do exist, following Definition 5.31). Now obviously

$$\Big\| \sup_{t\in[0,1]} \big|Z_n(f_{k_n})-Z_n(f_m)\big| \Big\|_p \le \sum_{i=m+1}^{k_n} \Big\| \sup_{g_i\in\mathfrak{I}_{2^{-i}}} \big|Z_n(h_i)-Z_n(h_{i-1})\big| \Big\|_p.$$

Using again (5.30) and (5.31) and taking into account that $\mathcal{C}(2^{-\ell})$ increases with ℓ, we obtain

$$\begin{aligned}&\Big\| \sup_{t\in[0,1]} \big|Z_n(f_{k_n})-Z_n(f_m)\big| \Big\|_p \\ &\le C_{p,\mu}\sum_{i=m+1}^{k_n} \mathcal{C}^{1/p}\big(2^{-i}\big)\big(2^{2i/p}n^{(1+\mu)/p-1/2}+2^{-i/4}\big) \\ &\le C_{p,\mu}\sum_{i=m+1}^{k_n} \big(\mathcal{C}^{1/p}\big(2^{-k_n}\big)2^{2i/p}n^{(1+\mu)/p-1/2}+\mathcal{C}^{1/p}\big(2^{-i}\big)2^{-i/4}\big).\end{aligned}$$

The first term above behaves like $\mathcal{C}^{1/p}(2^{-k_n})2^{(2k_n+1)/p}n^{(1+\mu)/p-1/2}$ and thus converges to 0, by the first assumption in (5.32), while the second term is, up to multiplication by a constant, bounded above by $\int_0^{2^{-m}} x^{-3/4}\mathcal{C}^{1/p}(x)\,dx$ and thus may be made arbitrarily small by choosing m large enough. That is, given any $\varepsilon>0$, we may choose m such that for n large enough,

$$\Big\| \sup_{t\in[0,1]} \big|Z_n(f_{k_n})-Z_n(f_m)\big| \Big\|_p \le \varepsilon.$$

The previous approximations imply that, given $\varepsilon>0$, there exists m such that

$$\lim_{n\to+\infty} \Big\| \sup_{t\in[0,1]} \big|Z_n(t)-Z_n(f_m)\big| \Big\|_p \le \varepsilon. \tag{5.35}$$

Define an equivalence relation on $[0, 1]$ as follows: $t \sim s$ if $f_m = f'_m$, where f_m and f'_m are the functions considered in Definition 5.31 corresponding to t and s, respectively. This partitions $[0, 1]$ into $\mathcal{C}(2^{-m})$ equivalence classes $E_1, \ldots, E_{\mathcal{C}(2^{-m})}$. Using (5.35) twice, we have that, for n large enough,

$$\Big\| \sup_{t,s\in E_i} \big|Z_n(t) - Z_n(s)\big| \Big\|_p \le 2\varepsilon, \quad i = 1, \ldots, E_{\mathcal{N}(2^{-m})}.$$

Define $d(E_i, E_j) = \inf\{|t - s|, t \in E_i, s \in E_j\}$ and, given $\delta > 0$, choose $t_{ij} \in E_i$ and $t_{ji} \in E_j$ such that $|t_{ij} - t_{ji}| < d(E_i, E_j) + \delta$. Then, if $t \in E_i$, $s \in E_j$ are such that $|t - s| < \delta$, it follows that $|t_{ij} - t_{ji}| < 2\delta$ and

$$\begin{aligned}\big|Z_n(t) - Z_n(s)\big| &\le 2 \sup_{u\sim v} \big|Z_n(u) - Z_n(v)\big| \\ &\quad + \max_{1\le i,j\le\mathcal{C}(2^{-m})} \big\{\big|Z_n(t_{ij}) - Z_n(t_{ji})\big| : |t_{ij} - t_{ji}|\big\}.\end{aligned}$$

Taking norms, we get (5.33), thus concluding the proof of the lemma. □

The maximal inequality proved in Lemma 5.33 suggests the need of controlling $\|Z_n(t) - Z_n(s)\|_p$.

Lemma 5.34 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables uniformly distributed on $[0, 1]$ such that $\mathrm{Cov}(X_1, X_n) = O(n^{-b})$ for some $b > 1$. Then, for each $p \ge 2$ and small enough $\eta > 0$, there exists a constant $K > 0$, independent of n, t and s, such that,*

$$\begin{aligned}&\mathrm{E}\big|Z_n(t) - Z_n(s)\big|^p \\ &\quad\le K\left[n^{-p(p-4-2\eta)/(2(p+2))} + \left(\sum_{j=1}^n \mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1), \mathbb{I}_{[s,t]}(X_j)\big)\right)^{p/2}\right].\end{aligned}$$

Remark 5.35 Before embarking in the proof of Lemma 5.34, we make a comment on the assumption on the covariance decay rate connected to the use of Theorem 2.18 in the proof. The assumption $\mathrm{Cov}(X_1, X_n) = O(n^{-b})$ implies that $u(n) = O(n^{-b+1})$. In order to simplify the use of Theorem 2.18, we will take, with the notation corresponding to this theorem, $r = +\infty$ and $p = b + 1$. This choice makes the exponent in the third term of (2.5) equal to the exponent in the first term. So, taking into account that the functions $f_{s,t,h}$ and $f_{s+h,t-h,h}$ are bounded by 1 and their derivatives uniformly bounded by $\frac{1}{h}$, the upper bound that follows from Theorem 2.18 may be rewritten as

$$\begin{aligned}&\mathrm{E}\left|\sum_{i=1}^n \big(f_{s,t,h}(X_i) - \mathrm{E} f_{s,t,h}(X_i)\big)\right|^p \\ &\quad\le K\left[\frac{n^{1+\eta}}{h^2} + \left(n \max_{i\le n} \sum_{j=1}^n \big|\mathrm{Cov}\big(f_{s,t,h}(X_1), f_{s,t,h}(X_j)\big)\big|\right)^{p/2}\right] \qquad (5.36)\end{aligned}$$

for some $\eta > 0$ small enough. Of course, an analogous upper bound holds for $f_{s+h,t-h,h}$. These bounds will be used in the proof of Lemma 5.34.

Proof of Lemma 5.34 Given $0 \le s < t \le 1$ and $0 < h < \frac{t-s}{2}$, consider the two functions $f_{s,t,h}$ and $f_{s+h,t-h,h}$, approximating $\mathbb{I}_{(s,t]}$ from above and below, respectively, that is $f_{s+h,t-h,h}(x) \le \mathbb{I}_{(s,t]}(x) \le f_{s,t,h}(x)$. Moreover,

$$0 \le f_{s,t,h}(x) - f_{s+h,t-h,h}(x) \le \mathbb{I}_{(s-h,s+h)}(x) - \mathbb{I}_{(t-h,t+h)}(x).$$

It is also clear that, for all $q \ge 1$ and $n \in \mathbb{N}$,

$$\begin{aligned} \mathrm{E}\big|f_{s,t,h}(X_n) - \mathrm{E}f_{s,t,h}(X_n) - \big(\mathbb{I}_{(s,t](X_n)} - (t-s)\big)\big|^q &\le 2h, \\ \mathrm{E}\big|f_{s+h,t-h,h}(X_n) - \mathrm{E}f_{s+h,t-h,h}(X_n) - \big(\mathbb{I}_{(s,t](X_n)} - (t-s)\big)\big|^q &\le 2h. \end{aligned} \tag{5.37}$$

It follows then that

$$\begin{aligned} \left|\sum_{i=1}^n \big(\mathbb{I}_{(s,t](X_i)} - (t-s)\big)\right| \le 4nh + \left|\sum_{i=1}^n \big(f_{s,t,h}(X_i) - \mathrm{E}f_{s,t,h}(X_i)\big)\right| \\ + \left|\sum_{i=1}^n \big(f_{s+h,t-h,h}(X_i) - \mathrm{E}f_{s+h,t-h,h}(X_i)\big)\right|. \end{aligned}$$

Taking the power p and mathematical expectations on the expression above means that we need to control $\mathrm{E}|\sum_{i=1}^n (f_{s,t,h}(X_i) - \mathrm{E}f_{s,t,h}(X_i))|^p$ and likewise with the function $f_{s+h,t-h,h}$. For this purpose, we use Theorem 2.18 as argued in Remark 5.35, so that (5.36) holds. Hence, we need to control the sums of covariances appearing on the right side of (5.36). Taking into account the stationarity, we have

$$\begin{aligned} &\max_{i \le n} \sum_{j=1}^n \big|\mathrm{Cov}\big(f_{s,t,h}(X_1), f_{s,t,h}(X_j)\big)\big| \\ &\quad \le \mathrm{Var}\big(f_{s,t,h}(X_1)\big) + 2\sum_{i=2}^n \big|\mathrm{Cov}\big(f_{s,t,h}(X_1), f_{s,t,h}(X_i)\big)\big|. \end{aligned} \tag{5.38}$$

Now, using (5.37), we have

$$\begin{aligned} \mathrm{Var}\big(f_{s,t,h}(X_1)\big) &\le 2\,\mathrm{Var}\big(\mathbb{I}_{(s,t]}(X_1)\big) + 4h, \\ \big|\mathrm{Cov}\big(f_{s,t,h}(X_1), f_{s,t,h}(X_i)\big)\big| &\le 2\big|\mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1), \mathbb{I}_{(s,t]}(X_i)\big)\big| + 4h. \end{aligned}$$

Inserting this in (5.38) and using Theorem 2.10, we find that

$$\begin{aligned} &\max_{i \le n} \sum_{j=1}^n \big|\mathrm{Cov}\big(f_{s,t,h}(X_1), f_{s,t,h}(X_j)\big)\big| \\ &\quad \le 4\Bigg[\mathrm{Var}\big(\mathbb{I}_{(s,t]}(X_1)\big) + h \\ &\qquad + \sum_{i=2}^n \max\bigg(\big|\mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1), \mathbb{I}_{(s,t]}(X_i)\big)\big| + h, \frac{\mathrm{Cov}(X_1, X_i)}{h^2}\bigg)\Bigg] \end{aligned}$$

$$\leq 4\left[\sum_{i=1}^{n}\left|\mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1),\mathbb{I}_{(s,t]}(X_i)\big)\right| + \sum_{i\leq h^{-3/(\nu+\varepsilon)}} h + \sum_{i>h^{-3/(\nu+\varepsilon)}} \frac{i^{-\nu-\varepsilon}}{h^2}\right]$$

$$\leq 4\left[\sum_{i=1}^{n}\left|\mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1),\mathbb{I}_{(s,t]}(X_i)\big)\right| + h^{1-3/(\nu+\varepsilon)}\right].$$

This upper bound is now used on (5.36) to find:

$$\begin{aligned}
&\mathrm{E}\left|\sum_{i=1}^{n}\big(\mathbb{I}_{(s,t](X_i)} - (t-s)\big)\right|^p \\
&\quad\leq 3^{p-1}\left[(4nh)^p + \mathrm{E}\left|\sum_{i=1}^{n}\big(f_{s,t,h}(X_i) - \mathrm{E}f_{s,t,h}(X_i)\big)\right|^p\right. \\
&\qquad\left. + \mathrm{E}\left|\sum_{i=1}^{n}\big(f_{s+h,t-h,h}(X_i) - \mathrm{E}f_{s+h,t-h,h}(X_i)\big)\right|^p\right] \\
&\quad\leq 2\times 4^p\times 3^{p-1}\left[(nh)^p + \frac{n^{1+\eta}}{h^2} + \big(nh^{1-3/(\nu+\varepsilon)}\big)^{p/2}\right. \\
&\qquad\left. + \left(n\sum_{i=1}^{n}\left|\mathrm{Cov}\big(\mathbb{I}_{(s,t]}(X_1),\mathbb{I}_{(s,t]}(X_i)\big)\right|\right)^{p/2}\right].
\end{aligned}$$

Choosing $h = n^{-(p-1+\eta)/(p+2)}$, we conclude the proof. □

The main ingredients for the proof of the convergence of the empirical process are now available. Thus, we may state the main result of this subsection.

