Springer Series in Operations Research
and Financial Engineering

Gennady Samorodnitsky

Stochastic Processes and Long Range Dependence

Springer

Springer Series in Operations Research and Financial Engineering

Series Editors

Thomas V. Mikosch
Sidney I. Resnick
Stephen M. Robinson

More information about this series at http://www.springer.com/series/3182

Gennady Samorodnitsky

Stochastic Processes and Long Range Dependence

Springer

Gennady Samorodnitsky
School of Operations Research
and Information Engineering
Cornell University
Ithaca, NY, USA

ISSN 1431-8598 ISSN 2197-1773 (electronic)
Springer Series in Operations Research and Financial Engineering
ISBN 978-3-319-45574-7 ISBN 978-3-319-45575-4 (eBook)
DOI 10.1007/978-3-319-45575-4

Library of Congress Control Number: 2016951256

Mathematics Subject Classification (2010): 60G10, 60G22, 60G18, 60G52, 60F17, 60E07

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

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

To the only loves of my life: to Julia, Eric, Danny, and Sarah

Preface

I first heard about long-range dependence while working on a book on stable processes with Murad Taqqu. Initially, the notion did not seem to stand out among other notions I was familiar with at the time. It seemed to describe simply situations in which covariance functions (or related functions) decayed at a slow rate. Why were so many other people excited about long-range dependence? At best, it seemed to require us to prove some more theorems. With time, I came to understand that I was wrong, and the people who got excited about long-range dependence were right. The content of this phenomenon is truly special, even if somewhat difficult to define precisely. This book is a product of many years of thinking about long memory (this term is synonymous with long-range dependence). It is my hope that it will serve as a useful complement to the existing books on long-range dependence such as Palma (2007), Giraitis et al. (2012), and Beran et al. (2013), and numerous surveys and collections.

I firmly believe that the main importance of the notion of long-range dependence is in statistical applications. However, I think of long-range dependence as a property of stationary stochastic processes, and this book is, accordingly, organized around probabilistic properties of stationary processes that are important for the presence or absence of long memory. The first four chapters of this book are therefore not really about long-range dependence, but deal with several topics in the general theory of stochastic processes. These chapters provide background, language, and models for the subsequent discussion of long memory. The subsequent five chapters deal with long-range dependence proper. This explains the title of the book: *Stochastic Processes and Long-Range Dependence*.

The four general chapters begin with a chapter on stationarity and invariance. The property of long-range dependence is by definition a property of stationary processes, so including such a chapter is necessary. Information on stationary processes is available from many sources, but some of the material in this chapter is less standard. The second chapter presents elements of ergodic theory of stationary processes. Ergodic theory intersects our journey through long-range dependence multiple times, so this chapter is also necessary. There are plenty of books on ergodic theory, but this literature is largely disjoint from books on stochastic

processes. Chapter 3 is a crash course on infinitely divisible processes. These processes provide a crucial source of examples on which to study the presence or absence of long memory. Much of the material in this chapter is not easily available from a single alternative source. Chapter 4 presents basic information on heavy tailed models. There is significant difference in the way long-range dependence expresses itself in stationary processes with light tails and those with heavy tails, particularly processes with infinite second moment. Therefore, including this chapter seems useful.

Chapter 5 is the first chapter specifically on long-range dependence. It is of an introductory and historical character. The best-known approach to long-range dependence, applicable to stationary processes with a finite second moment, is presented in Chapter 6. The vast majority of the literature on long-memory processes falls within this second-order approach. The chapter we include contains results not easily available elsewhere. Long-range dependence is sometimes associated with fractional integration, and Chapter 7 discusses this connection in some detail. Long-range dependence is also frequently associated with self-similarity. The connection is deep, and much of its power is due to the Lamperti theorem, which guarantees self-similarity of the limit in certain functional limit theorems. Chapter 8 presents the theory of self-similar processes, particularly self-similar processes with stationary increments. Finally, Chapter 9 introduces a less-standard point of view on long memory. It is the point of view that I have come to adopt over the years. It views the phenomenon of long-range dependence as a phase transition. In this chapter, we illustrate the phenomenon in a number of situations. Some of the results in this chapter have not appeared before.

The book concludes with an appendix. I have chosen to include it for convenience of the reader. It describes a number of notions and results belonging to the topics used frequently throughout this book.

The book can be used for a one-semester graduate topics course, even though the amount of material it contains is probably enough for a semester and a half, so the instructor has to be selective. There are exercises at the end of each chapter.

Writing this book took me a long time. I started working on it during my sabbatical in the Department of Mathematics of the University of Copenhagen and finished it during my following sabbatical (!) in the Department of Statistics of Columbia University. Most of it was, of course, written between those two visits, in my home department, School of Operations Research and Information Engineering of Cornell University. I am grateful to all these institutions for providing me with wonderful facilities and colleagues that greatly facilitated writing this book.

A number of people have read through portions of the manuscript and contributed useful comments and corrections. My particular thanks go to Richard Davis, Emily Fisher, Eugene Seneta, Julian Sun, and Phyllis Wan.

Ithaca, NY, USA Gennady Samorodnitsky

Contents

Chapter 1
Stationary Processes

1.1 Stationarity and Invariance

The stationarity of a stochastic process means the invariance of its finite-dimensional distributions under certain transformations of its parameter space. The classical definitions apply to the situations in which the parameter is one-dimensional, and has the interpretation of time.

Definition 1.1.1. A discrete-time stochastic process $\big(X_n,\, n \in \mathbb{Z}\big)$ is stationary if $\big(X_{n+k},\, n \in \mathbb{Z}\big) \stackrel{d}{=} \big(X_n,\, n \in \mathbb{Z}\big)$ for all $k \in \mathbb{Z}$.

A continuous-time stochastic process $\big(X(t),\, -\infty < t < \infty\big)$ is stationary if $\big(X(t+s),\, -\infty < t < \infty\big) \stackrel{d}{=} \big(X(t),\, -\infty < t < \infty\big)$ for all $s \in \mathbb{R}$.

In this case, the transformations of the (one-dimensional) parameter space form the group of shifts $g_s : T \to T$, $s \in T$, defined by $g_s t = t + s$ for $t \in T = \mathbb{Z}$ or $T = \mathbb{R}$.

Sometimes, the stationarity of a stochastic process with one-dimensional time is defined "halfway," so to speak: the process is defined only on the positive half-line, and shifts by only a positive amount are allowed. The following proposition shows that the two notions are equivalent.

Proposition 1.1.2. (i) *A discrete-time stochastic process* $\big(X_n,\, n = 0, 1, 2, \ldots\big)$ *has the property that* $\big(X_{n+k},\, n = 0, 1, 2, \ldots\big) \stackrel{d}{=} \big(X_n,\, n = 0, 1, 2, \ldots\big)$ *for all* $k = 0, 1, 2, \ldots$ *if and only if there exists a stationary stochastic process* $\big(Y_n,\, n \in \mathbb{Z}\big)$ *such that* $\big(X_n,\, n = 0, 1, 2, \ldots\big) \stackrel{d}{=} \big(Y_n,\, n = 0, 1, 2, \ldots\big)$.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_1

(ii) *A continuous-time stochastic process* $\big(X(t),\, t \geq 0\big)$ *has the property that* $\big(X(t+s),\, t \geq 0\big) \stackrel{d}{=} \big(X(t),\, t \geq 0\big)$ *for all* $s \geq 0$ *if and only if there exists a stationary stochastic process* $\big(Y(t),\, -\infty < t < \infty\big)$ *such that* $\big(X(t),\, t \geq 0\big) \stackrel{d}{=} \big(Y(t),\, t \geq 0\big)$.

Proof. We will prove the second part of the proposition. The proof in the discrete-time case is the same. Clearly, the existence of a stationary process $\big(Y(t),\, -\infty < t < \infty\big)$ as in the proposition guarantees that the process $\big(X(t),\, t \geq 0\big)$ has the required shift-invariance of its finite-dimensional distributions. Conversely, suppose that the finite-dimensional distributions of $\big(X(t),\, t \geq 0\big)$ are invariant under positive shifts. Define a family of finite-dimensional distributions on $\mathbb{R}$ by

$$F_{t_1,\ldots,t_k}(A) = P\Big(\big(X(0), X(t_2 - t_1), \ldots, X(t_k - t_1)\big) \in A\Big) \tag{1.1}$$

for $k \geq 1$, $t_1 < t_2 < \ldots < t_k$, and A a k-dimensional Borel set. This family is clearly consistent and invariant under shifting all the time points by any real number. By the Kolmogorov existence theorem (see, e.g., Theorem 6.16 in Kallenberg (2002)), there exists a stochastic process $\big(Y(t),\, -\infty < t < \infty\big)$ whose finite-dimensional distributions are given by (1.1). The shift-invariance of the family (1.1) means that this stochastic process is stationary, and by construction, its restriction to nonnegative times has the same finite-dimensional distributions as the process $\big(X(t),\, t \geq 0\big)$. □

Remark 1.1.3. Sometimes, the distributional invariance under shifts of Definition 1.1.1 is referred to as *strict stationarity*, to distinguish it (for stochastic processes with a finite second moment) from the invariance of the mean of the process and its covariance function when the time of the process is shifted. This weaker invariance property is then called "stationarity." In this book, stationarity means exclusively the distributional invariance of Definition 1.1.1, and we will refer to stochastic processes possessing the weaker invariance property as "weakly stationary" or "second-order stationary."

For stochastic processes $\big(X(t),\, t \in T\big)$ whose parameter space is not necessarily one-dimensional, the notion of stationarity is typically connected to a group of transformations of T. Let G be a group of transformations $g : T \to T$ (the transformations are then automatically one-to-one and onto).

Definition 1.1.4. A stochastic process $\big(X(t),\, t \in T\big)$ is called G-stationary (or stationary with respect to the group G of transformations of T) if $\big(X(g(t)),\, t \in T\big) \stackrel{d}{=} \big(X(t),\, t \in T\big)$ for all $g \in G$.

The most common examples are those of stochastic processes indexed by a finite-dimensional Euclidian space.

Example 1.1.5. Let $T = \mathbb{R}^d$ for $d = 1, 2, \ldots$. One says that a stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is stationary if $\big(X(\mathbf{t}+\mathbf{s}),\, \mathbf{t} \in \mathbb{R}^d\big) \stackrel{d}{=} \big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ for all $\mathbf{s} \in \mathbb{R}^d$.

Here the group G of transformations in Definition 1.1.4 is the group of shifts $g_{\mathbf{s}}(\cdot) = \cdot + \mathbf{s}$ for $\mathbf{s} \in \mathbb{R}^d$. Stationarity with respect to the group of shifts will be our default notion of stationarity of stochastic processes indexed by $\mathbb{R}^d$, unless a different group of transformations of $\mathbb{R}^d$ is specified, as in the next example.

Example 1.1.6. Let, once again, $T = \mathbb{R}^d$ for $d = 1, 2, \ldots$. A stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is said to be *isotropic* if $\big(X(U(\mathbf{t})),\, \mathbf{t} \in \mathbb{R}^d\big) \stackrel{d}{=} \big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ for each $d \times d$ orthogonal matrix U. According to Definition 1.1.4, an isotropic stochastic process is stationary with respect to the group $G = \mathbf{SO}(d)$ of rotations of $\mathbb{R}^d$.

A certain amount of ambiguity is connected to the notion of isotropy, because one sometimes combines it with the stationarity under shifts of Example 1.1.5 and reserves the adjective "isotropic" for stochastic processes $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ that, according to Definition 1.1.4, are stationary with respect to the group G of rigid motions of $\mathbb{R}^d$, consisting of transformations $g_{U,\mathbf{s}} : \mathbb{R}^d \to \mathbb{R}^d$, $U \in \mathbf{SO}(d)$, $\mathbf{s} \in \mathbb{R}^d$, defined by $g_{U,\mathbf{s}}\mathbf{t} = U(\mathbf{t}) + \mathbf{s}$, $t \in \mathbb{R}^d$.

It is easy to understand what the various notions of stationarity and invariance mean in the case of Gaussian processes.

Example 1.1.7. Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ be a Gaussian stochastic process (see Example 3.10). The finite-dimensional distributions of such a process are determined by the mean function $m(\mathbf{t}) = EX(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^d$ and the covariance function $R(\mathbf{s}, \mathbf{t}) = \mathrm{Cov}\big(X(\mathbf{s}), X(\mathbf{t})\big)$, $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$. Therefore, a Gaussian process is stationary if and only if the mean function $m(\mathbf{t}) \equiv m \in \mathbb{R}$ is constant on $\mathbb{R}^d$ and the covariance function $R(\mathbf{s}, \mathbf{t}) = R(\mathbf{t} - \mathbf{s})$ depends only on the difference between its arguments (we are committing here, and will continue committing in the sequel, the usual sin of using the same name for two slightly different functions).

A Gaussian process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is isotropic if and only if its mean function $m(\mathbf{t}) = m(\|\mathbf{t}\|)$, $\mathbf{t} \in \mathbb{R}^d$, depends only on the length of the parameter t, and the covariance function $R(\mathbf{s}, \mathbf{t}) = R\big(U(\mathbf{t}), U(\mathbf{s})\big)$ remains unchanged if both of its arguments undergo the same rotation. In one dimension, this all means only that the mean function is even.