Theorem 5.36 *Let X_n, $n\in\mathbb{N}$, be strictly stationary and associated random variables uniformly distributed on* $[0,1]$, *satisfying* (5.28). *Then Z_n converges in distribution in the space $D[0,1]$ to a centred Gaussian process with covariance function*

$$\varGamma(s,t) = s\wedge t - st + 2\sum_{k=2}^{\infty}\mathrm{Cov}\big(\mathbb{I}_{(-\infty,s]}(X_1),\mathbb{I}_{(-\infty,t]}(X_k)\big). \tag{5.39}$$

Proof The weak convergence of the finite-dimensional distributions follows as in the proof of Theorem 5.8, so we are left with the proof of the tightness of the sequence. For this purpose, according to Theorem 15.5 in Billingsley [10], it is enough to prove that, for some $p>2$,

$$\limsup_{n\to+\infty}\left\|\sup_{|t-s|\leq\delta}\left|Z_n(t)-Z_n(s)\right|\right\|_p = 0. \tag{5.40}$$

Denote $\mathcal{I} = \{\mathbb{I}_{(-\infty,t]} : t\in[0,1]\}$, the set of indicator functions. For each $k\in\mathbb{N}$, define $D_k = \{\frac{i}{2^k}, i=1,\ldots,2^k\}$ and, for each $d\in D_k$, the function $f_d(x) = \mathbb{I}_{(-\infty,d-2^{-k}]}(x) - 2^k(x-d)\mathbb{I}_{[d-2^{-k},d]}(x)$. Each f_d is differentiable except at $d-2^{-k}$ and d. Let f_d' be equal to the derivative of f_d where this derivative exists and to 0 at

$d - 2^{-k}$ and d. Then, for $d \in D_k$, $\|f_d\|_\infty \le 1$, $\|f'_d\|_\infty \le 2^k$. Moreover, the following approximations hold:

$$f_d(x) \le \mathbb{I}_{(-\infty,t]}(x) \le f_{d+2^{1-k}}(x) \quad \text{with } d = 2^{-k}\big[t2^k\big] \in D_k,$$
$$\mathrm{E}\big(f_{d+2^{1-k}}(X_1) - f_d(X_1)\big) \le 2^{1-k}.$$

Given the correspondence between the classes of functions considered, the above approximations imply that $\mathcal{C}(x) = O(x^{-1})$. Inserting this into (5.32), these assumptions are verified if $p > 5+2\mu$, so we need to prove that (5.31) holds for some $p > 5$ for our class of functions $\mathcal{I}_{2^{-k}}$. Assuming that $d, d' \in D_k$ are such that $d < d'$, the difference $f_{d'} - f_d$ is a function of the form defined in (5.25). In fact, it is easily checked that, referring to the notation of (5.25), $f_{d'} - f_d = f_{d,d'-2^{-k},2^{-k}}$. Choosing $\nu = b-4$, $p = 5+3\mu$, where $\mu \in (0, \frac{\nu}{3})$, Theorem 2.18 yields

$$\begin{aligned}
&\mathrm{E}\big|S_n(f_d - f_{d'}) - \mathrm{E}S_n(f_d - f_{d'})\big|^p \\
&\quad \le C\left[n^{1+\mu}2^{2k} + n^{p/2}\left(\sum_{j=1}^{\infty}\big|\mathrm{Cov}\big(f_{d,d'-2^{-k},2^{-k}}(X_1), f_{d,d'-2^{-k},2^{-k}}(X_j)\big)\big|\right)^{p/2}\right],
\end{aligned}$$

where the constant C does not depend on the class of functions. From this inequality it follows, by Lemma 5.30, that

$$\mathrm{E}\big|S_n(f_d - f_{d'}) - \mathrm{E}S_n(f_d - f_{d'})\big|^p \le C\big[n^{1+\mu}2^{2k} + C^* n^{p/2}\rho^{p/2}(f_d - f_{d'})\big],$$

that is, (5.31) really holds. So, from Lemma 5.33 it follows that

$$\Big\|\sup_{|t-s|\le\delta}\big|Z_n(t) - Z_n(s)\big|\Big\|_p \le \varepsilon + \mathcal{C}^{2/p}\big(2^{-m}\big)\sup_{|t-s|\le 2\delta}\big\|Z_n(t) - Z_n(s)\big\|_p.$$

We need now to estimate $\|Z_n(t) - Z_n(s)\|_p$, but this was done in Lemma 5.34, from which follows that

$$\begin{aligned}
&\big\|Z_n(t) - Z_n(s)\big\|_p \\
&\quad \le K\left(n^{-(p-4-2\mu)/(2(p+2))} + \left(\sum_{j=1}^{n}\mathrm{Cov}\big(f_{d,d,t-s}(X_1), f_{d,d,t-s}(X_j)\big)\right)^{1/2}\right).
\end{aligned}$$

Then, using Lemma 5.30 again, we have

$$\Big\|\sup_{|t-s|\le\delta}\big|Z_n(t) - Z_n(s)\big|\Big\|_p \le \varepsilon + K\mathcal{C}^{2/p}\big(2^{-m}\big)\big(n^{-(p-4-2\mu)/(2(p+2))} + \delta^{1/2}\big).$$

Therefore, recalling that $p > 5$, we have

$$\limsup_{n\to+\infty}\Big\|\sup_{|t-s|\le\delta}|Z_n(t) - Z_n(s)|\Big\|_p \le \varepsilon + K\mathcal{C}^{2/p}\big(2^{-m}\big)\delta^{1/2}$$

and now choose $\delta > 0$ such that $K\mathcal{C}^{2/p}(2^{-m})\delta^{1/2} \le \varepsilon$. As $\varepsilon > 0$ is arbitrarily chosen, (5.40) follows, so the theorem is proved. □

5.4.2 Convergence in $L^p[0, 1]$

As already argued in Sect. 5.3, one might be interested in establishing the convergence in distribution in a weaker space than the one studied in the previous subsection. The main results concerning the characterization of convergence in distribution have been addressed in Sect. 5.2, so we proceed directly to the proof of the appropriate versions of the convergence of the empirical process. As usual, we will look for assumptions on the decay rate of the covariances, proving the convergence in $L^p[0, 1]$ with weaker conditions than those obtained for $D[0, 1]$, when $p \in [2, \frac{8}{3})$.

Theorem 5.37 *Let X_n, $n \in \mathbb{N}$, be strictly stationary and associated random variables uniformly distributed on* $[0, 1]$. *Assume that*

$$\sum_{n=1}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_n) < \infty. \tag{5.41}$$

Then, the uniform empirical process Z_n converges weakly in $L^2[0, 1]$ to a centred Gaussian process Z with covariance Γ given by (5.39).

Proof We first check the tightness of Z_n, $n \in \mathbb{N}$. As the variables are stationary, defining $g_k(s,t) = \mathrm{Cov}(\mathbb{I}_{(s,t]}(X_1), \mathbb{I}_{(s,t]}(X_k))$, we have

$$\mathrm{E}\big|Z_n(t) - Z_n(s)\big|^2 = g_1(s,t) + 2\sum_{k=2}^{n}\left(1 - \frac{k}{n}\right) g_k(s,t).$$

As the random variables are uniform and associated, we may use now Corollary 2.36 to conclude that there exists a constant $M > 0$, independent of n, satisfying

$$\big|g_k(s,t)\big| \le M\,\mathrm{Cov}^{1/3}(X_1, X_k), \quad s, t \in [0, 1]. \tag{5.42}$$

So, it follows that

$$\begin{aligned} \mathrm{E}\big|Z_n(t) - Z_n(s)\big|^2 &\le g_1(s,t) + 2\sum_{k=2}^{\infty}\big|g_k(s,t)\big| \\ &\le M\left(\mathrm{Var}(X_1) + \sum_{k=2}^{\infty} \mathrm{Cov}^{1/3}(X_1, X_k)\right) < \infty. \end{aligned} \tag{5.43}$$

The functions $g_k(s,t)$ are continuous, so, as the series

$$g(s,t) = g_1(s,t) + 2\sum_{k=2}^{\infty}\big|g_k(s,t)\big|$$

is uniformly absolutely convergent, its sum is a continuous function on $[0, 1]^2$. Moreover, is immediate to check that each g_k vanishes on the diagonal of $[0, 1]^2$, so g vanishes on this diagonal as well. Thus, the following uniform estimate holds:

$$\mathrm{E}\big|Z_n(t+h) - Z_n(t)\big|^2 \le \sup_{0 \le s \le 1} g(s, s+h) = \varepsilon(h), \quad t \in [0, 1], n \ge 1. \tag{5.44}$$

So we have proved that condition (b) of Corollary 5.3 holds, and hence Theorem 5.2(b) is verified. To prove the tightness, it remains to prove that Theorem 5.2(a) also holds. It is a simple matter to check that it suffices to choose $s=0$ in (5.43) to have Theorem 5.2(a) verified.