Finally, a Gaussian process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is stationary with respect to the group of rigid motions of $\mathbb{R}^d$ if and only if its mean function is constant and its covariance function $R(\mathbf{s}, \mathbf{t}) = R(\|\mathbf{t} - \mathbf{s}\|)$ depends only on the length of the difference between its arguments.

Two major classes of stationary stochastic processes are the linear processes of Section 1.4 and the stationary infinitely divisible processes of Section 3.1, of which the Gaussian processes of Example 1.1.7 form a special case.

Definition 1.1.8. A stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ has stationary increments if the finite-dimensional distributions of $\big(X(\mathbf{t} + \mathbf{s}) - X(\mathbf{s}),\, \mathbf{t} \in \mathbb{R}^d\big)$ do not depend on $\mathbf{s} \in \mathbb{R}^d$.

A similar definition applies to stochastic processes indexed by $\mathbf{t} \in \mathbb{Z}^d$, in which case the finite-dimensional distributions of the increment process should not depend on the initial time point $\mathbf{s} \in \mathbb{Z}^d$.

Clearly, every stationary process has stationary increments as well, but the converse is not true: there are nonstationary stochastic processes with stationary increments.

Example 1.1.9. Let $d = 1$. A Brownian motion has stationary increments. More generally, every Lévy process $\big(X(t),\ -\infty < t < \infty\big)$ of Example 3.1.2 below has stationary increments. Such a process is obviously nonstationary (unless it degenerates to the zero process).

Example 1.1.10. Let $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ be a Gaussian process. It is clear that if its mean function is linear, $m(\mathbf{t}) = (\mathbf{c}, \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^d$, for some $\mathbf{c} \in \mathbb{R}^d$, and the incremental variance depends only on the difference between the two points, $\mathrm{Var}\big(X(\mathbf{t}) - X(\mathbf{s})\big) = H(\mathbf{t} - \mathbf{s})$, $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, for some function H, then the process has stationary increments. The latter condition is also necessary for stationarity of the increments. The former condition is necessary as well if, for example, the mean function is continuous (which will be the case if the process is measurable; see below), but in general, there are "pathological" nonlinear mean functions consistent with stationarity of the increments, given as solutions of the Cauchy functional equation. See Bingham et al. (1987). An example of a Gaussian process with stationary increments is the fractional Brownian motion of Example 3.5.1, including the usual Brownian motion as a special case.

The following, somewhat unexpected, result shows that stochastic processes with stationary increments become stationary if one "randomizes" them additionally by a shift chosen according to the Lebesgue measure. We will always denote the Lebesgue measure in any dimension by λ.

Proposition 1.1.11. *A stochastic process $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ on a probability space $\big(\Omega, \mathcal{F}, P\big)$ has stationary increments if and only if the image on $\mathbb{R}^{\mathbb{R}^d}$ of the σ-finite measure $\lambda \times P$ on $\mathbb{R} \times \Omega$ by the map $\big(u + X(\mathbf{t} + \mathbf{s}),\ \mathbf{t} \in \mathbb{R}^d\big) : \mathbb{R} \times \Omega \to \mathbb{R}^{\mathbb{R}^d}$ is independent of $\mathbf{s} \in \mathbb{R}^d$.*

Proof. Suppose first that $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ has stationary increments. Let $\mathbf{t}_1, \dots, \mathbf{t}_k, \mathbf{s}$ be arbitrary points in $\mathbb{R}^d$, and let A be a $(k+1)$-dimensional Borel set. Let F be the $(k+1)$-dimensional law of the random vector $\big(X(\mathbf{s}), X(\mathbf{t}_1 + \mathbf{s}) - X(\mathbf{s}), X(\mathbf{t}_2 + \mathbf{s}) - X(\mathbf{t}_1 + \mathbf{s}), \dots, X(\mathbf{t}_k + \mathbf{s}) - X(\mathbf{t}_{k-1} + \mathbf{s})\big)$. Notice that

$$\lambda \times P\Big((u, \omega) \in \mathbb{R} \times \Omega : \tag{1.2}$$

$$\big(u + X(\mathbf{s}), X(\mathbf{t}_1 + \mathbf{s}) - X(\mathbf{s}), X(\mathbf{t}_2 + \mathbf{s}) - X(\mathbf{t}_1 + \mathbf{s}), \dots, X(\mathbf{t}_k + \mathbf{s}) - X(\mathbf{t}_{k-1} + \mathbf{s})\big) \in A\Big)$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \int_{\mathbb{R}^k} \mathbf{1}\Big((u + x, y_1, \dots, y_k) \in A\Big)\, F(dx, dy_1, \dots, dy_k) \right) du$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \int_{\mathbb{R}^k} \mathbf{1}\Big((u, y_1, \dots, y_k) \in A \Big) F(dx, dy_1, \dots, dy_k) \right) du$$

$$= \int_{\mathbb{R}} P\Big(\big(X(\mathbf{t}_1 + \mathbf{s}) - X(\mathbf{s}), X(\mathbf{t}_2 + \mathbf{s}) - X(\mathbf{t}_1 + \mathbf{s}), \dots, X(\mathbf{t}_k + \mathbf{s}) - X(\mathbf{t}_{k-1} + \mathbf{s}) \big) \in A_u \Big) \, du \,,$$

where $A_u = \big\{ (y_1, \dots, y_k) \in \mathbb{R}^k : (u, y_1, \dots, y_k) \in A \big\}$ is the u-section of A. By the stationarity of the increments, this last expression is independent of $\mathbf{s} \in \mathbb{R}^d$. Since for every $(k+1)$-dimensional Borel set B, there is a $(k+1)$-dimensional Borel set A such that

$$\big\{ (x_0, x_1 \dots, x_k) : (x_0, x_1 \dots, x_k) \in B \big\}$$
$$= \big\{ (x_0, x_1 \dots, x_k) : (x_0, x_1 - x_0, \dots, x_k - x_{k-1}) \in A \big\} \,,$$

this proves the required invariance of the infinite "law" of $\big(u + X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d \big)$ under shifts.

In the opposite direction, given the invariance of the above infinite "law," the very first expression in (1.2) is independent of $\mathbf{s} \in \mathbb{R}^d$ for all $\mathbf{t}_1, \dots, \mathbf{t}_k$ in $\mathbb{R}^d$ and $(k+1)$-dimensional Borel sets A. Choosing $A = [0, 1] \times C$, where C is a k-dimensional Borel set, it follows from (1.2) that

$$P\Big(\big(X(\mathbf{t}_1 + \mathbf{s}) - X(\mathbf{s}), X(\mathbf{t}_2 + \mathbf{s}) - X(\mathbf{t}_1 + \mathbf{s}), \dots, X(\mathbf{t}_k + \mathbf{s}) - X(\mathbf{t}_{k-1} + \mathbf{s}) \big) \in C \Big)$$

does not depend on $\mathbf{s} \in \mathbb{R}^d$ either, which means that the process $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d \big)$ has stationary increments. □

1.2 Stationary Processes with a Finite Variance

For a stochastic process $\big(X(t),\ t \in T \big)$ such that $EX(t)^2 < \infty$ for all $t \in T$, we denote by

$$R_X(s, t) = \mathrm{Cov}(X(s), X(t)),\ s, t \in T,$$

its covariance function. If T is a topological space and the process $\big(X(t),\ t \in T \big)$ is continuous in L^2, then the bound

$$\Big| R_X(s, t) - R_X(s_1, t_1) \Big|$$
$$\le \big(\mathrm{Var}(X(s)) \big)^{1/2} \big(\mathrm{Var}(X(t) - X(t_1)) \big)^{1/2} + \big(\mathrm{Var}(X(t_1)) \big)^{1/2} \big(\mathrm{Var}(X(s) - X(s_1)) \big)^{1/2}$$

(it is an easy consequence of the Cauchy–Schwarz inequality) implies that the covariance function is continuous in both of its arguments.

If $\big(X(t),\ t \in T\big)$ is stationary with respect to a group G of transformations of T, then the covariance function automatically satisfies the relation

$$R_X(s,t) = \mathrm{Cov}(X(s),X(t)) = \mathrm{Cov}\big(X(g(s)),X(g(t))\big) = R_X\big(g(s),g(t)\big) \tag{1.3}$$

for all $s,t \in T$, $g \in G$.

In the particular case that $G = T$ is an abelian group (e.g., $G = T = \mathbb{R}^d$, or $G = T = \mathbb{Z}^d$), we will denote the identity element by 0 and use additive notation. Selecting $g = -s$ in (1.3) then gives us

$$R_X(s,t) = R_X(0,t-s),\quad s,t \in T\,.$$

The covariance function of a G-stationary process then becomes a function of one variable. It is very common simply to drop the unnecessary variable in the covariance function and write $R_X(t)$ when the meaning is $R_X(s,s+t)$ for some (equivalently, any) $s \in T$. We will adopt this reasonably innocent abuse of notation in this book. If $G = T$ is a topological abelian group, and a process $\big(X(t),\ t \in T\big)$ is both G-stationary and continuous in L^2, then its one-variable covariance function clearly inherits continuity from its two-variable counterpart.

If $\big(X(t),\ t \in T\big)$ is a stochastic process with a finite variance, then for every $n \geq 1$, $t_1,\ldots,t_n \in T$, and complex numbers $z_1,\ldots,z_n$, one has $E\big|\sum_{j=1}^n z_j X(t_j)\big|^2 \geq 0$, which gives the relation

$$\sum_{j=1}^n \sum_{k=1}^n z_j \bar{z}_k R_X(t_j,t_k) \geq 0\,. \tag{1.4}$$

This is the nonnegative definiteness property of the covariance function. In the cases in which stationarity allows us to use a one-variable notation for the covariance function, the nonnegative definiteness property takes the form

$$\sum_{j=1}^n \sum_{k=1}^n z_j \bar{z}_k R_X(t_j - t_k) \geq 0 \tag{1.5}$$

for all $n \geq 1$, $t_1,\ldots,t_n \in T$, and complex numbers $z_1,\ldots,z_n$.

Suppose now that $G = T$ is a locally compact abelian group, and a G-stationary process $\big(X(t),\ t \in T\big)$ is continuous in L^2. By Theorem 10.1.2, we know that there is a uniquely defined finite measure μ_X on the dual group Γ of G such that

$$R_X(t) = \int_\Gamma \gamma(t)\,\mu_X(d\gamma),\ t \in T \tag{1.6}$$

(recall that $\gamma \in \Gamma$ are the continuous characters of G). Since we work with real-valued stochastic processes, the covariance function is real, and the uniqueness of

μ_X guarantees that it is invariant under the transformation of the dual group Γ that sends every continuous character to its complex conjugate. The measure μ_X is the *spectral measure* of the (covariance function of the) process $\big(X(t),\, t \in T\big)$.

Applying the general representation (1.6) to the cases $G = T = \mathbb{R}^d$ and $G = T = \mathbb{Z}^d$ and using Example 10.1.1 proves the following theorem.

Theorem 1.2.1. *(i) Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ be a stationary stochastic process with a finite variance, continuous in L^2. Then there is a unique finite symmetric measure μ_X on $\mathbb{R}^d$ such that*

$$R_X(\mathbf{t}) = \int_{\mathbb{R}^d} e^{i(\mathbf{a},\mathbf{t})}\, \mu_X(d\mathbf{a}),\ \mathbf{t} \in \mathbb{R}^d\,. \tag{1.7}$$

(ii) Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{Z}^d\big)$ be a stationary stochastic process with a finite variance. Then there is a unique finite symmetric measure μ_X on $(-\pi,\pi]^d$ such that

$$R_X(\mathbf{t}) = \int_{(-\pi,\pi]^d} e^{i(\mathbf{a},\mathbf{t})}\, \mu_X(d\mathbf{a}),\ \mathbf{t} \in \mathbb{Z}^d\,. \tag{1.8}$$

In the second part of the theorem, the symmetry of the spectral measure μ_X is understood to mean invariance under the map $(-\pi,\pi]^d \to (-\pi,\pi]^d$ with $\mathbf{a} \to \big(H(a^{(1)}),\ldots,H(a^{(d)})\big)$ for $\mathbf{a} = \big(a^{(1)},\ldots,a^{(d)}\big)$, and $H : (-\pi,\pi] \to (-\pi,\pi]$ is defined by

$$H(a) = \begin{cases} -a \text{ if } -\pi < a < \pi, \\ \pi \quad \text{if } a = \pi. \end{cases}$$

The measure μ_X in both parts of Theorem 1.2.1 is also called the spectral measure of the (covariance function of the) process. Part (i) of Theorem 1.2.1 with $d = 1$ is often referred to as the *Bochner theorem*, while part (ii) of Theorem 1.2.1 with $d = 1$ is often referred to as the *Hergoltz theorem*.

Example 1.2.2. Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ be a stochastic process with a finite variance, continuous in L^2. Suppose that the process is G-stationary with respect to the group G of rigid motions of $\mathbb{R}^d$ (see Example 1.1.6). That is, $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is both stationary and isotropic. Let μ_X be its spectral measure in part (i) of Theorem 1.2.1. For every rotation $U \in \mathbf{SO}(d)$, we have by (1.3),

$$R_X(\mathbf{t}) = R_X\big(U(\mathbf{t})\big) = \int_{\mathbb{R}^d} e^{i(\mathbf{a},U(\mathbf{t}))}\, \mu_X(d\mathbf{a})$$

$$= \int_{\mathbb{R}^d} e^{i(U^{-1}(\mathbf{a}),\mathbf{t})}\, \mu_X(d\mathbf{a}) = \int_{\mathbb{R}^d} e^{i(\mathbf{a},\mathbf{t})}\, \mu_X \circ U(d\mathbf{a})$$

for every $\mathbf{t} \in \mathbb{R}^d$. By the uniqueness of the spectral measure, we conclude that $\mu_X = \mu_X \circ U$ for all $U \in \mathbf{SO}(d)$, and so the spectral measure of $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is invariant under the rotations of $\mathbb{R}^d$.

If the spectral measure of a stationary process in either part of Theorem 1.2.1 has a density h_X with respect to the d-dimensional Lebesgue measure λ_d, then h_X is called the *spectral density* of the (covariance function of the) process. It a symmetric (outside of a set of λ_d-measure 0) function such that

$$R_X(\mathbf{t}) = \int_{\mathbb{R}^d} e^{i(\mathbf{a},\mathbf{t})}\, h_X(\mathbf{a})\, d\mathbf{a},\ \mathbf{t} \in \mathbb{R}^d\,, \tag{1.9}$$

if $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is a stationary stochastic process with a finite variance, continuous in L^2, and

$$R_X(\mathbf{t}) = \int_{(-\pi,\pi)^d} e^{i(\mathbf{a},\mathbf{t})}\, h_X(\mathbf{a})\, d\mathbf{a},\ \mathbf{t} \in \mathbb{Z}^d\,, \tag{1.10}$$

if $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{Z}^d\big)$ is a stationary stochastic process with a finite variance.

A spectral density always exists if the covariance function of the process decays sufficiently fast to zero at large lags.

Proposition 1.2.3. *(*i*) Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{Z}^d\big)$ be a stationary stochastic process with a finite variance. Assume that*

$$\sum_{\mathbf{t}\in\mathbb{Z}^d} \Big|R_X(\mathbf{t})\Big| < \infty\,. \tag{1.11}$$

Then the process has a bounded and continuous spectral density given by

$$h_X(\mathbf{a}) = (2\pi)^{-d} \sum_{\mathbf{t}\in\mathbb{Z}^d} e^{i(\mathbf{t},\mathbf{a})} R_X(\mathbf{t}) \tag{1.12}$$

for $\mathbf{a} \in (-\pi, \pi)^d$.

*(*ii*) Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ be a stationary stochastic process with a finite variance, continuous in L^2. Assume that*

$$\int_{\mathbb{R}^d} \Big|R_X(\mathbf{t})\Big|\, d\mathbf{t} < \infty\,. \tag{1.13}$$

Then the process has a bounded and continuous spectral density given by

$$h_X(\mathbf{a}) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{i(\mathbf{t},\mathbf{a})} R_X(\mathbf{t})\, d\mathbf{t} \tag{1.14}$$

for $\mathbf{a} \in \mathbb{R}^d$.

Proof. (i) For $\mathbf{a} = \left(a^{(1)}, \ldots, a^{(d)}\right) \in (-\pi, \pi)^d$, the function h_X defined by the right-hand side of (1.12) is well defined and is clearly bounded and continuous. Since the covariance function R_X is nonnegative definite, for every $K = 1, 2 \ldots$,

$$0 \le \frac{1}{K^d} \sum_{\mathbf{t} \in \{-K, \ldots, K\}^d} \sum_{\mathbf{s} \in \{-K, \ldots, K\}^d} e^{i(\mathbf{t}-\mathbf{s}, \mathbf{a})} R_X(\mathbf{t}-\mathbf{s})$$

$$= \sum_{t^{(1)}=-K}^{K} \cdots \sum_{t^{(d)}=-K}^{K} \frac{1}{K^d} \prod_{m=1}^{d} \left(2K - |t^{(m)}|\right) e^{i \sum_{j=1}^{d} t^{(j)} j a^{(j)}} R_X\left((t^{(1)}, \ldots, t^{(d)})\right),$$

and the last expression converges to $2^d h_X(\mathbf{a})$ as $K \to \infty$. Hence h_X is a nonnegative function. Finally, for every $\mathbf{t} \in \mathbb{Z}^d$,

$$\int_{(-\pi,\pi)^d} e^{i(\mathbf{a},\mathbf{t})} h_X(\mathbf{a})\, d\mathbf{a} = \sum_{\mathbf{s} \in \mathbb{Z}^d} R_X(\mathbf{s}) (2\pi)^{-d} \int_{(-\pi,\pi)^d} e^{i(\mathbf{a},\mathbf{t}+\mathbf{s})}\, d\mathbf{a}$$

$$= R_X(-\mathbf{t}) = R_X(\mathbf{t}),$$

since

$$\int_{-\pi}^{\pi} e^{ian}\, da = \begin{cases} 2\pi & \text{if } n = 0, \\ 0 & \text{if } n \in \mathbb{Z},\ n \ne 0. \end{cases}$$

Therefore, h_X is the spectral density of the process $\left(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{Z}^d\right)$.

(ii) Notice that the L^2 continuity of the process implies continuity, hence measurability, of the covariance function, so the condition (1.13) makes sense. As in part (i), for $\mathbf{a} = \left(a^{(1)}, \ldots, a^{(d)}\right) \in \mathbb{R}^d$, the function h_X defined by the right-hand side of (1.14) is well defined, bounded, and continuous. By the symmetry of the covariance function, h_X is also real-valued. If μ_X is the spectral measure of the process, then all we need to prove is that for all $-\infty < y_j < z_j < \infty$, $j = 1, \ldots, d$,

$$\int_{\prod_{j=1}^{d} (y_j, z_j]} h(\mathbf{a})\, d\mathbf{a} = \mu_X\left(\prod_{j=1}^{d} (y_j, z_j]\right), \tag{1.15}$$

and it is clearly enough to prove (1.15) for "boxes" $\prod_{j=1}^{d} (y_j, z_j]$ such that μ_X assigns zero measure to the set $\{\mathbf{x} \in \mathbb{R}^d : x^{(j)} = -y_j \text{ or } -z_j\} = 0$ for each $j = 1, \ldots, d$.

The integrability assumption (1.13) justifies the application of Fubini's theorem in

$$\int_{\prod_{j=1}^d (y_j,z_j]} h(\mathbf{a})\,d\mathbf{a} = (2\pi)^{-d} \int_{\mathbb{R}^d} R_X(\mathbf{t}) \left[\int_{\prod_{j=1}^d (y_j,z_j]} e^{i(\mathbf{t},\mathbf{a})}\,d\mathbf{a} \right] d\mathbf{t}$$

$$= (2\pi)^{-d} \lim_{T\to\infty} \int_{[-T,T]^d} R_X(\mathbf{t}) \prod_{j=1}^d \frac{e^{it^{(j)}z_j} - e^{it^{(j)}y_j}}{it^{(j)}}\,d\mathbf{t}\,.$$

For every $T > 0$, another application of Fubini's theorem gives us

$$\int_{[-T,T]^d} R_X(\mathbf{t}) \prod_{j=1}^d \frac{e^{it^{(j)}z_j} - e^{it^{(j)}y_j}}{it^{(j)}}\,d\mathbf{t}$$

$$= \int_{\mathbb{R}^d} \left[\int_{[-T,T]^d} e^{i(\mathbf{t},\mathbf{w})} \left(\prod_{j=1}^d \frac{e^{it^{(j)}z_j} - e^{it^{(j)}y_j}}{it^{(j)}} \right) d\mathbf{t} \right] \mu_X(d\mathbf{w})$$

$$= \int_{\mathbb{R}^d} \prod_{j=1}^d \big(\varphi_T(z_j + w^{(j)}) - \varphi_T(y_j + w^{(j)})\big)\,\mu_X(d\mathbf{w})\,,$$

where for $T > 0$ and $h \in \mathbb{R}$,

$$\varphi_T(h) = \int_{-T}^T \frac{\sin(th)}{t}\,dt\,.$$

Notice that $\varphi_T(0) = 0$, $\varphi_T(-h) = -\varphi_T(h)$ for $h > 0$, and further, for $h \in (0,\infty)$, $\varphi_T(h) = \varphi_{Th}(1)$ is a uniformly (for $T > 0$ and $h > 0$) bounded function satisfying $\lim_{T\to\infty} \varphi_T(h) = \pi$. It follows that the function of $T > 0$ and $\mathbf{w} \in \mathbb{R}^d$

$$\prod_{j=1}^d \big(\varphi_T(z_j + w^{(j)}) - \varphi_T(y_j + w^{(j)})\big)$$

is uniformly bounded and converges, as $T \to \infty$, to

$$2\pi\,\mathbf{1}_{\prod_{j=1}^d (-z_j,-y_j)}(\mathbf{w})\,,$$

apart from the points in the set $\{\mathbf{w} \in \mathbb{R}^d : w^{(j)} = -y_j \text{ or } -z_j \text{ for some } j = 1,\dots,d\}$, which has, by the assumption, μ_X measure zero.

By the bounded convergence theorem, we obtain

$$\lim_{T\to\infty} \int_{[-T,T]^d} R_X(\mathbf{t}) \prod_{j=1}^d \frac{e^{it^{(j)}z_j} - e^{it^{(j)}y_j}}{it^{(j)}}\,d\mathbf{t}$$

$$= (2\pi)^d \mu_X\Big(\prod_{j=1}^{d}(-z_j, -y_j)\Big) = (2\pi)^d \mu_X\Big(\prod_{j=1}^{d}[-z_j, -y_j)\Big)$$

$$= (2\pi)^d \mu_X\Big(\prod_{j=1}^{d}(y_j, z_j]\Big)$$

by the symmetry of the spectral measure. The relation (1.15) follows. Therefore, h_X is the spectral density of the process $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$. □

We finish this section with an example presenting several frequently encountered one-dimensional spectral measures and densities.

Example 1.2.4. The first two examples refer to a continuous-time stationary stochastic process $\big(X(t),\ -\infty < t < \infty\big)$ with a finite variance.

If $R_X(t) = e^{-b^2t^2/2}$, $t \in \mathbb{R}$, $b > 0$ (the so-called Gaussian covariance function), then the spectral measure of the process is, of course, none other than the law of a zero-mean Gaussian random variable with variance b^2. Therefore, the process has in this case a spectral density given by

$$h_X(a) = \frac{1}{b\sqrt{2\pi}} e^{-a^2/(2b^2)},\ -\infty < a < \infty\,.$$

If $R_X(t) = e^{-b|t|}$, $t \in \mathbb{R}$, $b > 0$ (the so-called Ornstein–Uhlenbeck covariance function), then the spectral measure of the process is the law of a Cauchy random variable with scale b; see Samorodnitsky and Taqqu (1994). Therefore, in this case the process has a spectral density given by

$$h_X(a) = \frac{b}{\pi}\frac{1}{b^2 + a^2},\ -\infty < a < \infty\,.$$

The final example is that of the fractional Gaussian noise, a centered discrete-time Gaussian process $\big(X(t),\ t = \ldots, -1, 0, 1, 2, \ldots\big)$ with covariance function

$$R_X(t) = \frac{\sigma^2}{2}\Big[|t+1|^{2H} + |t-1|^{2H} - 2|t|^{2H}\Big],\ t \in \mathbb{Z} \tag{1.16}$$

for $\sigma > 0$ and $0 < H < 1$ (*the Hurst exponent*). This is a legitimate covariance function, as we will see in Section 5.1.

We claim that the fractional Gaussian noise has a spectral density, and if $H \neq 1/2$, the density is given by

$$h_X(a) = C(H)\sigma^2(1 - \cos a)\sum_{j=-\infty}^{\infty} |2\pi j + a|^{-(1+2H)},\ -\pi < a < \pi\,, \tag{1.17}$$

where

$$C(H) = \frac{H(1-2H)}{\Gamma(2-2H)\cos(\pi H)}\,.$$

Notice that if $H = 1/2$, then (1.16) reduces to $R_X(0) = \sigma^2$ and $R_X(t) = 0$ for $t \neq 0$, so that the fractional Gaussian noise with the Hurst exponent $H = 1/2$ is simply an i.i.d. centered Gaussian sequence with variance σ^2. As such, it has a constant spectral density $h_X(a) = \sigma^2/(2\pi)$, $a \in (-\pi, \pi)$.