Now that the tightness of Z_n has been established, to prove the weak $L^2[0,1]$ convergence to Z, it suffices, taking into account Theorem 5.1, to verify the convergence in distribution of $\int_0^1 f(t)Z_n(t)\,dt$ to $\int_0^1 f(t)Z(t)\,dt$ for each $f\in L^2[0,1]$. Observe that we may rewrite

$$\int_0^1 f(t)Z_n(t)\,dt=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\int_{X_i}^1 f(t)\,dt-\mathrm{E}\int_{X_i}^1 f(t)\,dt\right).$$

Thus, we need to verify that a Central Limit Theorem holds for the random variables $\int_{X_i}^1 f(t)\,dt$. Now, as $L^2[0,1]\subset L^1[0,1]$, the function $F(x)=\int_x^1 f(t)\,dt$ is absolutely continuous, so the required Central Limit Theorem follows immediately from Theorem 4.6. □

The corresponding result for $L^p[0,1]$ is as follows.

Theorem 5.38 *Let X_n, $n\in\mathbb{N}$, be strictly stationary and associated random variables uniformly distributed on* $[0,1]$, *and* $p>2$. *Assume that*

$$\mathrm{Cov}(X_1,X_n)=O\bigl(n^{-b}\bigr),\quad b>\frac{3p}{2}. \tag{5.45}$$

Then, the uniform empirical process Z_n converges weakly in $L^p[0,1]$ *to a centred Gaussian process Z with covariance Γ given by* (5.39).

Proof Inequality (5.44) still holds for this case, as the arguments leading to this inequality still apply. To prove Theorem 5.2(b), we will prove that Corollary 5.3(a) also holds. First notice that the random variables $\mathbb{I}_{[0,t]}(X_i)$ are nonincreasing functions of associated variables X_i, so they are associated as well. Recall that $Z_n(t)=n^{-1/2}\sum_{i=1}^n(\mathbb{I}_{[0,t]}(X_i)-t)$. The random variables in this summation are obviously uniformly bounded by 2. On the other hand, taking into account Corollary 2.36, we have that there exists a constant $M>0$, independent of n, such that

$$\begin{aligned}\sum_{k=n}^\infty \mathrm{Cov}\bigl(\mathbb{I}_{[0,t]}(X_1),\mathbb{I}_{[0,t]}(X_k)\bigr)&\le M\sum_{k=n}^\infty \mathrm{Cov}^{1/3}(X_1,X_k)\\&\le M\sum_{k=n}^\infty k^{-b/3}=O\bigl(n^{1-b/3}\bigr).\end{aligned}$$

Now, if we choose $r\in(p,\frac{2b}{3})$, the assumptions of Corollary 2.21 are satisfied, so it follows that $\mathrm{E}|\sum_{i=1}^n(\mathbb{I}_{[0,t]}(X_i)-t)|^r=O(n^{r/2})$, and Corollary 5.3(a) holds. Finally, the finite-dimensional distributions are convergent as in the previous theorem. In fact, it is sufficient now to check that $\int_0^1 f(t)Z_n(t)\,dt\xrightarrow{d}\int_0^1 f(t)Z(t)\,dt$ for each $f \in L^q(0,1)$, the dual space of $L^p[0,1]$. As $L^q[0,1] \subset L^1[0,1]$, the arguments used in the final part of the proof of Theorem 5.37 still hold. □

Getting back to the convergence in $L^2[0,1]$, there is an interesting characterization of the convergence of the uniform empirical process not using assumptions on covariances. This was proved by Morel and Suquet [66] and explores some facts directly depending on the uniform distribution of the random variables. Assume that X and Y are uniformly distributed on $[0,1]$. Then, the following calculation is obvious:

$$\begin{aligned}\int_0^1 &\operatorname{Cov}\big(\mathbb{I}_{(-\infty,t]}(X), \mathbb{I}_{(-\infty,t]}(Y)\big)\,dt \\ &= \int_0^1 \mathbf{P}\big(\max(X,Y) \le t\big) - t^2\,dt \\ &= \int_0^1 1 - \mathbf{P}\big(\max(X,Y) > t\big) - t^2 = \frac{2}{3} - \mathrm{E}\max(X,Y). \end{aligned} \tag{5.46}$$

Theorem 5.39 *Let X_n, $n \in \mathbb{N}$, be stationary and associated random variables uniformly distributed on* $[0,1]$. *The series in*

$$\Gamma(s,t) = s \wedge t - st + 2\sum_{k=2}^{\infty} \operatorname{Cov}\big(\mathbb{I}_{(-\infty,s]}(X_1), \mathbb{I}_{(-\infty,t]}(X_k)\big) \tag{5.47}$$

converges almost everywhere in $[0,1]^2$. *It represents the covariance operator of some Gaussian random process in* $L^2[0,1]$ *if and only if*

$$\sum_{k=2}^{\infty}\left(\frac{2}{3} - \mathrm{E}\max(X_1,X_k)\right) < \infty. \tag{5.48}$$

Moreover, if (5.48) *holds, the sequence of covariance operators*

$$\begin{aligned}\Gamma_n(s,t) &= \mathrm{E}\big(Z_n(s)Z_n(t)\big) \\ &= \mathrm{E}X_1^2 + 2\sum_{k=2}^{n}\left(1-\frac{k}{n}\right)\operatorname{Cov}\big(\mathbb{I}_{(-\infty,s]}(X_1), \mathbb{I}_{(-\infty,t]}(X_k)\big)\end{aligned} \tag{5.49}$$

converges in $L^2([0,1]^2)$ *to* Γ.

Proof If Γ is the covariance operator of a Gaussian random process, then, by Theorem 4.9 in Parthasarathy [80], it has finite trace, that is, it verifies

$$\int_0^1 \Gamma(t,t)\,dt = \sum_{j=1}^{\infty}\left(\frac{2}{3} - \mathrm{E}\max(X_1,X_j)\right) < \infty,$$

and thus, (5.48) is satisfied.

Assume now that (5.48) holds. Using the Cauchy–Schwarz inequality, it follows that $\Gamma_n(s,t)^2 \le \Gamma_n(s,s)\Gamma_n(t,t)$, so, taking into account that, due to the association of the random variables X_n, $\Gamma_n \le \Gamma$, we have

$$\int_0^1\int_0^1 \Gamma_n(s,t)^2\,ds\,dt \le \left(\int_0^1 \Gamma_n(t,t)\,dt\right)^2 \le \left(\int_0^1 \Gamma(t,t)\,dt\right)^2 < \infty.$$

It follows now, using monotone convergence, that

$$\int_0^1\int_0^1 \Gamma_n(s,t)^2\,ds\,dt \le \left(\int_0^1 \Gamma_n(t,t)\,dt\right)^2 < \infty.$$

Hence, Γ is almost everywhere finite on $[0,1]^2$ and, again by the monotonicity of $\Gamma_n(s,t)$ for all $s,t \in [0,1]$, it follows that $\Gamma_n \nearrow \Gamma$ almost everywhere on $[0,1]^2$. Finally, as

$$\int_0^1\int_0^1 \left|\Gamma(s,t) - \Gamma_1(s,t)\right|^2 ds\,dt \le 4\left(\int_0^1 \Gamma(t,t)\,dt\right)^2 < \infty,$$

we can apply the Monotone Convergence Theorem to conclude

$$\int_0^1\int_0^1 \left|\Gamma(s,t) - \Gamma_n(s,t)\right|^2 ds\,dt \longrightarrow 0,$$

so we have the $L^2([0,1]^2)$ convergence to 0 of $\Gamma - \Gamma_n$, and hence, $\Gamma \in L^2([0,1]^2)$. □

Theorem 5.40 *Let X_n, $n \in \mathbb{N}$, be stationary and associated random variables uniformly distributed on* $[0,1]$. *Then, the uniform empirical process Z_n converges weakly in $L^2[0,1]$ to a centred Gaussian process Z with covariance operator Γ defined by* (5.47) *if and only if* (5.48) *holds.*

Proof We start by proving the tightness of the sequence Z_n using Corollary 5.4. Let e_n, $n \in \mathbb{N}$, be an orthonormal basis of $L^2[0,1]$. Let Γ_n be defined by (5.49). Then, due to the association of the variables, (5.46) and (5.48), we have that

$$\mathrm{E}\|Z_n\|_2^2 = \int_0^1 \Gamma_n(t,t)\,dt \le \int_0^1 \Gamma(t,t)\,dt < \infty.$$

To prove condition (b) of Corollary 5.4, denote, as before,

$$r_N^2(f) = \sum_{i=N}^{+\infty}\left(\int_0^1 e_i(t)f(t)\,dt\right)^2.$$

It is obvious that

$$\begin{aligned}\mathrm{E}r_N^2(Z_n) &= \mathrm{E}\|Z_n\|_2^2 - \mathrm{E}\left(\sum_{i=1}^{N-1}\left(\int_0^1 e_i(t)Z_n(t)\,dt\right)^2\right)\\ &= \int_0^1 \Gamma_n(t,t)\,dt - \int_0^1\int_0^1 \sum_{i=1}^{N-1} e_i(s)e_i(t)\Gamma_n(s,t)\,ds\,dt.\end{aligned}$$

The Monotone Convergence Theorem implies that

$$\int_0^1 \Gamma_n(t,t)\,dt \longrightarrow \int_0^1 \Gamma(t,t)\,dt,$$

while the $L^2([0,1]^2)$ convergence of Γ_n to Γ, proved on Theorem 5.39, implies, as $n \longrightarrow +\infty$,

$$\int_0^1 \int_0^1 \sum_{i=1}^{N-1} e_i(s)e_i(t)\Gamma_n(s,t)\,ds\,dt \longrightarrow \int_0^1 \int_0^1 \sum_{i=1}^{N-1} e_i(s)e_i(t)\Gamma(s,t)\,ds\,dt.$$

Thus, it follows that

$$\lim_{n\to+\infty} \mathrm{E} r_N^2(Z_n) = \mathrm{E} r_N^2(Z).$$

Finally, as $\mathrm{E}\|Z\|_2^2 < \infty$, again the Monotone Convergence Theorem implies

$$\lim_{N\to+\infty} \mathrm{E} r_N^2(Z) = 0,$$

concluding the proof. $\square$

Appendix A
General Inequalities

A.1 Berry–Esséen Inequalities

The following is the classical Berry–Esséen inequality, proved independently by Berry [9] and Esséen [37], which is at the basis of the classical bounds for the convergence rates in Central Limit Theorems.

Theorem A.1 *Let F_1 and F_2 be distribution functions with characteristic functions φ_1 and φ_2, respectively, and assume that F_2 has a bounded derivative f_2. Then, for every $T > 0$,*

$$\sup_{x\in\mathbb{R}}\bigl|F_1(x)-F_2(x)\bigr| \leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\varphi_1(t)-\varphi_2(t)}{t}\right| dt + \frac{24}{\pi T}\sup_{x\in\mathbb{R}}\bigl|f_2(x)\bigr|. \tag{A.1}$$

The following two-dimensional extension of the classical Berry–Esséen inequality was proved by Sadikova [91].