In order to check the validity of (1.17) for $H \neq 1/2$, observe first that the function defined in (1.17) is symmetric, bounded, and continuous away from the origin. Its behavior at the origin is determined by the term in the sum corresponding to $j = 0$. Since

$$(1-\cos a)|a|^{-(1+2H)} \sim \frac{1}{2}|a|^{1-2H},\ a \to 0\,,$$

we see that the function h_X defined in (1.17) is integrable. Next, for every $t \in \mathbb{Z}$,

$$\begin{aligned}\int_{-\pi}^{\pi} e^{ita} h_X(a)\,da &= C(H)\sigma^2 \int_{-\pi}^{\pi} \cos(ta)(1-\cos a) \sum_{j=-\infty}^{\infty} |2\pi j + a|^{-(1+2H)}\,da \\ &= C(H)\sigma^2 \sum_{j=-\infty}^{\infty} \int_{\pi(2j-1)}^{\pi(2j+1)} \cos(ta)(1-\cos a)|a|^{-(1+2H)}\,da \\ &= C(H)\sigma^2 \int_{-\infty}^{\infty} \cos(ta)(1-\cos a)|a|^{-(1+2H)}\,da\,.\end{aligned}$$

We will use the well-known integral formula

$$\int_0^{\infty} a^{-2H} \sin a\,da = \frac{\Gamma(2-2H)\cos(\pi H)}{1-2H} \tag{1.18}$$

for $0 < H < 1$, $H \neq 1/2$; see, e.g., (7.3.8) and (7.3.9) in Titchmarsh (1986). Since

$$\cos(ta)(1-\cos a) = \frac{1}{2}\big[\big(1-\cos((t-1)a)\big) - 2\big(1-\cos(ta)\big) + \big(1-\cos((t+1)a)\big)\big]\,,$$

we can continue the above computation to obtain

$$\begin{aligned}&2C(H)\sigma^2\Big[|t+1|^{2H} + |t-1|^{2H} - |t|^{2H}\Big]\int_0^{\infty} a^{-(1+2H)}(1-\cos a)\,da \\ &= \frac{\sigma^2}{2}\Big[|t+1|^{2H} + |t-1|^{2H} - 2|t|^{2H}\Big]\,.\end{aligned}$$

Since this coincides with the covariance function of the fractional Gaussian noise in (1.16), it follows that the function h_X defined in (1.17) is the spectral density of this process.

1.3 Measurability and Continuity in Probability

A continuous-time stationary stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ may lack even very basic regularity properties, as the following example shows.

Example 1.3.1. The so-called i.i.d. process is a process for which every finite-dimensional distribution is the distribution of the appropriate number of i.i.d. random variables. As an illustration, consider that case in which these i.i.d. random variables are standard normal. Then the process is an uncountable collection of i.i.d. standard normal random variables. Clearly, this process is very irregular: it is not continuous in probability, and its realizations are unbounded on every infinite set of time points.

In spite of the previous example, most stationary and stationary increment processes one encounters possess as least some level of regularity. For example, a certain degree of regularity is guaranteed for stationary and stationary increment processes that are also *measurable*.

Definition 1.3.2. A stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ defined on a probability space $\big(\Omega, \mathcal{F}, P\big)$ is measurable if the map $X : \mathbb{R}^d \times \Omega \to \mathbb{R}$ is product measurable.

Two stochastic processes, one measurable and the other not measurable, can have the same finite-dimensional distributions. Nonetheless, the finite-dimensional distributions of a process determine whether the process has a measurable version. Explicit necessary and sufficient conditions on the finite-dimensional distributions of a process for a measurable version to exist can be found in Section 9.4 of Samorodnitsky and Taqqu (1994). When a measurable version of a process exists, we always identify the process with such a version, and, with some abuse of terminology, simply call that process measurable.

Theorem 1.3.3. *Every measurable stochastic process* $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ *with stationary increments is continuous in probability.*

Proof. Let $\mu_{\mathbf{X}}$ be the "infinite law," described in Proposition 1.1.11, of the shifted process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ on $\mathbb{R}^{\mathbb{R}^d}$ equipped with the cylindrical σ-field.

Consider the σ-finite measure space $\big(\mathbb{R}^{\mathbb{R}^d}, \mu_{\mathbf{X}}\big)$, and let U_0 be the group of invertible measure-preserving transformations on that space. The group operation is the composition of two transformations, which we denote by $\circ$. We equip U_0 with the metric

$$\rho(U,V)=\sum_{n=1}^{\infty}2^{-n}\big(\|f_n\circ U-f_n\circ V\|_2+\|f_n\circ U^{-1}-f_n\circ V^{-1}\|_2\big)\,, \tag{1.19}$$

where (f_n) is an orthonormal basis in the separable Hilbert space $L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$. Then (U_0,ρ) is a complete separable metric space; both the composition of two transformations and taking the inverse of a transformation are continuous in the metric ρ. In short, U_0 equipped with the metric ρ is a *Polish group*; see Aaronson (1997), pp. 13–14.

By Proposition 1.1.11, we can define a map $T:\mathbb{R}^d\to U_0$ by

$$\big(T(\mathbf{s})\big)\big((x(\mathbf{t}),\,\mathbf{t}\in\mathbb{R}^d)\big)=\big((x(\mathbf{t}+\mathbf{s}),\,\mathbf{t}\in\mathbb{R}^d)\big),\,\mathbf{s}\in\mathbb{R}^d\,. \tag{1.20}$$

This map satisfies $T(\mathbf{s}_1+\mathbf{s}_2)=T(\mathbf{s}_1)\circ T(\mathbf{s}_2)$ for $\mathbf{s}_1,\mathbf{s}_2\in\mathbb{R}^d$, i.e., T is a group homomorphism from $\mathbb{R}^d$ to U_0. We claim that this map is also Borel measurable. To show this, it is enough to prove that for every open ball B in the metric ρ, the set $\big\{\mathbf{s}\in\mathbb{R}^d:\,T(\mathbf{s})\in B\big\}$ is measurable. By the definition of the metric ρ, this will follow if we show that for every fixed functions (g_n), (h_n) in $L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$ and $\epsilon>0$, the set

$$\left\{\mathbf{s}\in\mathbb{R}^d:\,\sum_{n=1}^{\infty}2^{-n}\big(\|f_n\circ T(\mathbf{s})-g_n\|_2+\|f_n\circ T(-\mathbf{s})-h_n\|_2\big)<\epsilon\right\} \tag{1.21}$$

is measurable.

To this end, we will check that for every function $f\in L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$, the map T from $\mathbb{R}^d$ to $L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$ that maps a point $\mathbf{s}\in\mathbb{R}^d$ to the shifted function f,

$$\big(T(\mathbf{s})f\big)\big((x(\mathbf{t}),\,\mathbf{t}\in\mathbb{R}^d)\big)=f\big((x(\mathbf{t}+\mathbf{s}),\,\mathbf{t}\in\mathbb{R}^d)\big),$$

is Borel measurable. Indeed, such measurability will imply that the function $\mathbf{s}\to\|f\circ T(\mathbf{s})-g\|$ is measurable for every $f,g\in L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$, and the sum in (1.21) is a countable sum of measurable functions, hence itself measurable. By the definition of the measure $\mu_{\mathbf{X}}$, the measurability of the shift T on $L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$ will follow once we prove that for all $f,g\in L^2\big(\mathbb{R}^{\mathbb{R}^d},\mu_{\mathbf{X}}\big)$ and $\epsilon>0$, the set

$$\left\{\mathbf{s}\in\mathbb{R}^d:\,\int_{-\infty}^{\infty}E\Big(f\big(u+(X(\mathbf{t}+\mathbf{s}),\mathbf{t}\in\mathbb{R}^d)\big)-g\big(u+(X(\mathbf{t}),\mathbf{t}\in\mathbb{R}^d)\big)\Big)^2du<\epsilon\right\}$$

is Borel measurable. However, this is an immediate consequence of Fubini's theorem once we notice that by the measurability of the process $\big(X(\mathbf{t}),\,\mathbf{t}\in\mathbb{R}^d\big)$, the map

$$\big(u,\omega,\mathbf{s}\big)\to\big(u+X(\mathbf{t}+\mathbf{s}),\mathbf{t}\in\mathbb{R}^d\big)$$

from $\mathbb{R}\times\Omega\times\mathbb{R}^d$ to $\mathbb{R}^{\mathbb{R}^d}$ is measurable.

Since measurable group homomorphisms of a locally compact group and a separable topological group ($\mathbb{R}^d$ and U_0, respectively, in this case) are continuous (Corollary I.9 in Neeb (1997)), the map (1.20) is continuous. By the definition of the metric (1.19), this means that for every function f among the basis functions f_n in (1.19), and for every $\mathbf{t} \in \mathbb{R}^d$,

$$\int_{-\infty}^{\infty} E\Big(f\big(u + X(\cdot + \mathbf{t})\big) - f\big(u + X(\cdot + \mathbf{s})\big)\Big)^2 \, du \to 0 \ \text{ as } \mathbf{s} \to \mathbf{t}. \tag{1.22}$$

Since every function $f \in L^2\big(\mathbb{R}^{\mathbb{R}^d}, \mu_{\mathbf{X}}\big)$ of norm 1 can be chosen to be in a basis of $L^2\big(\mathbb{R}^{\mathbb{R}^d}, \mu_{\mathbf{X}}\big)$, we conclude that (1.22) holds for all $f \in L^2\big(\mathbb{R}^{\mathbb{R}^d}, \mu_{\mathbf{X}}\big)$.

Choosing $f\big((x(\mathbf{r}),\ \mathbf{r} \in \mathbb{R}^d)\big) = \mathbf{1}\big(|x(0)| \le 1\big)$, the statement (1.22) becomes

$$E\big(|X(\mathbf{t}) - X(\mathbf{s})| \wedge 2\big) \to 0 \ \text{ as } \mathbf{s} \to \mathbf{t},$$

which is, of course, equivalent to saying that $X(\mathbf{s}) \to X(\mathbf{t})$ in probability. □

Remark 1.3.4. There is a simple and intuitive way to view the statement of Theorem 1.3.3. Lusin's theorem of real analysis says that a measurable function is "nearly continuous"; see, e.g., Folland (1999). This allows for a "small number" of "bad points." Since each point of a process with stationary increment is equally "good" or "bad" as far as continuity in probability is concerned, it is easy to believe that every point must be a point of continuity in probability.

Remark 1.3.5. Note that Theorem 1.3.3 guarantees only that a measurable stochastic process with stationary increments is continuous in probability. No claim regarding sample continuity is made, and in fact, there exist measurable stationary processes $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ whose sample functions are, on an event of probability 1, unbounded in every d-dimensional ball of a positive radius; see e.g. Maejima (1983).

Interestingly, sometimes continuity and almost sure unboundedness in every ball of positive radius are the only options for a measurable stationary stochastic process. For example, for measurable stationary Gaussian processes, this is the statement of Belyayev's theorem; see Itô and Nisio (1968).

1.4 Linear Processes

Linear processes, otherwise known as infinite moving average processes, are discrete-time processes of the form

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\, \varepsilon_j = \sum_{j=-\infty}^{\infty} \varphi_j\, \varepsilon_{n-j}\,, \quad n = 1, 2, \dots\,, \tag{1.23}$$

where $(\varepsilon_n,\ n = \ldots, -1, 0, 1, 2, \ldots)$ are i.i.d. *noise variables*, or *innovations*, and (φ_n) are deterministic coefficients. The coefficients, clearly, have to satisfy certain conditions for the series to converge and the process to be well defined. It is obvious that whenever the process is well defined, it is stationary. If $\varphi_j = 0$ for all $j < 0$, then the moving average is sometimes called *causal* (with respect to the noise sequence (ε_n)). In that case, each X_n is a function of the noise variables ε_j with $j \le n$. Similarly, if $\varphi_j = 0$ for all $j > 0$, then the moving average is sometimes called *purely noncausal* (with respect to the noise sequence). In that case, each X_n is a function of the noise variables ε_j with $j \ge n$.

Linear processes form a very attractive class of stationary processes because of their clear and intuitive (though not necessarily simple) structure. As a result, they have been very well studied.

The actual conditions needed for the series (1.23) to converge depend mostly on how heavy the tails of the noise variables are. In the situation that the noise variables are known to possess a finite moment of a certain order, the next theorem provides explicit sufficient conditions for convergence of that series. Let ε be a generic noise variable.

Theorem 1.4.1. *Suppose that $E|\varepsilon|^p < \infty$ for some $p > 0$.*

(i) *If $0 < p \le 1$, then the condition*

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^p < \infty \tag{1.24}$$

is sufficient for convergence of the series (1.23).

(ii) *If $1 < p \le 2$ and $E\varepsilon = 0$, then condition* (1.24) *is sufficient for convergence of the series* (1.23). *If $E\varepsilon \ne 0$, then* (1.24) *and the condition*

$$\text{the series } \sum_{j=-\infty}^{\infty} \varphi_j \text{ converges} \tag{1.25}$$

are sufficient for convergence of the series (1.23).

(iii) *If $p > 2$ and $E\varepsilon = 0$, then the condition*

$$\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty \tag{1.26}$$

is sufficient for convergence of the series (1.23). *If $E\varepsilon \ne 0$, then conditions* (1.26) *and* (1.25) *are sufficient for convergence of the series* (1.23).

Under the sufficient conditions for convergence of the series (1.23) *in all three parts of the theorem, one also has $E|X_n|^p < \infty$, $n = 1, 2, \ldots$.*

Proof. We begin by assuming that $E\varepsilon = 0$ whenever $p > 1$. We will prove that the series (1.23) converges in L^p. This will imply $E|X_n|^p < \infty$, $n = 1, 2, \ldots$. Furthermore, it will also imply convergence of the series (1.23) in probability, and for series of independent random variables, convergence in probability implies almost sure convergence.

In order to prove the L^p convergence of the series (1.23), we need to show that

$$\lim_{m\to\infty} \sup_{k\geq 0} E \left| \sum_{m\leq|j|\leq m+k} \varphi_j \varepsilon_j \right|^p = 0 \,. \tag{1.27}$$

We will prove (1.27) by induction on $K = 1, 2, \ldots$, by considering $K - 1 < p \leq K$.

Suppose first that $K = 1$. We have by the triangle inequality in L^p that

$$E \left| \sum_{m\leq|j|\leq m+k} \varphi_j \varepsilon_j \right|^p \leq E \sum_{m\leq|j|\leq m+k} |\varphi_j|^p |\varepsilon_j|^p = E|\varepsilon|^p \sum_{m\leq|j|\leq m+k} |\varphi_j|^p \,,$$

and (1.27) follows from (1.24).

Next, suppose that $K = 2$. Since we are assuming that $E\varepsilon = 0$, we may use the Marcinkiewicz–Zygmund inequalities of Theorem 10.7.2 to obtain

$$E \left| \sum_{m\leq|j|\leq m+k} \varphi_j \varepsilon_j \right|^p \leq B_p E \left(\sum_{m\leq|j|\leq m+k} \varphi_j^2 \varepsilon_j^2 \right)^{p/2} . \tag{1.28}$$

Since $p/2 \leq 1$, we may use, once again, the triangle inequality in $L^{p/2}$ to conclude that

$$E \left| \sum_{m\leq|j|\leq m+k} \varphi_j \varepsilon_j \right|^p \leq B_p E|\varepsilon|^p \sum_{m\leq|j|\leq m+k} |\varphi_j|^p \,,$$

and (1.27) follows, once again, from (1.24).

Assume now that (1.27) holds for $0 < p \leq K$, for some $K \geq 2$, and consider $K < p \leq K + 1$. Note that the Marcinkiewicz–Zygmund inequalities still apply, and (1.28) holds. Subtracting and adding $E\varepsilon^2$ inside the sum on the right-hand side, we can further bound the right-hand side of (1.28) by

$$B_p 2^{p/2-1} \big(E|\varepsilon|^2\big)^{p/2} \left(\sum_{m\leq|j|\leq m+k} \varphi_j^2 \right)^{p/2} + B_p 2^{p/2-1} E \left(\sum_{m\leq|j|\leq m+k} \varphi_j^2 (\varepsilon_j^2 - \mathbb{E}\varepsilon^2) \right)^{p/2} .$$

The assumption (1.26) implies that

$$\lim_{m\to\infty} \sup_{k\geq 0} \left(\sum_{m\leq|j|\leq m+k} \varphi_j^2 \right)^{p/2} = 0\,.$$

Furthermore, $p/2 \leq K$, and so the induction hypothesis (which we apply while replacing φ_j with φ_j^2 and ε_j with $\varepsilon_j^2 - E\varepsilon^2$) tells us that

$$\lim_{m\to\infty} \sup_{k\geq 0} E \left(\sum_{m\leq|j|\leq m+k} \varphi_j^2(\varepsilon_j^2 - \mathbb{E}\varepsilon^2) \right)^{p/2} = 0$$

as well. This proves (1.27) for $K < p \leq K+1$ and thereby completes the inductive argument.

The above argument proves the statement of the theorem if one assumes that $E\varepsilon = 0$ whenever $p > 1$. If $p > 1$ but $E\varepsilon \neq 0$, then we write (for $n = 0$)

$$\sum_{j=-\infty}^{\infty} \varphi_{-j}\, \varepsilon_j = \sum_{j=-\infty}^{\infty} \varphi_{-j}\Big(\big(\varepsilon_j - E\varepsilon\big) + E\varepsilon \Big)$$

and thereby reduce the situation to the case already considered. □

A partial converse to the statement of Theorem 1.4.1 is in Exercise 1.6.2. See also Exercise 1.6.3.

Remark 1.4.2. Note that we have actually proved that in the case $p \leq 1$, the series (1.23) converges absolutely in L^p and with probability 1. In the case $p > 1$, absolute convergence may not hold, but the series converges *unconditionally*. This means that for every *deterministic permutation* of the terms of the series, the resulting series converges in L^p and with probability 1, and the limit is almost surely equal to the sum of the original series.

Let $\big(X_n,\, n \in \mathbb{Z}\big)$ be a linear process (1.23), and suppose that the noise variables have a finite second moment. The conditions of part (iii) of Theorem 1.4.1 are, according to Exercise 1.6.2, necessary and sufficient for the linear process to be well defined in this case; the fact that they are satisfied will be assumed in the sequel every time we deal with finite-variance linear processes. The series defining the process converges in L^2, and therefore, the linear process has the covariance function

$$R_X(n) = \mathrm{Var}(\varepsilon) \sum_{j=-\infty}^{\infty} \varphi_j \varphi_{j+n}, \quad n = 0, 1, \ldots\,. \tag{1.29}$$

It turns out that a finite-variance linear process has a spectral density, as described in the following theorem.

Theorem 1.4.3. *Let* $\big(X_n,\ n \in \mathbb{Z}\big)$ *be a linear process* (1.23) *with a finite variance. Then the process has a spectral density given by*

$$h(a) = \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|\sum_{j=-\infty}^{\infty} \varphi_j e^{ija}\right|^2,\ a \in (-\pi, \pi)\,, \tag{1.30}$$

where the series in (1.30) *converges in* $L^2\big((-\pi.\pi), \lambda_1\big)$.

Proof. For $m \geq 1$, consider a finite-variance linear process $\big(X_n^{(m)},\ n \in \mathbb{Z}\big)$ with finitely many nonzero coefficients, $\varphi_n^{(m)} = \varphi_n \mathbf{1}(|n| \leq m)$, $n \in \mathbb{Z}$. Define also

$$h^{(m)}(a) = \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|\sum_{j=-m}^{m} \varphi_j e^{ija}\right|^2 := \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|g^{(m)}(a)\right|^2,\ a \in (-\pi, \pi)\,.$$

Note that for every $n = 0, 1, \ldots,$

$$\int_{-\pi}^{\pi} e^{ian} h^{(m)}(a)\, da = \frac{\mathrm{Var}(\varepsilon)}{2\pi} \sum_{j_1=-m}^{m} \sum_{j_2=-m}^{m} \varphi_{j_1}\varphi_{j_2} \int_{-\pi}^{\pi} e^{ia(n+j_1-j_2)}\, da$$

$$= \frac{\mathrm{Var}(\varepsilon)}{2\pi} \sum_{j_1=-m}^{m} \sum_{j_2=-m}^{m} \varphi_{j_1}\varphi_{j_2} \mathbf{1}\big(j_2 - j_1 = n\big)(2\pi) = R_{X^{(m)}}(n)$$

by (1.29). That is, $h^{(m)}$ is the spectral density of the linear process $\big(X_n^{(m)},\ n \in \mathbb{Z}\big)$. For $m, k \geq 1$, $\big(X_n^{(m+k)} - X_n^{(m)},\ n \in \mathbb{Z}\big)$ is also a finite-variance moving average, with finitely many nonzero coefficients, $\big(\varphi_n \mathbf{1}(m < |n| \leq m + k),\ n \in \mathbb{Z}\big)$, and hence has the spectral density

$$h^{(m,k)}(a) = \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|g^{(m+k)}(a) - g^{(m)}(a)\right|^2,\ a \in (-\pi, \pi)\,.$$

Therefore,

$$\mathrm{Var}\Big(X_0^{(m+k)} - X_0^{(m)}\Big) = \frac{\mathrm{Var}(\varepsilon)}{2\pi} \int_{-\pi}^{\pi} h^{(m,k)}(a)\, da \tag{1.31}$$

$$= \frac{\mathrm{Var}(\varepsilon)}{2\pi} \int_{-\pi}^{\pi} \left|g^{(m+k)}(a) - g^{(m)}(a)\right|^2 da\,.$$

Since by the L^2 convergence of the series (1.23), the right-hand side of (1.31) converges to zero as $m \to \infty$, uniformly in k, we conclude that

$$\lim_{m\to\infty} \sup_{k\geq 0} \int_{-\pi}^{\pi} \left| g^{(m+k)}(a) - g^{(m)}(a) \right|^2 da = 0\,,$$

and so the sequence of functions

$$g^{(m)}(a) = \sum_{j=-m}^{m} \varphi_j e^{ija},\ a \in (-\pi, \pi)\,,$$

$m = 1, 2, \ldots$, converges in $L^2\big((-\pi.\pi), \lambda_1\big)$. If we denote its limit by

$$g(a) = \sum_{j=-\infty}^{\infty} \varphi_j e^{ija},\ a \in (-\pi, \pi)\,,$$

then we can use twice the L^2 convergence, first on the probability space, and then on $\big((-\pi, \pi), \lambda_1\big)$ to obtain, for every $n = 0, 1, \ldots$,

$$\begin{aligned} R_X(n) &= \lim_{m\to\infty} R_{X^{(m)}}(n) = \lim_{m\to\infty} \int_{-\pi}^{\pi} e^{ian} h^{(m)}(a)\, da \\ &= \lim_{m\to\infty} \int_{-\pi}^{\pi} e^{ian} \frac{\mathrm{Var}(\varepsilon)}{2\pi} \left| g^{(m)}(a) \right|^2 da = \int_{-\pi}^{\pi} e^{ian} \frac{\mathrm{Var}(\varepsilon)}{2\pi} |g(a)|^2\, da \\ &= \int_{-\pi}^{\pi} e^{ian} h(a)\, da\,, \end{aligned}$$

which shows that h is the spectral density of the finite-variance linear process $\big(X_n,\ n \in \mathbb{Z}\big)$. □

The function $g(a) = \sum_{j=-\infty}^{\infty} \varphi_j e^{ija}$, $-\pi < a < \pi$, is sometimes called the *transfer function of the linear filter defined by the coefficients* (φ_n), and the function $|g|^2$ is called the *power transfer function* of that filter.

Example 1.4.4. A special class of linear processes consists of *stationary AutoRegressive Moving Average*, or *ARMA*, *processes*. Let $r, q \geq 0$ be two nonnegative integers, and let $(\phi_0, \ldots, \phi_r)$ and $(\theta_0, \ldots, \theta_q)$ be real numbers such that $\phi_0 = \theta_0 = 1$.

Given a sequence $\big(\varepsilon_n,\ n \in \mathbb{Z}\big)$ of i.i.d. random variables, a stationary ARMA(r, q) process $\big(X_n\ n \in \mathbb{Z}\big)$ is defined to be a stationary process that satisfies the recurrence

$$\sum_{j=0}^{r} \phi_j X_{n-j} = \sum_{j=0}^{q} \theta_j \varepsilon_{n-j} \quad \text{for all } n \in \mathbb{Z}. \tag{1.32}$$

We will assume that the noise variables satisfy $E|\varepsilon|^p < \infty$ for some $p > 0$.

In order to answer the obvious existence and uniqueness questions, and to see the connection with the linear processes, we introduce two polynomials: the *autoregressive polynomial* ϕ and the *moving average polynomial* θ, of degrees r and q, respectively, defined by

$$\phi(x) = \sum_{j=0}^{r} \phi_j x^j, \quad \theta(x) = \sum_{j=0}^{q} \theta_j x^j, \quad x \in \mathbb{C}\,.$$

We assume that

$$\phi(x) \neq 0 \ \text{ for all } x \in \mathbb{C} \text{ with } |x| = 1. \tag{1.33}$$

Under the assumption (1.33), the polynomial ϕ does not vanish in the annulus $\{x \in \mathbb{C} : R^{-1} < |x| < R\}$ for some $R > 1$. Therefore, $1/\phi$ is an analytic function in that region and hence has a power series expansion

$$\frac{1}{\phi(x)} = \sum_{j=-\infty}^{\infty} \psi_j x^j, \quad x \in \mathbb{C},\, R^{-1} < |x| < R\,; \tag{1.34}$$

see Section 2.3 in Chapter VI of Ahlfors (1953). In particular, the coefficients (ψ_j) decay exponentially fast: for every $\tau \in (R^{-1}, 1)$, there is a finite C such that $|\psi_j| \leq C\tau^{|j|}$ for all j. The analytic function equal identically to 1 satisfies in the annulus $\{x \in \mathbb{C} : R^{-1} < |x| < R\}$ the relation

$$1 = \sum_{j=0}^{r} \phi_j x^j \sum_{j=-\infty}^{\infty} \psi_j x^j\,,$$

and the uniqueness of the power series expansion of an analytic function means that the coefficients at the like powers of x in the two series are the same:

$$\sum_{k=0}^{r} \phi_k \psi_{j-k} = \mathbf{1}(j = 0)\,. \tag{1.35}$$

Similarly, the expansion of the analytic function θ/ϕ satisfies

$$\sum_{j=0}^{q} \theta_j x^j \sum_{j=-\infty}^{\infty} \psi_j x^j = \frac{\theta(x)}{\phi(x)} := \sum_{j=-\infty}^{\infty} \varphi_j x^j \tag{1.36}$$

in the annulus $\{x \in \mathbb{C} : R^{-1} < |x| < R\}$, and using once again the uniqueness of the power series expansion of an analytic function tells us that the coefficients of like powers of x in the two series around θ/ϕ are the same:

$$\varphi_j = \sum_{k=0}^{q} \theta_k \psi_{j-k} \quad \text{for each } j \in \mathbb{Z}. \tag{1.37}$$

We conclude that the coefficients (φ_j) decay exponentially fast as well. Furthermore, the relation

$$\theta(x) = \phi(x)\frac{\theta(x)}{\phi(x)}, \quad x \in \mathbb{C},\ R^{-1} < |x| < R,$$

between three analytic functions leads to the corresponding relation between the three series expansions in that annulus:

$$\sum_{j=0}^{q} \theta_j x^j = \sum_{j=0}^{r} \phi_j x^j \sum_{j=-\infty}^{\infty} \varphi_j x^j .$$

Appealing yet again to the uniqueness of the power series expansion of an analytic function, we may equate the coefficients of like powers of x to conclude that

$$\theta_j = \sum_{k=0}^{r} \phi_k \varphi_{j-k}, \quad j \in \mathbb{Z}\,, \tag{1.38}$$

which one interprets as $0 = 0$ if $j \notin \{0, \ldots, q\}$.