Theorem A.2 *Let F and G be two-dimensional distribution functions with characteristic functions φ and ϕ, respectively. Define, for all $u, v \in \mathbb{R}$, $\widehat{\varphi}(u,v) = \varphi(u,v) - \varphi(u,0)\varphi(0,v)$ and $\widehat{\phi}(u,v) = \phi(u,v) - \phi(u,0)\phi(0,v)$. Assume that G has bounded first-order partial derivatives and put $A_1 = \sup \frac{\partial G}{\partial u}$, $A_2 = \sup \frac{\partial G}{\partial v}$. Then, for every $T > 0$,*

$$\begin{aligned}
&\sup_{u,v\in\mathbb{R}}\bigl|F(u,v)-G(u,v)\bigr| \\
&\quad\leq \frac{1}{2\pi^2}\int_{[-T,T]^2}\left|\frac{\widehat{\varphi}(s,t)-\widehat{\phi}(s,t)}{st}\right| ds\,dt \\
&\qquad + 2\sup_{u\in\mathbb{R}}\bigl|F(u,+\infty)-G(u,+\infty)\bigr| + 2\sup_{v\in\mathbb{R}}\bigl|F(+\infty,v)-G(+\infty,v)\bigr| \\
&\qquad + 2(3\sqrt{2}+4\sqrt{3})\frac{A_1+A_2}{T}.
\end{aligned}$$

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8, © Springer-Verlag Berlin Heidelberg 2012

Proof Define the density function

$$\kappa(x) = \frac{3}{8\pi}\left(\frac{\sin(x/4)}{x/4}\right)^4.$$

This density verifies

$$\int x^2\kappa(x)\,dx = 12 \quad \text{and} \quad \int |x|\kappa(x)\,dx < 3\sqrt{2}$$

and has the characteristic function

$$\widehat{\kappa}(t) = \begin{cases} 1 - 6t^2 + 6|t|^3 & \text{if } |t| \le \frac{1}{2}, \\ 2(1-|t|)^3 & \text{if } \frac{1}{2} \le |t| < 1, \\ 0 & \text{if } |t| \ge 1. \end{cases}$$

Define $\lambda > 0$ such that

$$\int_{-\infty}^{-\lambda} \kappa(x)\,dx = \int_{\lambda}^{+\infty} \kappa(x)\,dx = 1 - \frac{\sqrt{3}}{2}.$$

It is easily verified, using Chebyshev's inequality, that $\lambda \le 4\sqrt{3}$. Given $T > 0$, construct the two-dimensional density function $q_T(x, y) = T^2\kappa(Tx)\kappa(Ty)$, whose characteristic function is $\widehat{q_T}(s, t) = \widehat{\kappa}(\frac{s}{T})\widehat{\kappa}(\frac{t}{T})$, which is null outside the square $\max(|s|, |t|) \le T$. Moreover,

$$\begin{aligned} &\int\int |x| q_T\left(x \pm \frac{\lambda}{T}, t \pm \frac{\lambda}{T}\right) dx\,dy \\ &\quad = \int\int |y| q_T\left(x \pm \frac{\lambda}{T}, t \pm \frac{\lambda}{T}\right) dx\,dy \le \frac{3\sqrt{2}}{T} + \frac{\lambda}{T} \end{aligned}$$

and

$$\begin{aligned} &\int\int_{(-\infty,0]\times(-\infty,0]} q_T\left(x + \frac{\lambda}{T}, t + \frac{\lambda}{T}\right) dx\,dy \\ &\quad = \int\int_{[0,+\infty)\times[0,+\infty)} q_T\left(x - \frac{\lambda}{T}, t - \frac{\lambda}{T}\right) dx\,dy = \frac{3}{4}. \end{aligned}$$

Introduce random vectors (X_1, X_2), (Y_1, Y_2) with distribution functions F and G, respectively, a vector (Z_1, Z_2), independent of the previous ones, with density function q_T, and the function

$$R(u, v) = \mathbf{P}(X_1 < u, X_2 < v) - \mathbf{P}(Y_1 < u, Y_2 < v).$$

Let $\gamma = \sup_{u,v\in\mathbb{R}} |R(u, v)| = \sup_{u,v\in\mathbb{R}} R(u, v)$, $\theta = 1$ and

$$R_1(x, y) = \mathbf{P}(X_1 + Z_1 \le x, X_2 + Z_2 \le y) - \mathbf{P}(Y_1 + Z_1 \le x, Y_2 + Z_2 \le y).$$

Given the independence of (Z_1, Z_2) with respect to the other vectors, it follows that

$$R_1\left(x + \frac{\theta\lambda}{T}, y + \frac{\theta\lambda}{T}\right) = \int\int R(x-u, y-v) q_T\left(u + \frac{\theta\lambda}{T}, v + \frac{\theta\lambda}{T}\right) du\,dv.$$

Then, obviously, $\gamma_1 \sup_{u,v\in\mathbb{R}} |R(u,v)| \le \gamma$. Consider now $\varepsilon > 0$ and consider $x_\varepsilon, y_\varepsilon \in \mathbb{R}$ such that $R(x_\varepsilon, y_\varepsilon) \ge \gamma - \varepsilon$. For all $u, v \le 0$, we have

$$\begin{aligned}
R(x_\varepsilon - u, y_\varepsilon - v) &= F(x_\varepsilon - u, y_\varepsilon - v) - G(x_\varepsilon - u, y_\varepsilon - v)\\
&\ge F(x_\varepsilon, y_\varepsilon) - G(x_\varepsilon, y_\varepsilon) - \big(G(x_\varepsilon - u, y_\varepsilon - v) - G(x_\varepsilon, y_\varepsilon)\big)\\
&\ge R(x_\varepsilon, y_\varepsilon) - A_1|u| - A_2|v|\\
&\ge \gamma - \varepsilon - A_1|u| - A_2|v|.
\end{aligned}$$

Multiplying the inequality above by $q_T(u + \frac{\lambda}{T}, v + \frac{\lambda}{T})$ and integrating over the set $(-\infty, 0] \times (-\infty, 0]$, we find

$$\begin{aligned}
J_1 &= \int\int_{(-\infty,0]\times(-\infty,0]} R(x_\varepsilon - u, y_\varepsilon - v) q_T\left(u + \frac{\lambda}{T}, v + \frac{\lambda}{T}\right) du\, dv\\
&\ge \frac{3}{4}(\gamma - \varepsilon) - \frac{A_1 + A_2}{T}(3\sqrt{2} + \lambda)
\end{aligned}$$

and

$$J_2 = \left| \int\int_{\mathbb{R}\setminus(-\infty,0]\times(-\infty,0]} R(x_\varepsilon - u, y_\varepsilon - v) q_T\left(u + \frac{\lambda}{T}, v + \frac{\lambda}{T}\right) du\, dv \right| \le \frac{\gamma}{4}.$$

Hence,

$$\begin{aligned}
\gamma_1 &\ge R_1\left(x_\varepsilon + \frac{\lambda}{T}, y_\varepsilon + \frac{\lambda}{T}\right)\\
&\ge |J_1| - J_2\\
&\ge \frac{\gamma}{2} - \frac{3\varepsilon}{4} - \frac{A_1 + A_2}{T}(3\sqrt{2} + \lambda).
\end{aligned}$$

Thus, as $\varepsilon > 0$ was chosen arbitrarily, it follows that

$$2\gamma_1 + \frac{2(A_1 + A_2)}{T}(3\sqrt{2} + \lambda) \ge \gamma. \tag{A.2}$$

Of course, it may happen that $\gamma = \sup_{u,v\in\mathbb{R}} |R(u,v)| = -\sup_{u,v\in\mathbb{R}} R(u,v)$. But in this case we repeat the steps above by choosing $\theta = -1$. Hence (A.2) holds in both cases.

We now consider the approximation of γ_1. Using the inversion formula, we have that

$$\begin{aligned}
&\int\int \big(F(x-u, y-v) - F(x-u, +\infty)F(+\infty, y-v)\big) q_T(u,v)\, du\, dv\\
&= \frac{1}{(2\pi)^2} \int\int \frac{\varphi(s,t) - \varphi(s,0)\varphi(0,t)}{-st} e^{isx - ity} \widehat{\kappa}\left(\frac{s}{T}\right)\widehat{\kappa}\left(\frac{t}{T}\right) ds\, dt\\
&= \frac{1}{(2\pi)^2} \int\int \frac{\widehat{\varphi}(s,t)}{-st} e^{isx - ity} \widehat{\kappa}\left(\frac{s}{T}\right)\widehat{\kappa}\left(\frac{t}{T}\right) ds\, dt,
\end{aligned}$$

and analogously with respect to the distribution function G. Making the difference between these two representations, we find that

$$\begin{aligned}
&\int\int R(x-u,y-v)q_T(u,v)\,du\,dv \\
&\quad=\int\int\big(F(x-u,+\infty)F(+\infty,y-v) \\
&\qquad-G(x-u,+\infty)G(+\infty,y-v)\big)q_T(u,v)\,du\,dv \\
&\quad=\frac{1}{(2\pi)^2}\int\int\frac{\widehat{\varphi}(s,t)-\widehat{\phi}(s,t)}{-st}e^{isx-ity}\widehat{\kappa}\left(\frac{s}{T}\right)\widehat{\kappa}\left(\frac{t}{T}\right)ds\,dt. \qquad (A.3)
\end{aligned}$$

Finally, remarking that

$$\begin{aligned}
&F(x-u,+\infty)F(+\infty,y-v)-G(x-u,+\infty)G(+\infty,y-v) \\
&\quad\le\big|F(x-y,+\infty)-G(x-u,+\infty)\big|+\big|F(+\infty,y-v)-G(+\infty,y-v)\big|
\end{aligned}$$

and taking into account (A.2) and (A.3), we conclude the proof. □

The following corollary is an immediate consequence of Theorem A.2.