Recall that the noise variables satisfy $E|\varepsilon|^p < \infty$ for some $p > 0$. Theorem 1.4.1 applies regardless of the value of $p > 0$, and the infinite moving average $X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\ \varepsilon_j$, $n \in \mathbb{Z}$, in (1.23) with the coefficients given by (1.37) is well defined. Furthermore, for every $n \in \mathbb{Z}$, we can use (1.38) to see that

$$\sum_{j=0}^{r} \phi_j X_{n-j} = \sum_{j=0}^{r} \phi_j \sum_{k=-\infty}^{\infty} \varphi_k\ \varepsilon_{n-j-k}$$

$$= \sum_{j=0}^{r} \phi_j \sum_{k=-\infty}^{\infty} \varphi_{k-j}\ \varepsilon_{n-k} = \sum_{k=-\infty}^{\infty} \left(\sum_{j=0}^{r} \phi_j \varphi_{k-j} \right) \varepsilon_{n-k}$$

$$= \sum_{k=0}^{q} \theta_k \varepsilon_{n-k}\,,$$

and so the linear process with the coefficients given by (1.37) satisfies the ARMA equation (1.32).

In the other direction, suppose that $\big(X_n\ n \in \mathbb{Z}\big)$ is a stationary process satisfying the ARMA equation (1.32), and denote by W_n the random variable appearing on both the left-hand and right-hand sides of (1.32), $n \in \mathbb{Z}$. Since $E|\varepsilon|^p < \infty$, we also have $E|W|^p < \infty$. The first Borel–Canteli lemma then shows that with the

coefficients (ψ_j) defined by (1.34), the sum $\sum_j \psi_j W_{n-j}$ converges with probability 1 for each n, and, by (1.32), this sum can be written in two different ways:

$$\sum_{j=-\infty}^{\infty} \psi_j W_{n-j} = \sum_{j=-\infty}^{\infty} \psi_j \sum_{k=0}^{r} \phi_k X_{n-j-k} = \sum_{j=-\infty}^{\infty} \psi_j \sum_{k=0}^{q} \theta_k \varepsilon_{n-j-k} \,. \tag{1.39}$$

We first treat the last sum in (1.39). Interchanging twice a finite sum and a convergent sum, and using (1.37), we obtain

$$\begin{aligned} \sum_{j=-\infty}^{\infty} \psi_j \sum_{k=0}^{q} \theta_k \varepsilon_{n-j-k} &= \sum_{k=0}^{q} \theta_k \sum_{j=-\infty}^{\infty} \psi_{j-k} \varepsilon_{n-j} \\ &= \sum_{j=-\infty}^{\infty} \left(\sum_{k=0}^{q} \theta_k \psi_{j-k} \right) \varepsilon_{n-j} = \sum_{j=-\infty}^{\infty} \varphi_{n-j} \, \varepsilon_j \,. \end{aligned}$$

We would like to offer a similar treatment to the middle sum in (1.39), but a problem arises, since we do not know a priori whether the stationary process $\big(X_n \; n \in \mathbb{Z}\big)$ has any finite moments and hence cannot guarantee that a sum of the type $\sum_j \psi_j X_{n-j}$ converges. To overcome this problem, we write

$$\sum_{j=-\infty}^{\infty} \psi_j \sum_{k=0}^{r} \phi_k X_{n-j-k} = \lim_{M \to \infty} \sum_{j=-M}^{M} \psi_j \sum_{k=0}^{r} \phi_k X_{n-j-k} \,,$$

and we work first with the finite double sum above. Write for $M \geq r$,

$$\begin{aligned} \sum_{j=-M}^{M} \psi_j \sum_{k=0}^{r} \phi_k X_{n-j-k} &= \sum_{k=0}^{r} \phi_k \sum_{j=k-M}^{k+M} \psi_{j-k} X_{n-j} \\ &= \sum_{j=-\infty}^{\infty} \left(\sum_{k=\max(0,j-M)}^{\min(r,j+M)} \phi_k \psi_{j-k} \right) X_{n-j} \,. \end{aligned}$$

Appealing to (1.35), we see that for $r - M \leq j \leq M$, the sum over k on the right-hand side above is equal to $\mathbf{1}\big(j = 0\big)$. Therefore, the middle sum in (1.39) can be written as

$$\begin{aligned} X_n + \lim_{M \to \infty} \Bigg[& \sum_{j=-\infty}^{r-M-1} \left(\sum_{k=\max(0,j-M)}^{\min(r,j+M)} \phi_k \psi_{j-k} \right) X_{n-j} \\ & + \sum_{j=-M+1}^{\infty} \left(\sum_{k=\max(0,j-M)}^{\min(r,j+M)} \phi_k \psi_{j-k} \right) X_{n-j} \Bigg] \,. \end{aligned} \tag{1.40}$$

Next, observe that the inner sums over k vanish if $j < -M$ in the first sum over j and $j > M + r$ in the second sum over j. If we set $\tilde{\psi}(M) = \sup_{|j|>M-r} |\psi_j|$ and $\tilde{\phi} = \sum_{k=0}^{r} |\phi_k|$, then each of the two double sums above is stochastically bounded by $\tilde{\psi}(M)\tilde{\phi} \sum_{j=1}^{r} |X_j|$. Clearly, $\tilde{\psi}(M) \to 0$ as $M \to \infty$, and the limit in (1.40) is equal to zero, as a limit in probability. We conclude that the middle sum in (1.39) is equal to X_n, and therefore, the stationary process $\big(X_n\ n \in \mathbb{Z}\big)$ satisfies $X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\ \varepsilon_j,\ n \in \mathbb{Z}$, with the coefficients defined by (1.37). We have, therefore, the following result.

Theorem 1.4.5. *Suppose that the autoregressive polynomial ϕ has no roots on the unit circle of the complex plane (i.e., satisfies the assumption (1.33)), and that for some $p > 0$, $E|\varepsilon|^p < \infty$. Then the ARMA equation (1.32) has a unique stationary solution. The solution is a linear process*

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\ \varepsilon_j,\ n \in \mathbb{Z}$$

with the coefficients (φ_j) defined as the coefficients of the series expansion (1.36) of the ratio of the moving average and autoregressive polynomials in an annulus $\{x \in \mathbb{C} : R^{-1} < |x| < R\}$ in which the autoregressive polynomial ϕ does not vanish. Alternatively, the coefficients (φ_j) are given by (1.37). The unique stationary solution of the ARMA equation (1.32) is a process with a finite absolute pth moment.

Furthermore, if the autoregressive polynomial ϕ has no roots on or inside the unit circle of the complex plane, i.e., if

$$\phi(x) \neq 0 \text{ for all } x \in \mathbb{C} \text{ with } |x| \leq 1, \tag{1.41}$$

then $\varphi_j = 0$ for $j < 0$, and the unique stationary solution of the ARMA equation (1.32) is a causal moving average

$$X_n = \sum_{j=-\infty}^{n} \varphi_{n-j}\ \varepsilon_j,\ n \in \mathbb{Z}\,.$$

If the autoregressive polynomial ϕ has no roots on or outside of the unit circle of the complex plane, i.e., if

$$\phi(x) \neq 0 \text{ for any } x \in \mathbb{C} \text{ with } |x| \geq 1, \tag{1.42}$$

then $\varphi_j = 0$ for $j > 0$, and the unique stationary solution of the ARMA equation (1.32) is a purely noncausal moving average

$$X_n = \sum_{j=n}^{\infty} \varphi_{n-j}\ \varepsilon_j,\ n \in \mathbb{Z}\,.$$

The only parts of Theorem 1.4.5 that have not yet been proved are the facts that the unique stationary solution is purely causal under the assumption (1.41) and purely noncausal under the assumption (1.42). These, however, follow immediately from the facts that a function analytic inside a circle has a convergent series expansion inside that circle into nonnegative powers of the argument, while a function analytic outside a circle has a convergent series expansion outside that circle into nonpositive powers of the argument; see Ahlfors (1953).

If we now assume that the noise variables have a finite second moment, then Theorem 1.4.3 applies to the stationary ARMA process, and we conclude that it has a spectral density given by

$$h(a) = \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|\sum_{j=-\infty}^{\infty}\varphi_j e^{ija}\right|^2 = \frac{\mathrm{Var}(\varepsilon)}{2\pi}\left|\frac{\theta\left(e^{ia}\right)}{\phi\left(e^{ia}\right)}\right|^2$$

$$= \frac{\mathrm{Var}(\varepsilon)}{2\pi}\frac{\left|\theta\left(e^{ia}\right)\right|^2}{\left|\phi\left(e^{ia}\right)\right|^2},\ a \in (-\pi,\pi)\,. \tag{1.43}$$

Since the spectral density of a stationary ARMA process is the ratio of (the norms of) two finite polynomials of the complex exponential, one sometimes says that such processes have a rational spectral density.

1.5 Comments on Chapter 1

Comments on Section 1.3

Statements of the type "measurability implies continuity in the presence of homogeneity" have been proved in different areas of mathematics. A common reference is Banach's theorem on homomorphisms between Polish groups (Theorem 4 in Banach (1955), p. 23). Banach assumes, however, Baire measurability instead of Borel measurability, which is more common in probability theory.

Comments on Section 1.4

A wealth of information on the ARMA processes of Example 1.4.4 is in Brockwell and Davis (1991). This book also contains useful additional information on general finite-variance stationary processes.

1.6 Exercises to Chapter 1

Exercise 1.6.1. *Let* $\left(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\right)$ *be a stationary stochastic process with a finite variance, continuous in* L^2*, with a spectral measure* μ_X*. Then the restriction of the*

domain of the process to the integers, $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{Z}^d\big)$*, is also a stationary stochastic process with a finite variance. How is its spectral measure related to the "continuous time" spectral measure* μ_X*?*

Exercise 1.6.2. *Suppose that the series in* (1.23) *converges. Show that the coefficients* (φ_n) *must satisfy* (1.26)*.*

Exercise 1.6.3. *It is tempting to guess that if the series in* (1.23) *converges and* $E|\varepsilon|^p = \infty$ *for some* $0 < p < 2$*, then* (1.24) *has to hold. The following provides a counterexample.*

Let (ε_j) *be i.i.d. symmetric random variables taking values* $\pm n!$ *for* $n = 1, 2, \ldots$ *such that* $P(\varepsilon = n!) = P(\varepsilon = -n!) = c/(n+1)!$*, where* $c > 0$ *is a normalizing constant. Suppose that the sequence of the coefficients* (φ_n) *is piecewise constant, with* $(n-1)!$ *of the coefficients taking the value* $1/n!$ *for* $n = 1, 2, \ldots$*. Show that the series in* (1.23) *converges, that* $E|\varepsilon| = \infty$*, and that* (1.24) *fails for* $p = 1$*.*

Exercise 1.6.4. *Give an alternative proof of the convergence part of Theorem 1.4.1 using the three series theorem.*

Exercise 1.6.5. *Prove the following extension of Theorem 1.4.3. Let* $\mathbf{Y}$ *be a finite-variance stationary process with a bounded spectral density* f_Y*. Let* (φ_j) *be real coefficients satisfying* (1.26)*. Then*

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_j Y_{n-j},\ n = 1, 2, \ldots,$$

is a well-defined finite-variance stationary process, and it has a spectral density, a version of which is given by

$$f_X(x) = f_Y(x) \left| \sum_{j=-\infty}^{\infty} \varphi_j e^{ijx} \right|^2,\ x \in (-\pi, \pi)\,.$$

Chapter 2
Elements of Ergodic Theory of Stationary Processes and Strong Mixing

2.1 Basic Definitions and Ergodicity

Let $\big(X_n,\, n \in \mathbb{Z}\big)$ be a discrete-time stationary stochastic process. Consider the space $\mathbb{R}^{\mathbb{Z}}$ of the doubly infinite sequences $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$ of real numbers, and equip this space with the usual cylindrical σ-field $\mathcal{B}^{\mathbb{Z}}$. The stochastic process naturally induces a probability measure $\mu_{\mathbf{X}}$ on this space via

$$\mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\, \big(x_i, \ldots, x_j\big) \in B\Big\} = P\Big(\big(X_i, \ldots, X_j\big) \in B\Big) \tag{2.1}$$

for all $i \le j$ and Borel sets $B \in \mathbb{R}^{j-i+1}$. The space $\mathbb{R}^{\mathbb{Z}}$ has a natural *left shift* operation $\theta : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$. For $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{Z}}$, the shifted sequence $\theta\mathbf{x}$ is the sequence of real numbers whose ith coordinate is the $(i+1)$st coordinate x_{i+1} of $\mathbf{x}$ for each $i \in \mathbb{Z}$. Formally,

$$\theta\big((\ldots, x_{-1}, x_0, x_1, x_2, \ldots)\big) = (\ldots, x_0, x_1, x_2, x_3 \ldots)\,.$$

Clearly, the left shift is a one-to-one transformation of $\mathbb{R}^{\mathbb{Z}}$ onto itself, and both θ and its inverse, the *right shift* θ^{-1}, are measurable with respect to the cylindrical σ-field. Note that the left shift θ leaves the measure $\mu_{\mathbf{X}}$ on $\mathbb{R}^{\mathbb{Z}}$ unchanged, because for all $i \le j$ and Borel sets $B \in \mathbb{R}^{j-i+1}$,

$$\begin{aligned}
\mu_{\mathbf{X}} \circ \theta^{-1}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\, \big(x_i, \ldots, x_j\big) \in B\Big\} &= \mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\, \big(x_{i+1}, \ldots, x_{j+1}\big) \in B\Big\} \\
&= P\Big(\big(X_{i+1}, \ldots, X_{j+1}\big) \in B\Big) = P\Big(\big(X_i, \ldots, X_j\big) \in B\Big) \\
&= \mu_{\mathbf{X}}\Big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\, \big(x_i, \ldots, x_j\big) \in B\Big\}\,,
\end{aligned}$$

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_2

where the third equality follows from the stationarity of the process. In other words, the left shift preserves the measure $\mu_{\mathbf{X}}$ induced by a stationary process on $\mathbb{R}^{\mathbb{Z}}$. This is, of course, not particularly exciting. On the other hand, in spite of this preservation of the measure $\mu_{\mathbf{X}}$ by the left shift, if we choose a point (sequence) $\mathbf{x} \in \mathbb{R}^{\mathbb{Z}}$ according to the probability measure $\mu_{\mathbf{X}}$, there is no reason to expect that the *trajectory* $\theta^n\mathbf{x}$, $n = 0, 1, 2, \ldots$, of the point $\mathbf{x}$ should be in some way trivial. Here $\theta^n = \theta \circ \ldots \circ \theta$ is the composition of n left shifts, $n = 1, 2, \ldots$ (which is, of course, simply a left shift by n time units), while θ^0 is the identity operator on $\mathbb{R}^{\mathbb{Z}}$.

In fact, for most stationary stochastic processes, a "typical point" $\mathbf{x}$ selected according to the measure $\mu_{\mathbf{X}}$ follows a highly nontrivial trajectory. Such trajectories are, obviously, closely related to interesting probabilistic properties of a stationary process. Therefore, ergodic theory that studies measure-preserving transformations (as well as more general transformations) of a measure space provides an important point of view on stationary processes. In this and the following sections of this chapter, we describe certain basic notions of ergodic theory and discuss what they mean for stationary stochastic processes. Much more detail can be found in, for example, Krengel (1985) and Aaronson (1997).

We commence by noting that the connection between a stationary stochastic process $\big(X_n,\, n \in \mathbb{Z}\big)$ and the probability measure $\mu_{\mathbf{X}}$ it induces on the cylindrical σ-field on $\mathbb{R}^{\mathbb{Z}}$ is not a one-way affair, in which the stationary process, defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$, is the "primary actor" while the induced measure $\mu_{\mathbf{X}}$ is "secondary." In fact, if we start with *any* probability measure μ on $\mathbb{R}^{\mathbb{Z}}$ that is invariant under the left shift θ, then we can define a stochastic process $\big(X_n,\, n \in \mathbb{Z}\big)$ on the probability space $\big(\mathbb{Z}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ by

$$X_n(\mathbf{x}) = x_n,\ n \in \mathbb{Z},\ \text{for } \mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{Z}}, \tag{2.2}$$

and then the invariance of the measure μ causes stationarity of the process $\big(X_n,\, n \in \mathbb{Z}\big)$. Moreover, the measure $\mu_{\mathbf{X}}$ induced by this process coincides with μ. Recall also that the definition of the cylindrical σ-field shows that there is a one-to-one correspondence between shift-invariant probability measures on $\mathbb{R}^{\mathbb{Z}}$ and collections of the finite-dimensional distributions of stationary stochastic processes indexed by $\mathbb{Z}$.

We conclude that, given a collection of the finite-dimensional distributions of a stationary stochastic process, we can define a stochastic process with these finite-dimensional distributions on the space $\mathbb{R}^{\mathbb{Z}}$ equipped with the cylindrical σ-field and appropriate shift-invariant probability measure via the coordinate evaluation scheme (2.2). Since the ergodic properties of stationary stochastic processes we discuss (such as ergodicity and mixing) depend only on the finite-dimensional distributions of these processes, it is completely unimportant on what probability space a stochastic process is defined. However, the sequence space $\mathbb{R}^{\mathbb{Z}}$ has a built-in left shift operation, which provides a convenient language for discussing ergodic properties. Therefore, in this section we assume, unless stated otherwise, that a stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is defined on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$

by (2.2), and that the probability measure μ is shift-invariant. We emphasize that our conclusions about ergodic properties of stationary stochastic processes applies regardless of the actual probability space on which a process is defined.

For now, however, we consider an arbitrary σ-finite measure space $(E, \mathcal{E}, m)$. Let $\phi : E \to E$ be a measurable map. The powers of ϕ are defined in the sense of repeated application: ϕ^n is a map from E to E given by $\phi^n(x) = \phi(\phi(\ldots\phi(x)))$ for $n \geq 1$ (applying ϕ n times). The operator ϕ^0 is, by definition, the identity operator on E. A set $A \in \mathcal{E}$ is called ϕ-invariant if $m\big(A \triangle \phi^{-1}A\big) = 0$, where $\triangle$ denotes the symmetric difference of two sets. It is easy to check that the collection $\mathcal{I}$ of all ϕ-invariant sets is a sub-σ-field of $\mathcal{E}$ (see Exercise 2.6.2); we call $\mathcal{I}$ the (ϕ-) invariant σ-field. Invariant σ-fields naturally appear in the following key result of ergodic theory.

Theorem 2.1.1 (Birkhoff's pointwise ergodic theorem). *Suppose that $\phi : E \to E$ is a measurable map, preserving the measure m. Let $f \in L_1(m)$. Then there is a function $g_f : E \to \mathbb{R}$, measurable with respect to the invariant σ-field $\mathcal{I}$, such that*

$$\frac{1}{n}\sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to g_f(x) \quad \text{as } n \to \infty$$

for m-almost every $x \in E$. The function g_f satisfies $g_f \in L_1(m)$, $\|g_f\|_1 \leq \|f\|_1$, and

$$\int_A g_f(x)\, m(dx) = \int_A f(x)\, m(dx)$$

for every set $A \in \mathcal{I}$ of finite measure m.

See, e.g., Theorem 2.3 in Krengel (1985, p. 9). Note, for example, that if the measure m in Theorem 2.1.1 is actually a probability measure, then the properties of the function g_f in the theorem identify that function as the conditional expectation of f (viewed as a random variable on the probability space $(E, \mathcal{E}, m)$) given the invariant σ-field $\mathcal{I}$.

A map $\phi : E \to E$ is called *nonsingular* if ϕ is both onto and one-to-one, both ϕ and its inverse $\phi^{-1} : E \to E$ are measurable, and the induced measure $m \circ \phi^{-1}$ is equivalent to the original measure m. Clearly, if ϕ preserves the measure m, it is also nonsingular, as long as it satisfies the other requirements of nonsingularity.

Definition 2.1.2. A nonsingular map ϕ on $(E, \mathcal{E}, m)$ is called ergodic if every ϕ-invariant set A is such that either $m(A) = 0$ or $m(A^c) = 0$.

Note that every measurable set A such that either $m(A) = 0$ or $m(A^c) = 0$ is invariant for every nonsingular map (see Exercise 2.6.3). They are trivially invariant, so to speak. What distinguishes ergodic nonsingular maps is that no other measurable sets are invariant for these maps.

Example 2.1.3. Let $E = \mathbb{Z}$, let $\mathcal{E}$ be the collection of all subsets of $\mathbb{Z}$, and let m be the counting measure. Consider two nonsingular (actually, measure-preserving) maps: $\phi_1(x) = x + 1$, $\phi_2(x) = x$, $x \in \mathbb{Z}$.

Note that the only ϕ_1-invariant sets are the empty set and the entire space $\mathbb{Z}$. These are trivially invariant sets, and hence ϕ_1 is an ergodic map.

On the other hand, ϕ_2 is the identity map, so that every measurable set is invariant with respect to ϕ_2. Since this includes many nontrivially invariant sets (such as the set of all even numbers, for example), the map ϕ_2 is not ergodic.

If ϕ is a nonsingular measure-preserving ergodic map, then Exercise 2.6.4 tells us that the function g_f in Birkhoff's pointwise ergodic theorem must be a constant function regardless of what function $f \in L_1(m)$ we choose. In particular, if m is an infinite measure, then for every such f, the function g_f must be, up to a set of measure zero, the zero function, since $g_f \in L_1(m)$, and the only constant function in the space $L_1(m)$ of an infinite measure m is the zero function.

Now let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary stochastic process defined by (2.2) on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ with a shift-invariant μ. We say that the stochastic process is ergodic if the left shift θ is an ergodic map, $\mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$, in the sense of Definition 2.1.2. Since ergodicity of the left shift is determined by the probability measure μ, which is, in turn, determined by the finite-dimensional distributions of the process, the latter determine whether a given stationary process is ergodic. Notice that a stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$ is ergodic if and only if the time-reversed process $\big(X_{-n},\ n \in \mathbb{Z}\big)$ is ergodic (see Exercise 2.6.3).

Example 2.1.4. Tail and invariant σ-fields In the context of stationary stochastic processes on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$, the ($\theta$-)invariant σ-field $\mathcal{I}$ is a sub-σ-field of $\mathcal{B}^{\mathbb{Z}}$. Another important natural sub-σ-field of $\mathcal{B}^{\mathbb{Z}}$ is the *tail σ-field* $\mathcal{T}$, defined as the completion with respect to the measure μ of the σ-field

$$\bigcap_{n=1}^{\infty} \sigma\big(x_n, x_{n+1}, \dots\big)\,.$$

In general, not every tail event is an invariant event, as can be seen by choosing

$$\mu = \frac{1}{2}\delta_{(\dots,0,1,0,\dots)} + \frac{1}{2}\delta_{(\dots,1,0,1,\dots)} \tag{2.3}$$

($x_0 = 1$ in the first sequence and $x_0 = 0$ in the second sequence), and

$$A = \Big\{(\dots, x_{-1}, x_0, x_1, x_2, \dots) : x_{2n} = 1 \ \text{ for infinitely many } n = 0, 1, 2, \dots\Big\}\,.$$

On the other hand, we claim that every invariant event is a tail event, that is,

$$\mathcal{I} \subset \mathcal{T}\,. \tag{2.4}$$

To see this, let A be an invariant cylindrical set. Since the sets in a σ-field can be approximated arbitrarily closely with respect to a probability measure by the sets in a field that generates the σ-field, for every $\varepsilon > 0$ there is a finite-dimensional cylindrical set B such that

$$\mu\big(A \triangle B\big) \le \varepsilon$$

(finite-dimensionality of B means that $B \in \sigma\big(x_k, \ldots, x_m\big)$ for some $k \le m$); see, for example, Corollary 1, in Billingsley (1995, p. 169). Since the measure μ is shift-invariant and the set A is invariant as well, we conclude that for every $n \ge 0$,

$$\varepsilon \ge \mu\big(A \triangle B\big) = \mu\Big(\theta^{-1}\big(A \triangle B\big)\Big)$$
$$= \mu\Big(\theta^{-1}(A) \triangle \theta^{-1}(B)\Big) = \mu\Big(A \triangle \theta^{-1}(B)\Big),$$

and iterating this procedure, we see that

$$\mu\Big(A \triangle \theta^{-n}(B)\Big) \le \varepsilon$$

for all $n = 0, 1, \ldots$. Note that if $B \in \sigma\big(x_k, \ldots, x_m\big)$, then $\theta^{-n}(B) \in \sigma\big(x_{k+n}, \ldots, x_{m+n}\big)$.

We conclude that for every $\varepsilon > 0$, there are sets $B_n \in \sigma(x_n, x_{n+1}, \ldots)$ for $n = 0, 1, 2, \ldots$ such that for every n, $\mu\big(A \triangle B_n\big) \le \varepsilon/2^n$. This implies that for every $m = 0, 1, 2 \ldots$,

$$\mu\Big(A \triangle \cup_{n=m}^{\infty} B_n\Big) \le \sum_{n=m}^{\infty} \mu\big(A \triangle B_n\big) \le \varepsilon/2^{m-1}\,,$$

and then also

$$\mu\Big(A \triangle \cap_{m=0}^{\infty} \cup_{n=m}^{\infty} B_n\Big) \le \sum_{m=0}^{\infty} \mu\big(A \triangle \cup_{n=m}^{\infty} B_n\big) \le 4\varepsilon\,.$$

Therefore, for every $\varepsilon > 0$, there is a set $B \in \mathcal{T}$ such that $\mu\big(A \triangle B\big) \le \varepsilon$, and since $\mathcal{T}$ is a σ-field, there is also $B \in \mathcal{T}$ such that $\mu\big(A \triangle B\big) = 0$. Therefore, A itself is a tail event.