Corollary A.3 *Let X and Y be random variables, denote by $F_{(X,Y)}$, F_X and F_Y their joint and marginal distributions functions, and by $\varphi_{(X,Y)}$, φ_X and φ_Y their joint and marginal characteristic functions. Assume that the marginal densities f_X and f_Y exist and are bounded by M. Then, for every $T>0$,*

$$\begin{aligned}
&\sup_{u,v\in\mathbb{R}}\big|F_{(X,Y)}(u,v)-F_X(u)F_Y(v)\big| \\
&\quad\le\frac{1}{2\pi^2}\int_{[-T,T]^2}\left|\frac{\varphi_{(X,Y)}(s,t)-\varphi_X(s)\varphi_Y(t)}{st}\right|ds\,dt+\frac{45M}{T}.
\end{aligned}$$

Proof A direct application of Theorem A.2, taking into account that the sup terms in the upper bound become

$$\sup_{u\in\mathbb{R}}\big|F_{(X,Y)}(u,+\infty)-F_X(u)\big|=\sup_{v\in\mathbb{R}}\big|F_{(X,Y)}(+\infty,v)-F_Y(v)\big|=0,$$

gives

$$\begin{aligned}
&\sup_{u,v\in\mathbb{R}}\big|F_{(X,Y)}(u,v)-F_X(u)F_Y(v)\big| \\
&\quad\le\frac{1}{2\pi^2}\int_{[-T,T]^2}\left|\frac{\varphi_{(X,Y)}(s,t)-\varphi_X(s)\varphi_Y(t)}{st}\right|ds\,dt+4(3\sqrt{2}+3\sqrt{3})\frac{M}{T}.
\end{aligned}$$

Finally, remark that $4(3\sqrt{2}+4\sqrt{3})\approx 44.6834\le 45$. □

A.2 An Estimate for the Standard Gaussian Distribution

Lemma A.4 *Let X be a standard Gaussian random variable and denote by ϕ its density function. Then, for every $x > 0$,*

$$\mathbf{P}(X > x) \le \frac{1}{x}\phi(x) \quad \textit{and} \quad \mathbf{P}(X > x) \ge \left(x + \frac{1}{x}\right)^{-1}\phi(x).$$

Proof Noticing that $\phi'(x) = -x\phi(x)$ and ϕ is an even function, we have that

$$\begin{aligned}\phi(x) = \phi(-x) &= \int_{-\infty}^{-x} -y\phi(y)\,dy = \int_{x}^{+\infty} y\phi(y)\,dy \\ &\ge x\int_{x}^{+\infty} \phi(y)\,dy = x\mathbf{P}(X > x),\end{aligned}$$

so the first inequality follows. What concerns the second inequality, start by noticing that $(\frac{1}{x}\phi(x))' = -(1 + \frac{1}{x^2})\phi(x)$, so that

$$\begin{aligned}\frac{1}{x}\phi(x) &= \int_{x}^{+\infty}\left(1 + \frac{1}{y^2}\right)\phi(y)\,dy \\ &\le \left(1 + \frac{1}{x^2}\right)\int_{x}^{+\infty} \phi(y)\,dy = \left(1 + \frac{1}{x^2}\right)\mathbf{P}(X > x),\end{aligned}$$

which proves the second inequality. □

A.3 A Maximal Inequality

The following inequality is proved, in a more general form, in Stout [96] (see Theorem 3.7.5 therein).

Theorem A.5 *Let X_n, $n \in \mathbb{N}$, be random variables with finite pth moments. Assume that there exists a constant $K_1 > 0$, independent of n, such that, for every $n \in \mathbb{N}$, $\mathrm{E}|S_n|^p \le K_1 n^{p/2}$. Then, there exits a constant $K_2 > 0$, independent of n, such that, for every $n \in \mathbb{N}$,*

$$\mathrm{E}\Big(\max_{1\le k\le n} |S_k|\Big)^p \le K_2 n^{p/2}.$$

Appendix B
General Results on Large Deviations

We present here, without proofs, an account of general results on large deviation principles. For the details and more results on large deviations, we refer the interested reader to Dembo and Zeitouni [29]. In the sequel, X_n, $n \in \mathbb{N}$, is some sequence of real random variables, and, as usual, $S_n = X_1 + \cdots + X_n$. In order to obtain a general description of rate functions, define, for every $u \in \mathbb{R}$,

$$\Lambda(u) = \lim_{n\to+\infty} \frac{1}{n} \log \mathrm{E}\big(e^{uS_n}\big), \tag{B.1}$$

and recall the Fenchel–Legendre transform of Λ (see Definition 3.15),

$$\Lambda^*(x) = \sup_{u\in\mathbb{R}} \big(ux - \Lambda(u)\big). \tag{B.2}$$

Theorem B.1 (Gärtner–Ellis) *Assume that, for every $x \in \mathbb{R}$, $\Lambda(x)$ defined by* (B.1) *exists. Then, its Fenchel–Legendre transform verifies*:

(a) *for every closed $F \subset \mathbb{R}$,*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n} S_n \in F\right) \le -\inf_{x\in F} \Lambda^*(x);$$

(b) *if Λ is differentiable, then, for every open $G \subset \mathbb{R}$,*

$$\liminf_{n\to+\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n} S_n \in G\right) \ge -\inf_{x\in G} \Lambda^*(x).$$

That is, the Gärtner–Ellis theorem says that it is enough to have the existence of the limit in (B.1), together with the differentiability of Λ, to have a large deviation principle. Moreover, this result also identifies the rate function. Notice that there are no assumptions on the distributions of the random variables nor on their independence. It is thus a very general criterium.

The following characterization of the rate function, expressing some regularity, is useful.

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8,

Theorem B.2 *Let X_n, $n \in \mathbb{N}$, be random variables that satisfy the large deviation principle with rate Λ^*. Then, for any topological basis $\mathcal{A}$ of $\mathbb{R}$ and any $x \in \mathbb{R}$,*

$$\begin{aligned}\Lambda^*(x) &= \sup_{A\in\mathcal{A}:x\in A}\left\{-\liminf_{n\to+\infty}\frac{1}{n}\log\mathbf{P}\left(\frac{1}{n}S_n\in A\right)\right\}\\ &= \sup_{A\in\mathcal{A}:x\in A}\left\{-\limsup_{n\to+\infty}\frac{1}{n}\log\mathbf{P}\left(\frac{1}{n}S_n\in A\right)\right\}.\end{aligned}$$

Definition B.3 A rate function r is called a *good rate function* if, for all $\ell \in \mathbb{R}$, the sets $\{x : r(x) \le \ell\}$ are compact.

The Gärtner–Ellis theorem assumes the differentiability of the Λ before identifying the rate function as its transform. This differentiability is often hard to verify, so an alternative characterization of the rate function is convenient. The notion just introduced helps on this difficulty, as it allows identification of good rate functions as Fenchel–Legendre transforms.

Theorem B.4 *Assume that the large deviation principle is satisfied with a good rate function r. Moreover, assume that, for every $u \in \mathbb{R}$,*

$$\overline{\Lambda}(u) = \limsup_{n\to+\infty}\frac{1}{n}\log \mathrm{E}\big(e^{uS_n}\big) < \infty.$$

Then

(a) *for every $u \in \mathbb{R}$, the limit above exists and satisfies $\overline{\Lambda}(u) = \sup_{t\in\mathbb{R}}(ut - r(t))$;*
(b) *if the rate function r is convex, it is the Fenchel–Legendre transform of $\overline{\Lambda}$.*

Next we quote a result giving sufficient conditions for the large deviation principle to be verified. This was an assumption on the previous theorem, so an alternative condition is convenient. One solution requires some global regularity on the probability distributions of the random variables.

Definition B.5 The probability distributions $\mathbf{P}_n$ on $\mathbb{R}$ are said to be *exponentially tight* if, for each $\varepsilon > 0$, there exists a compact set K such that

$$\lim\sup_{n\to+\infty}\frac{1}{n}\log\mathbf{P}_n\big(K^c\big) < -\varepsilon.$$

This is a strengthened version of the usual tightness condition and plays a role similar to the one of tightness in functional limit theorems.

Theorem B.6 *Assume that $\mathbf{P}_n$, the distributions of $\frac{1}{n}S_n$, are exponentially tight and that, for every continuous and bounded function f, the following limit exists:*

$$\Lambda_f = \lim_{n\to+\infty}\frac{1}{n}\log \mathrm{E}\big(e^{nf(n^{-1}S_n)}\big).$$

Then, the random variables X_n, $n \in \mathbb{N}$, *satisfy the large deviation principle with good rate function* $r(x) = \sup(f(x) - \Lambda_f)$, *where the* sup *is taken over the family of bounded and continuous functions. Moreover, for every such* f,

$$\Lambda_f = \sup_{x \in \mathbb{R}} \bigl(f(x) - r(x)\bigr).$$

The verification of the existence of the limit Λ_f for the given family of functions may be relaxed, requiring such a verification to be done in a suitable smaller family of functions.

Definition B.7 A family $\mathcal{G}$ of continuous real valued functions is *well separated* if

(a) the constant functions are in $\mathcal{G}$;
(b) if $g_1, g_2 \in \mathcal{G}$, then $g(x) = \sup(g_1(x), g_2(x)) \in \mathcal{G}$;
(c) given $x, y \in \mathbb{R}$, $x \neq y$ and $a, b \in \mathbb{R}$, there exists $g \in \mathcal{G}$ such that $g(x) = a$ and $g(y) = b$.

The following result is helpful for checking that the large deviation principle holds.

Theorem B.8 *If, with the notation of the previous theorem, the limit* Λ_g *exists for a well-separated family of functions, then it exists for every continuous and bounded function, that is, the assumption of the previous theorem is satisfied.*

Finally, we refer a convenient well-separated family

$$\mathcal{G} = \Bigl\{ g : \mathbb{R} \longrightarrow \mathbb{R} \text{ continuous, concave, and } \sup_{x \in \mathbb{R}} g(x) < \infty \Bigr\}.$$

Appendix C
Miscellaneous

C.1 Generalized Inverse Functions

Definition C.1 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a monotone function. The *generalized inverse* of f, denoted $f^{\leftarrow}$, is

(a) if f is nondecreasing, $f^{\leftarrow}(u) = \inf\{x : u \leq f(x)\}$;
(b) if f is nonincreasing, $f^{\leftarrow}(u) = \inf\{x : u \geq f(x)\}$.

It is obvious that if f is an invertible function, then $f^{\leftarrow} = f^{-1}$.

C.2 Separation of Variables

It is usual to study centred random variables, and in several estimation problems one is interested in the centred estimator. When considering estimators defined by some quotient, centring by its mean introduces an extra difficulty for characterizing the behaviour of this centring term. It is more convenient to replace this term by the quotient between the means of the numerator and the denominator if we can then separate the corresponding term. This is achieved through the inclusion stated in Theorem C.3 below, proved by Jacob and Niéré [49]. An auxiliary inclusion will be needed in the proof.

Lemma C.2 *Let X and Y be nonnegative random variables, and $\varepsilon \in (0, 1)$. Then*

$$\big\{|X-1| > \varepsilon Y\big\} \subset \left\{|X-1| > \frac{\varepsilon}{2}\right\} \cup \left\{|Y-1| > \frac{\varepsilon}{2}\right\}.$$

Proof Let $\delta > 0$. Then, obviously,

$$\big\{|X-1| > \varepsilon Y\big\} = \big\{|X-1| > \varepsilon Y, |Y-1| < \delta\big\} \cup \big\{|X-1| > \varepsilon Y, |Y-1| \geq \delta\big\}.$$

If $|Y-1| < \delta$, then $Y > 1-\delta$, so that

$$\big\{|X-1| > \varepsilon Y\big\} \subset \big\{|X-1| > \varepsilon(1-\delta)\big\} \cup \big\{|Y-1| \geq \delta\big\}.$$

DOI 10.1007/978-3-642-25532-8,

Choose now δ satisfying $\varepsilon(1-\delta)=\delta$, that is $\delta=\frac{\varepsilon}{\varepsilon+1}>\frac{\varepsilon}{2}$, as $\varepsilon\in(0,1)$. Finally, we have

$$\{|X-Y|>\varepsilon Y\}\subset\left\{|X-1|>\frac{\varepsilon}{2}\right\}\cup\left\{|Y-1|>\frac{\varepsilon}{2}\right\}.$$
□

We now prove the inclusion enabling the separation of variables when dealing with quotients.

Theorem C.3 *Let X and Y be nonnegative integrable random variables. Given $\varepsilon\in(0,\frac{2\mathrm{E}X}{\mathrm{E}Y})$, we have*

$$\left\{\left|\frac{X}{Y}-\frac{\mathrm{E}X}{\mathrm{E}Y}\right|>\varepsilon\right\}\subset\left\{|X-\mathrm{E}X|>\frac{\varepsilon}{4}\mathrm{E}Y\right\}\cup\left\{|Y-\mathrm{E}Y|>\frac{\varepsilon}{4}\frac{(\mathrm{E}Y)^2}{\mathrm{E}X}\right\}. \tag{C.1}$$

Proof Put, for simplicity, $\alpha=\frac{\mathrm{E}Y}{\mathrm{E}X}$. Then, we have

$$\begin{aligned}\left\{\left|\frac{X}{Y}-\frac{\mathrm{E}X}{\mathrm{E}Y}\right|>\varepsilon\right\}&=\left\{\left|\frac{X}{Y}\frac{\mathrm{E}Y}{\mathrm{E}X}-1\right|>\varepsilon\alpha\right\}\\&=\left\{\left|\frac{X}{\mathrm{E}X}-\frac{Y}{\mathrm{E}Y}\right|>\varepsilon\alpha\frac{Y}{\mathrm{E}Y}\right\}\\&\subset\left\{\left|\frac{X}{\mathrm{E}X}-1\right|+\left|\frac{Y}{\mathrm{E}Y}-1\right|>\varepsilon\alpha\frac{Y}{\mathrm{E}Y}\right\}\\&\subset\left\{\left|\frac{X}{\mathrm{E}X}-1\right|>\frac{\varepsilon\alpha}{2}\frac{Y}{\mathrm{E}Y}\right\}\cup\left\{\left|\frac{Y}{\mathrm{E}Y}-1\right|>\frac{\varepsilon\alpha}{2}\frac{Y}{\mathrm{E}Y}\right\}\\&\subset\left\{\left|\frac{X}{\mathrm{E}X}-1\right|>\frac{\varepsilon\alpha}{4}\right\}\cup\left\{\left|\frac{Y}{\mathrm{E}Y}-1\right|>\frac{\varepsilon\alpha}{4}\right\},\end{aligned}$$

taking into account the previous lemma. □

C.3 Integration by Parts

Theorem C.4 *Let $f:\mathbb{R}^2\longrightarrow\mathbb{C}$ be twice continuously differentiable with bounded derivatives. If X and Y are square-integrable random variables, then*

$$\int_{\mathbb{R}^2}f(x,y)(\mathbf{P}_{(X,Y)}-\mathbf{P}_X\otimes\mathbf{P}_Y)(dx\,dy)=\int_{\mathbb{R}^2}\frac{\partial^2 f}{\partial x\,\partial y}(x,y)H(x,y)\,dx\,dy, \tag{C.2}$$

where $H(x,y)=\mathbf{P}(X>x,Y>y)-\mathbf{P}(X>x)\mathbf{P}(Y>y)$.

Proof Notice first that the integral on the right of (C.2) is finite as f has bounded derivatives and the random variables are square integrable. Let $u,v\in\mathbb{R}$. Then, for all $x\geq u$, $y\geq v$,

$$f(x,y)=f(u,y)+f(x,v)-f(u,v)+\int_{(u,x)\times(v,y)}\frac{\partial^2 f}{\partial x\,\partial y}(s,t)\,ds\,dt.$$

Integrating these expressions over $(u,+\infty)\times(v,+\infty)$ with respect to the measure $\mu=\mathbf{P}_{(X,Y)}-\mathbf{P}_X\otimes\mathbf{P}_Y$ and letting $u,v\longrightarrow-\infty$, we need to prove that

$$\lim_{u,v\to-\infty}\int_{(u,+\infty)\times(v,+\infty)} f(u,y)+f(x,v)-f(u,v)\mu(dx\,dy)=0 \tag{C.3}$$

and

$$\begin{aligned}&\int_{\mathbb{R}^2}\frac{\partial^2 f}{\partial x\,\partial y}(x,y)H(x,y)\,dx\,dy\\ &\quad=\lim_{u,v\to-\infty}\int_{(u,+\infty)\times(v,+\infty)}\int_{(u,x)\times(v,y)}\frac{\partial^2 f}{\partial x\,\partial y}(s,t)\,ds\,dt\,\mu(dx\,dy).\end{aligned}\tag{C.4}$$

Separate (C.3) into the sum of three terms in the obvious way. For the first term,

$$\begin{aligned}&\int_{(u,+\infty)\times(v,+\infty)} f(u,y)\mu(dx\,dy)\\ &\quad=\int_{(u,+\infty)\times(v,+\infty)} f(u,y)\mathbf{P}_{(X,Y)}(dx\,dy)\\ &\qquad-\int_{\mathbb{R}\times(v,+\infty)}\mathbf{P}(X>u)f(u,y)\mathbf{P}_{(X,Y)}(dx\,dy)\\ &\quad\le\int_{\mathbb{R}\times(v,+\infty)}\big|f(u,y)\big|\big|\mathbb{I}_{(u,+\infty)}(x)-\mathbf{P}(X>u)\big|\mathbf{P}_{(X,Y)}(dx\,dy).\end{aligned}$$

Taking into account the differentiability of f, up to the multiplication by a constant, this is still bounded above by

$$\begin{aligned}&\int_{\mathbb{R}\times(v,+\infty)}\big(1+u^2+y^2\big)\big|\mathbb{I}_{(u,+\infty)}(x)-\mathbf{P}(X>u)\big|\mathbf{P}_{(X,Y)}(dx\,dy)\\ &\quad\le 2\big(1+u^2+\mathrm{E}Y^2\big)\mathbf{P}(X\le u),\end{aligned}$$

remarking that $\mathbb{I}_{(u,+\infty)}(x)-\mathbf{P}(X>u)=\mathbb{I}_{(-\infty,u]}(x)-\mathbf{P}(X\le u)$. Thus, as Y is square integrable, $\lim_{u\to-\infty}(1+\mathrm{E}Y^2)\mathbf{P}(X\le u)=0$. On the other hand, for $u<0$, one has that $u^2\mathbf{P}(X\le u)\le\mathrm{E}(X^2\mathbb{I}_{(-\infty,u]}(X))\longrightarrow 0$ as $u\longrightarrow-\infty$. The second term in (C.3) is obviously identical to the one just treated. Finally, the third term is, up to the multiplication by a constant, bounded above by

$$\mathbf{P}(X\le u,Y\le y)-\mathbf{P}(X\le u)\mathbf{P}(Y\le v)+2u^2\mathbf{P}(X\le u)+2v^2\mathbf{P}(Y\le v)\longrightarrow 0$$

as $u,v\longrightarrow-\infty$. For (C.4), taking into account the boundedness of the derivatives and the integrability of H (as follows from Theorem 1.4), we may apply the Dominated Convergence Theorem and Fubini's theorem to find

$$\begin{aligned}&\int_{\mathbb{R}^2}\frac{\partial^2 f}{\partial x\,\partial y}(x,y)H(x,y)\,dx\,dy\\ &\quad=\lim_{u,v\to-\infty}\int_{(u,+\infty)\times(v,+\infty)}\frac{\partial^2 f}{\partial x\,\partial y}(x,y)H(x,y)\,dx\,dy\end{aligned}$$

$$
\begin{aligned}
&= \lim_{u,v\to-\infty} \int_{(u,+\infty)\times(v,+\infty)} \frac{\partial^2 f}{\partial x\,\partial y}(x,y) \\
&\quad \times \int_{(u,+\infty)\times(v,+\infty)} \mathbb{I}_{(x,+\infty)\times(y,+\infty)}(s,t)\mu(ds\,dt)\,dx\,dy \\
&= \lim_{u,v\to-\infty} \int_{(u,+\infty)\times(v,+\infty)} \int_{(u,x)\times(v,y)} \frac{\partial^2 f}{\partial x\,\partial y}(s,t)\,ds\,dt\mu(dx\,dy). \qquad \square
\end{aligned}
$$

C.4 Some Asymptotic Results on Real Sequences

In order to prepare the framework for the proof of the large deviation principle, the following result about subadditive sequences is useful.

Lemma C.5 *Let u_n and ε_n, $n \in \mathbb{N}$, be sequences of real numbers such that $u_{n+m} \le u_n + u_m + \varepsilon_{n+m}$, where, for some $\delta > 1$,*

$$
\limsup_{n\to+\infty} \frac{\varepsilon_n}{n} \log^\delta n < \infty.
$$

Then, $\overline{u} = \lim_{n\to+\infty} \frac{u_n}{n}$ exists.

The following two lemmas are standard results on convergence of sequences of real numbers.

Lemma C.6 (Cesàro lemma) *Let b_n, $n \in \mathbb{N}$, be an increasing sequence of positive real numbers such that $b_n \longrightarrow +\infty$, and x_n, $n \in \mathbb{N}$, a sequence of real numbers such that $x_n \longrightarrow x_\infty \in \mathbb{R}$. Then, with $b_0 = 0$,*

$$
\frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) x_k \longrightarrow x_\infty.
$$

Proof Fix $\varepsilon > 0$ and choose n_0 such that $x_n \ge x_\infty - \varepsilon$ for every $n \ge n_0$. Then

$$
\begin{aligned}
&\liminf_{n\to+\infty} \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) x_k \\
&\quad \ge \liminf_{n\to+\infty} \left(\frac{1}{b_n} \sum_{k=1}^{n_0} (b_k - b_{k-1}) x_k + \frac{b_n - b_{n_0}}{b_n} (x_\infty - \varepsilon) \right). \\
&\quad \ge x_\infty - \varepsilon.
\end{aligned}
$$

Thus, as $\varepsilon >$ is chosen arbitrarily, it follows that

$$
\liminf_{n\to+\infty} \frac{1}{b_n} \sum_{k=1}^{n} (b_k - b_{k-1}) x_k \ge x_\infty.
$$

Now repeat the argument with respect to lim sup to conclude that

$$\limsup_{n\to+\infty} \frac{1}{b_n}\sum_{k=1}^{n}(b_k - b_{k-1})x_k \le x_\infty,$$

so the result follows. □

Lemma C.7 (Kronecker lemma) *Let b_n, $n \in \mathbb{N}$, be an increasing sequence of positive real numbers such that $b_n \longrightarrow +\infty$, x_n, $n \in \mathbb{N}$, a sequence of real numbers and, for each $n \in \mathbb{N}$, $s_n = x_1 + \cdots + x_n$. Assume that $\sum_n \frac{x_n}{b_n}$ converges. Then*

$$\frac{s_n}{b_n} \longrightarrow 0.$$

Proof Define, for each $n \in \mathbb{N}$, $u_n = \sum_{k=1}^{n} \frac{x_n}{b_k}$. Thus, u_n is a convergent sequence, and $u_n - u_{n-1} = \frac{x_n}{b_n}$. Hence,

$$s_n = \sum_{k=1}^{n} b_k(u_k - u_{k-1}) = b_n u_n - \sum_{k=1}^{n}(b_k - b_{k-1})u_{k-1},$$

so that, dividing by b_n, we get the result by the Cesàro lemma. □

C.5 A Formula About Multivariate Gaussian Integrals

Lemma C.8 *For all $\lambda \in (0,1)$ and $x \in \mathbb{R}$, let $p(\lambda, x, y)$ be the Gaussian density function with mean λx and covariance matrix $\Sigma = [\sigma_{ij}]_{i,j=1,\dots,n}$. Then*

$$\frac{\partial p}{\partial \alpha} = -\frac{1}{\alpha}\left(\sum_{i,j=1}^{n} \sigma_{ij}\frac{\partial^2 p}{\partial x_i\,\partial x_j} - \sum_{i=1}^{n} x_j \frac{\partial p}{\partial x_i}\right). \tag{C.5}$$

Proof Denote by $\widehat{p}$ the Fourier transform of p with respect to the variable x, that is,

$$\widehat{p}(z) = \int_{\mathbb{R}^n} e^{i\sum_{j=1}^{n} z_j x_j}\, p(\alpha, x, y)\, dx, \quad z \in \mathbb{R}^n$$

(this is not a characteristic function because, as a function of x alone, p is not a density). If we rewrite (C.5) in terms of Fourier transforms, we find

$$\begin{aligned}\frac{\partial \widehat{p}}{\partial \alpha} &= -\frac{1}{\alpha}\left(\sum_{i,j=1}^{n} \sigma_{ij}\widehat{\frac{\partial^2 p}{\partial x_i\,\partial x_j}} - \sum_{i=1}^{n} \widehat{x_j \frac{\partial p}{\partial x_i}}\right)\\ &= \frac{1}{\alpha}\left(\sum_{i,j=1}^{n} \sigma_{ij} x_j x_k \widehat{p} - \sum_{i=1}^{n}\left(\widehat{p} + x_j \frac{\partial \widehat{p}}{\partial x_i}\right)\right). \end{aligned} \tag{C.6}$$

Now, to identify $\widehat{p}$, notice that

$$
\begin{aligned}
&p(\alpha, x, y)\\
&\quad = \frac{1}{(2\pi)^n (1-\alpha^2)^{n/2} |\Sigma|^{1/2}} \exp\left(\frac{1}{2(1-\alpha^2)} (y-\alpha x)^t \Sigma^{-1} (y-\alpha x)\right)\\
&\quad = \frac{1}{\alpha^n} \frac{\alpha^n}{(2\pi)^n (1-\alpha^2)^{n/2} |\Sigma|^{1/2}} \exp\left(\frac{\alpha^2}{2(1-\alpha^2)} \big(x-\alpha^{-1} y\big)^t \Sigma^{-1} \big(x-\alpha^{-1} y\big)\right),
\end{aligned}
$$

thus, as a function of x, $\alpha^n p(\alpha, x, y)$ is a Gaussian density with mean $\alpha^{-1} y$ and covariance matrix $\frac{1-\alpha^2}{\alpha^2} \Sigma$, so

$$
\alpha^n \widehat{p}(z) = \exp\left(i\alpha^{-1} \sum_{j=1}^n z_j y_j - \frac{1-\alpha^2}{\alpha^2} \sum_{j,k=1}^n z_j z_j \sigma_{jk}\right).
$$

The verification of (C.6) is now a simple matter of routine computation. □

References

1. Anderson, T., Darling, D.: Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. **23**, 193–212 (1952)
2. Andrews, D., Pollard, D.: An introduction to functional central limit theorems for dependent stochastic processes. Int. Stat. Rev. **62**, 119–132 (1994)
3. Azevedo, C., Oliveira, P.E.: Kernel-type estimation of bivariate distribution function for associated random variables. In: New Trends in Probability and Statistics, vol. 5. VSP, Utrecht (2000)
4. Bagai, I., Prakasa Rao, B.L.S.: Estimation of the survival function for stationary associated processes. Stat. Probab. Lett. **12**, 385–391 (1991)
5. Barlow, R., Proschan, F.: Statistical Theory of Reliability and Life Testing: Probability Models. Holt, Rinehart & Winston, New York (1975)
6. Bensaïd, N., Fabre, J.P.: Estimation par noyau d'une derivée de Radon–Nikodym sous des conditions de mélange. Can. J. Stat. **28**, 267–282 (1998)
7. Bensaïd, N., Oliveira, P.E.: Histogram estimation of Radon–Nikodym derivatives for strong mixing data. Statistics **35**, 569–592 (2001)
8. Berkes, I.: The functional law of iterated logarithm for dependent random variables. Z. Wahrscheinlichkeitstheor. Verw. Geb. **26**, 246–258 (1973)
9. Berry, A.: The accuracy of the Gaussian approximation to the sum of independent variables. Trans. Am. Math. Soc. **49**, 122–136 (1941)
10. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1968)
11. Billingsley, P.: Probability and Measure. Wiley, New York (1986)
12. Birkel, T.: The invariance principle for associated processes. Stoch. Process. Appl. **27**, 57–71 (1988)
13. Birkel, T.: Moment bounds for associated sequences. Ann. Probab. **16**, 1184–1193 (1988)
14. Birkel, T.: On the convergence rate in the central limit theorem for associated processes. Ann. Probab. **16**, 1685–1698 (1988)
15. Birkel, T.: A note on the strong law of large numbers for positively dependent random variables. Stat. Probab. Lett. **7**, 17–20 (1989)
16. Birkel, T.: A functional central limit theorem for positively dependent random variables. J. Multivar. Anal. **44**, 314–320 (1993)
17. Block, H., Savits, T., Shaked, M.: Some concepts of negative dependence. Ann. Probab. **10**, 765–772 (1982)
18. Blum, J., Kiefer, J., Rosenblatt, M.: Distribution free tests of independence based on the sample distribution. Ann. Math. Stat. **32**, 485–498 (1961)
19. Bochner, S.: Harmonic Analysis and the Theory of Probability. University of California Press, Berkeley (1955)
20. Bryc, W.: On large deviations for uniformly mixing sequences. Stoch. Process. Appl. **41**, 191–202 (1992)

DOI 10.1007/978-3-642-25532-8,

21. Bryc, W., Dembo, A.: Large deviations and strong mixing. Ann. Inst. Henri Poincaré Probab. Stat. **32**, 549–569 (1996)
22. Bulinski, A.: On the convergence rates in the CLT for positively and negatively dependent random fields. In: Ibragimov, I., Zaitov, Y. (eds.) Probability Theory and Mathematical Statistics, pp. 3–14. Gordon & Breach, Amsterdam (1996)
23. Cai, Z., Roussas, G.G.: Smooth estimate of quantiles under association. Stat. Probab. Lett. **36**, 275–287 (1997)
24. Cai, Z., Roussas, G.G.: Efficient estimation of a distribution function under quadrant dependence. Scand. J. Stat. **25**, 211–224 (1998)
25. Cox, J., Grimmett, G.: Central limit theorems for associated random variables and the percolation model. Ann. Probab. **12**, 514–528 (1984)
26. Cremers, H., Kadelka, D.: On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in L_p^E. Stoch. Process. Appl. **21**, 305–317 (1986)
27. Dabrowski, A.: A functional law of iterated logarithm for associated sequences. Stat. Probab. Lett. **3**, 209–212 (1985)
28. Dabrowski, A., Dehling, H.: A Berry–Esséen theorem and a functional law of the iterated logarithm for weakly associated random variables. Stoch. Process. Appl. **30**, 277–289 (1988)
29. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications. Springer, New York (1998)
30. Dewan, I., Prakasa Rao, B.L.S.: A general method of density estimation for associated random variables. J. Nonparametr. Stat. **10**, 405–420 (1999)
31. Donsker, M.: An invariance principle for certain probability limit theorems. Mem. Amer. Math. Soc. **6** (1951)
32. Donsker, M.: Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. Ann. Math. Stat. **23**, 277–281 (1952)
33. Doob, J.: Heuristic approach to the Kolmogorov–Smirnov theorems. Ann. Math. Stat. **20**, 393–403 (1949)
34. Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stoch. Process. Appl. **84**, 313–342 (1999)
35. Erdos, P., Käc, M.: On central limit theorems in the theory of probability. Bull. Am. Math. Soc. **52**, 292–302 (1946)
36. Esary, J., Proschan, F., Walkup, D.: Association of random variables with applications. Ann. Math. Stat. **38**, 1466–1474 (1967)
37. Esséen, C.: Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. Acta Math. **77**, 1–125 (1945)
38. Ferrieux, D.: Estimation de densités de mesures moyennes de processus ponctuels associées. Ph.D. thesis, Université de Montpellier, France (1996) (in French)
39. Ferrieux, D.: Estimation à noyau de densités moyennes de mesures aléatoires associées. C. R. Acad. Sci., Sér. 1 Math. **326**, 1131–1134 (1998) (in French)
40. Fortuin, C., Kasteleyn, P., Ginibre, J.: Correlation inequalities on some partially ordered sets. Commun. Math. Phys. **22**, 89–103 (1971)
41. Henriques, C., Oliveira, P.E.: Estimation of a two-dimensional distribution function under association. J. Stat. Plan. Inference **113**, 137–150 (2003)
42. Henriques, C., Oliveira, P.E.: Exponential rates for kernel density estimation under association. Stat. Neerl. **59**, 448–466 (2005)
43. Henriques, C., Oliveira, P.E.: Convergence rates for the strong law of large numbers under association. Technical report, Pré-Publicações do Departamento de Matemática da Universidade de Coimbra 08-14 (2008)
44. Henriques, C., Oliveira, P.E.: Large deviations for the empirical mean of associated random variables. Stat. Probab. Lett. **78**, 594–598 (2008)
45. Herrndorf, N.: The invariance principle for φ-mixing sequences. Z. Wahrscheinlichkeitstheor. Verw. Geb. **63**, 97–108 (1983)
46. Hoeffding, W.: A nonparametric test of independence. Ann. Math. Stat. **19**, 546–557 (1948)

47. Ibragimov, I.: Some limit theorems for stationary sequences. Theory Probab. Appl. **7**, 349–382 (1962)
48. Ioannides, D., Roussas, G.G.: Exponential inequality for associated random variables. Stat. Probab. Lett. **42**, 423–431 (1998)
49. Jacob, P., Niéré, L.: Contribution a l'estimation des lois de Palm d'une mesure aléatoire. Publ. Inst. Stat. Univ. Paris **2**, 39–49 (1990)
50. Jacob, P., Oliveira, P.E.: A general approach to non-parametric histogram estimation. Statistics **27**, 73–92 (1995)
51. Jacob, P., Oliveira, P.E.: Kernel estimators of general Radon–Nikodym derivatives. Statistics **30**, 25–46 (1997)
52. Jacob, P., Oliveira, P.E.: Histograms and associated point processes. Stat. Inference Stoch. Process. **2**, 227–251 (1999)
53. Joag-Dev, K.: Independence via uncorrelatedness under certain dependence structures. Ann. Probab. **11**, 1037–1041 (1983)
54. Joag-Dev, K., Proschan, F.: Negative association of random variables with applications. Ann. Stat. **11**, 286–295 (1983)
55. Lebowitz, J.: Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. Commun. Math. Phys. **28**, 313–321 (1972)
56. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes. Results in Mathematics and Related Areas (3), vol. 23. Springer-Verlag, Berlin (1991)
57. Lehmann, E.: Consistency and unbiasedness of certain nonparametric tests. Ann. Math. Stat. **22**, 165–179 (1951)
58. Lehmann, E.: Some concepts of dependence. Ann. Math. Stat. **37**, 1137–1153 (1966)
59. Li, Y.-X., Wang, J.-F.: The law of iterated logarithm for positively dependent random variables. J. Math. Anal. Appl. **339**, 259–265 (2008)
60. Lindqvist, H.: Association of probability measures on partially ordered spaces. J. Multivar. Anal. **26**, 111–132 (1988)
61. Louhichi, S.: Rosenthal's inequality for LPQD sequences. Stat. Probab. Lett. **42**, 139–144 (1999)
62. Louhichi, S.: Convergence rates in the strong law for associated random variables. Probab. Math. Stat. **20**, 203–214 (2000)
63. Louhichi, S.: Weak convergence for empirical processes of associated sequences. Ann. Inst. Henri Poincaré Probab. Stat. **36**, 547–567 (2000)
64. Louhichi, S., Soulier, P.: Marcinkiewicz–Zygmund strong laws for infinite variance time series. Stat. Inference Stoch. Process. **3**, 31–40 (2000). 19th Rencontres Franco-Belges de Statisticiens (Marseille, 1998)
65. Masry, E.: Multivariate probability density estimation for associated processes: strong consistency and rates. Stat. Probab. Lett. **58**, 205–219 (2002)
66. Morel, B., Suquet, C.: Hilbertian invariance principles for the empirical process under association. Math. Methods Stat. **11**, 203–220 (2002)
67. Nachbin, L.: Topology and Order. Van Nostrand, New York (1965)
68. Newman, C.: Normal fluctuations and the FKG inequalities. Commun. Math. Phys. **74**, 119–128 (1980)
69. Newman, C.: A general central limit theorem for FKG systems. Commun. Math. Phys. **91**, 75–80 (1983)
70. Newman, C.: Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: Tong, Y. (ed.) Inequalities in Statistics and Probability, vol. 5, pp. 127–140. Inst. Math. Stat., Hayward (1984)
71. Newman, C., Wright, A.: An invariance principle for certain dependent sequences. Ann. Probab. **9**, 671–675 (1981)
72. Newman, C., Wright, A.: Associated random variables and martingale inequalities. Z. Wahrscheinlichkeitstheor. Verw. Geb. **59**, 361–371 (1982)
73. Oliveira, P.E.: Invariance principles in $L^2[0,1]$. Comment. Math. Univ. Carol. **31**, 357–366 (1990)

74. Oliveira, P.E.: Density estimation for associated sampling: a point process influenced approach. J. Nonparametr. Stat. **14**, 495–509 (2002)
75. Oliveira, P.E.: An exponential inequality for associated variables. Stat. Probab. Lett. **73**, 189–197 (2005)
76. Oliveira, P.E., Suquet, C.: An invariance principle in $L^2[0, 1]$ for non stationary φ-mixing sequences. Comment. Math. Univ. Carol. **36**, 293–302 (1995)
77. Oliveira, P.E., Suquet, C.: $L^2[0, 1]$ weak convergence of the empirical process for dependent variables. In: Antoniadis, A., Oppenheim, G. (eds.) Actes des XVèmes Rencontres Franco-Belges de Statisticiens (Ondelettes et Statistique). Lecture Notes in Statistics, vol. 103, pp. 331–344. Springer, New York (1995)
78. Oliveira, P.E., Suquet, C.: An $L^2[0, 1]$ invariance principle for LPQD random variables. Port. Math. **53**, 367–379 (1996)
79. Oliveira, P.E., Suquet, C.: Weak convergence in $L^p(0, 1)$ of the uniform empirical process under dependence. Stat. Probab. Lett. **39**, 363–370 (1998)
80. Parthasarathy, K.: Probability Measures on Metric Spaces. Academic Press, San Diego (1967)
81. Pisier, G.: Some applications of the metric entropy condition to harmonic analysis. In: Banach Spaces, Harmonic Analysis and Probability Theory. Lecture Notes in Mathematics, vol. 995, pp. 123–154. Springer, New York (1983)
82. Pitt, L.: Positively correlated normal variables are associated. Ann. Probab. **10**, 496–499 (1982)
83. Prakasa Rao, B.L.S.: Bernstein-type inequality for associated sequences. In: Ghosh, J., Mitra, S., Parthasarathy, K., Prakasa Rao, B.L.S. (eds.) Statistics and Probability: A Raghu Raj Bahadur Festschrift, pp. 499–509. Wiley Eastern, New Delhi (1993)
84. Prokhorov, Y.V.: Convergence of random processes and limit theorems in probability theory. Theory Probab. Appl. **1**, 157–214 (1956)
85. Roussas, G.G.: Kernel estimates under association: strong uniform consistency. Stat. Probab. Lett. **12**, 215–224 (1991)
86. Roussas, G.G.: Curve estimation in random fields of associated processes. J. Nonparametr. Stat. **3**, 215–224 (1993)
87. Roussas, G.G.: Asymptotic normality of a smooth estimate of a random field distribution function under association. Stat. Probab. Lett. **24**, 77–90 (1995)
88. Roussas, G.G.: Asymptotic normality of the kernel estimate of a probability density function under association. Stat. Probab. Lett. **50**, 1–12 (2000)
89. Roussas, G.G.: An Esséen-type inequality for probability density functions, with an application. Stat. Probab. Lett. **51**, 397–408 (2001)
90. Roussas, G.G.: Erratum to "An Esséen-type inequality for probability density functions, with an application". Stat. Probab. Lett. **54**, 449 (2001)
91. Sadikova, S.: Two dimensional analogues of an inequality of Esséen with applications to the central limit theorem. Theory Probab. Appl. **11**, 325–335 (1966)
92. Shaked, M.: A family of concepts of positive dependence for bivariate distributions. J. Am. Stat. Assoc. **72**, 642–650 (1977)
93. Shaked, M.: A general theory of some positive dependence notions. J. Multivar. Anal. **12**, 199–218 (1982)
94. Shao, Q.M., Yu, H.: Weak convergence for weighted empirical processes of dependent sequences. Ann. Probab. **24**, 2098–2127 (1996)
95. Shepp, L.: On the integral of the absolute value of the pinned Wiener process. Ann. Probab. **10**, 234–239 (1982)
96. Stout, W.: Almost Sure Convergence. Academic Press, San Diego (1974)
97. Stroock, D.: A Concise Introduction to the Theory of Integration. Birkhäuser, Basel (1994)
98. Sung, S.: A note on the exponential inequality for associated random variables. Stat. Probab. Lett. **77**, 1730–1736 (2007)
99. Suquet, C.: Relecture des critères de realtive compacité d'une famille de probabilités sur un espace de Hilbert. Technical report, Publ. IRMA, Lille 28-III (1992)

100. Suquet, C.: Tightness in Schauder decomposable Banach spaces. In: Proc. St. Petersburg Mathematical Soc. American Mathematical Society Translations, Series 2, vol. 193 (1999)
101. Utev, S.A.: On the central limit theorem for φ-mixing arrays of random variables. Theory Probab. Appl. **35**, 131–139 (1990)
102. Vakhania, V.V., Tarieladze, V.I., Chobanyan, S.A.: Probability Distributions on Banach Spaces. Reidel, Dordrecht (1987)
103. Watson, G.: Goodness-of-fit tests on a circle. Biometrika **48**, 109–114 (1981)
104. Wood, T.: A Berry–Esséen theorem for associated random variables. Ann. Probab. **11**, 1042–1047 (1983)
105. Xing, G., Yang, S.: Notes on the rate of strong convergence for associated random variables. Commun. Stat., Theory Methods **38**, 1–4 (2009)
106. Xing, G., Yang, S., Liu, A.: Exponential inequalities for positively associated random variables and applications. J. Inequal. Appl. **2008**, 385362 (2008)
107. Yang, S.: Moment inequalities for partial sums of random variables. Sci. China Ser. A **44**, 1–6 (2001)
108. Yang, S., Su, C., Yu, K.: A general method to the strong law of large numbers and its applications. Stat. Probab. Lett. **78**, 794–803 (2008)
109. Yu, H.: A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences. Probab. Theory Relat. Fields **95**, 357–370 (1993)
110. Yu, H.: The law of the iterated logarithm for associated random variables. Acta Math. Sin. **29**, 507–511 (1996)

Index

P.E. Oliveira, *Asymptotics for Associated Random Variables*,
DOI 10.1007/978-3-642-25532-8,