As a corollary, we conclude that every stationary process for which the tail σ-field consists of trivial events (i.e., of the events of probability 0 or 1) must be ergodic, because every invariant event will also have probability 0 or 1. For example, if $\big(X_n,\ n \in \mathbb{Z}\big)$ consists of i.i.d. random variables, then by the Kolmogorov zero–one law, the tail σ-field consists of trivial events, and so every i.i.d. process is ergodic.

If $(X_n,\, n \in \mathbb{Z})$ is a stationary stochastic process defined by (2.2) on the probability space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ with a shift-invariant μ, then Birkhoff's pointwise ergodic theorem applies, of course, in the usual way, and the limiting function g_f in that theorem is, of course, the conditional expectation of f given the invariant σ-field just because m is a probability measure. It is interesting to see how the pointwise ergodic theorem applies if a stationary process $(X_n,\, n \in \mathbb{Z})$ is defined on an abstract probability space $(\Omega, \mathcal{F}, P)$.

We start by defining the invariant σ-sub-field $\mathcal{I}_{\mathbf{X}}$ of $\mathcal{F}$. Let $T_{\mathbf{X}} : \Omega \to \mathbb{R}^{\mathbb{Z}}$ be the measurable map $T_{\mathbf{X}}(w) = (\ldots, X_{-1}(\omega), X_0(\omega), X_1(\omega), \ldots)$. If $\mathcal{I}$ is the invariant σ-field in $\mathbb{R}^{\mathbb{Z}}$, then we set

$$\mathcal{I}_{\mathbf{X}} = T_{\mathbf{X}}^{-1}\mathcal{I} = \left\{T_{\mathbf{X}}^{-1}(A),\, A \in \mathcal{I}\right\} \subset \mathcal{F}\,. \tag{2.5}$$

If μ is the shift-invariant probability measure on $\mathbb{R}^{\mathbb{Z}}$ induced by the stationary process via (2.1), then for every measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ such that

$$E\big|f\big(\ldots, X_{-1}, X_0, X_1, \ldots\big)\big| < \infty, \tag{2.6}$$

we have

$$\frac{1}{n}\sum_{j=0}^{n-1} f\big(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots\big) = \frac{1}{n}\sum_{j=0}^{n-1} f\big(\theta^j(T_{\mathbf{X}}(\omega))\big)$$

$$\to E_\mu\big(f\big|\mathcal{I}\big)(T_{\mathbf{X}}(\omega)) = E\Big(f\big(\ldots, X_{-1}, X_0, X_1, \ldots\big)\Big|\mathcal{I}_{\mathbf{X}}\Big)(\omega) \tag{2.7}$$

with probability 1, where E_μ is the (conditional) expectation with respect to the probability measure μ on $\mathbb{R}^{\mathbb{Z}}$. The convergence follows from Birkhoff's pointwise ergodic theorem, and the last equality follows from the definition of the σ-field $\mathcal{I}_{\mathbf{X}}$.

In particular, a stationary process $(X_n,\, n \in \mathbb{Z})$ is ergodic if and only if the σ-field $\mathcal{I}_{\mathbf{X}}$ defined by (2.5) consists of trivial events. For an ergodic stationary process, one has

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f\big(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots\big) \tag{2.8}$$

$$= Ef\big(\ldots, X_{-1}, X_0, X_1, \ldots\big)$$

with probability 1, for every measurable f satisfying (2.6). Of course, the converse statement is also true: if (2.8) holds for every measurable f satisfying (2.6), then the process is ergodic.

Example 2.1.5. The Strong Law of Large Numbers Suppose that $(X_n,\, n \in \mathbb{Z})$ is an ergodic stationary process with a finite mean. Choosing

$$f\big(\ldots, x_{-1}, x_0, x_1, \ldots\big) = x_0\,,$$

an immediate application of (2.8) proves the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} X_j = EX_0 \tag{2.9}$$

with probability 1. Note, however, that the strong law of large numbers (2.9) may hold for nonergodic stationary processes and hence does not imply ergodicity. For example, let $\big(X_n,\, n \in \mathbb{Z}\big)$ be a canonical stationary process defined by (2.2) on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$, where a shift-invariant μ is the following modification of the measure in (2.3):

$$\mu = \frac{1}{4}\delta_{(\ldots,0,1,0,\ldots)} + \frac{1}{4}\delta_{(\ldots,1,0,1,\ldots)} + \frac{1}{2}\delta_{(\ldots,1/2,1/2,1/2,\ldots)}\,.$$

In this case, the law of large numbers trivially holds (with $EX_0 = 1/2$), but the process is not ergodic, since the event

$$A = \Big\{\text{infinitely many of } x_n \text{ are equal to zero}\Big\}$$

is obviously invariant, and its probability is equal to $1/2$.

The following proposition presents a characterization of ergodicity in the context of stationary stochastic processes that is particularly easy to visualize: a stationary process is ergodic unless it can be represented as a mixture of two stationary processes with different finite-dimensional distributions.

Proposition 2.1.6. *A stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is nonergodic if and only if there is a probability space supporting two stationary processes, $\big(Y_n,\, n \in \mathbb{Z}\big)$ and $\big(Z_n,\, n \in \mathbb{Z}\big)$, with different finite-dimensional distributions, and a Bernoulli(p) random variable with $0 < p < 1$ independent of them such that*

$$\big(X_n,\, n \in \mathbb{Z}\big) \stackrel{d}{=} \begin{cases} \big(Y_n,\, n \in \mathbb{Z}\big) & \text{with probability } p, \\ \big(Z_n,\, n \in \mathbb{Z}\big) & \text{with probability } 1-p. \end{cases} \tag{2.10}$$

Proof. Suppose first that the process $\big(X_n,\, n \in \mathbb{Z}\big)$ is not ergodic, and let μ be the probability measure induced by the process on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ via (2.1). By the lack of ergodicity, there is a set of sequences $A \in \mathcal{I}$ with $p := \mu(A) \in (0,1)$. Define two new probability measures on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ by

$$\mu_1(B) = p^{-1}\mu\big(B \cap A\big), \quad \mu_2(B) = (1-p)^{-1}\mu\big(B \cap A^c\big), \; B \in \mathcal{B}^{\mathbb{Z}}\,.$$

Using first the shift-invariance of A and then the shift-invariance of μ, we see that

$$\begin{aligned} \mu_1\big(\theta^{-1}(B)\big) &= p^{-1}\mu\big(\theta^{-1}(B) \cap A\big) = p^{-1}\mu\big(\theta^{-1}(B) \cap \theta^{-1}(A)\big) \\ &= p^{-1}\mu\big(\theta^{-1}(B \cap A)\big) = p^{-1}\mu\big(B \cap A\big) = \mu_1(B), \end{aligned}$$

and hence μ_1 is shift-invariant. Similarly, the probability measure μ_2 is shift-invariant as well.

Let now $\Omega = \mathbb{R}^{\mathbb{Z}} \cap \{0,1\} \cap \mathbb{R}^{\mathbb{Z}}$, with the product σ-field, and let

$$P = \mu_1 \times \big(p\delta_1 + (1-p)\delta_0\big) \times \mu_2 \,.$$

Using the obvious notation, we define three stochastic processes on this probability space by

$$W_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = \begin{cases} y_n \text{ if } b = 1 \\ z_n \text{ if } b = 0 \end{cases},$$

$$Y_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = y_n \,,$$

$$Z_n\big((\ldots, y_{-1}, y_0, y_1 \ldots),\, b,\, (\ldots, z_{-1}, z_0, z_1 \ldots)\big) = z_n, \quad n \in \mathbb{Z} \,.$$

Since the measures μ_1 and μ_2 are shift-invariant, the processes $\big(Y_n,\, n \in \mathbb{Z}\big)$ and $\big(Z_n,\, n \in \mathbb{Z}\big)$ are stationary. If the two processes had the same finite-dimensional distributions, then the probability measures they generate on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ would coincide. However, these probability measures are μ_1 and μ_2 respectively, and they cannot coincide, since they live on disjoint subsets A and A^c of $\mathbb{R}^{\mathbb{Z}}$. Therefore, the two processes have different finite-dimensional distributions. Finally, let A be a cylindrical subset of $\mathbb{R}^{\mathbb{Z}}$. Then

$$\begin{aligned} P\Big(\big(W_n,\, n \in \mathbb{Z}\big) \in A\Big) &= p\mu_1(A) + (1-p)\mu_2(A) \\ &= \mu(A) = P\Big(\big(X_n,\, n \in \mathbb{Z}\big) \in A\Big) \,, \end{aligned}$$

and we conclude that the relation (2.10) holds.

Conversely, suppose that (2.10) holds. Since the processes $\mathbf{Y} = \big(Y_n,\, n \in \mathbb{Z}\big)$ and $\mathbf{Z} = \big(Z_n,\, n \in \mathbb{Z}\big)$ have different finite-dimensional distributions, there is a bounded measurable function $f : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ such that $Ef(\mathbf{Y}) \neq Ef(\mathbf{Z})$ (one can take $f = \mathbf{1}_A$, with A a cylindrical set to which the laws of $\mathbf{Y} = \big(Y_n,\, n \in \mathbb{Z}\big)$ and $\mathbf{Z} = \big(Z_n,\, n \in \mathbb{Z}\big)$ assign different probabilities). If we call the right-hand side of (2.10) $\big(X_n,\, n \in \mathbb{Z}\big)$ (which is legitimate, since ergodicity depends only on the finite-dimensional distributions of a process), then the ergodic theorem (2.7) tells us that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, X_{j-1}(\omega), X_j(\omega), X_{j+1}(\omega), \ldots\big)$$

$$= \begin{cases} \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, Y_{j-1}(\omega), Y_j(\omega), Y_{j+1}(\omega), \ldots\big) \text{ with probability } p \\ \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\big(\ldots, Z_{j-1}(\omega), Z_j(\omega), Z_{j+1}(\omega), \ldots\big) \text{ with probability } 1-p \end{cases}$$

$$:= \begin{cases} L_1 \text{ with probability } p \\ L_2 \text{ with probability } 1-p \end{cases},$$

where

$$EL_1 = Ef(\mathbf{Y}) \neq Ef(\mathbf{Z}) = EL_2\,.$$

Therefore, L_1 and L_2 cannot be two identical constants, which would be the only possibility if $\big(X_n,\, n \in \mathbb{Z}\big)$ were ergodic. Hence (2.10) implies lack of ergodicity. □

We now switch back to an arbitrary σ-finite measure space $\big(E, \mathcal{E}, m\big)$ and a measurable map $\phi : E \to E$. Let $\big(G, \mathcal{G}\big)$ be another measurable space, and $\varphi : G \to G$ a measurable map. We say that a mapping $f : E \to G$ is compatible with the maps ϕ and φ if $f \circ \phi = \varphi \circ f$, or in other words, if $f(\phi(x)) = \varphi(f(x))$ for all $x \in E$. It turns out that compatible mappings preserve ergodicity, as the following proposition shows.

Proposition 2.1.7. *Let $\phi : E \to E$ be a nonsingular map on a σ-finite measure space $\big(E, \mathcal{E}, m\big)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $\big(G, \mathcal{G}\big)$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is ergodic on $\big(E, \mathcal{E}, m\big)$, then φ is ergodic on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$.*

Proof. The compatibility of f with ϕ and φ implies that for every subset B of G,

$$\phi^{-1}\big(f^{-1}(B)\big) = f^{-1}\big(\varphi^{-1}(B)\big)\,. \tag{2.11}$$

Since ϕ is nonsingular, for every $B \in \mathcal{G}$,

$$\begin{aligned} m \circ f^{-1}(B) = 0 \iff m\big(f^{-1}(B)\big) = 0 \iff m\big(\phi^{-1}\big(f^{-1}(B)\big)\big) = 0 \\ \iff m\big(f^{-1}\big(\varphi^{-1}(B)\big)\big) = 0 \iff m \circ f^{-1}\big(\varphi^{-1}(B)\big) = 0\,, \end{aligned}$$

and so φ is nonsingular on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$. Next, let B be a φ-invariant event. Using (2.11) once again, we see that

$$\begin{aligned} 0 = m \circ f^{-1}\big(B \,\triangle\, \varphi^{-1}(B)\big) &= m\big(f^{-1}(B) \,\triangle\, f^{-1}\big(\varphi^{-1}(B)\big)\big) \\ &= m\big(f^{-1}(B) \,\triangle\, \phi^{-1}\big(f^{-1}(B)\big)\big)\,, \end{aligned}$$

implying that $f^{-1}(B)$ is invariant for ϕ. Since ϕ is ergodic, we conclude that $m\big(f^{-1}(B)\big) = 0$ or 1, which is the same as $m \circ f^{-1}(B) = 0$ or 1, which means that every φ-invariant event is trivially invariant, and hence φ is ergodic. □

Naturally compatible maps are produced by the common transformations of stochastic processes. Let $\big(X_n,\, n \in \mathbb{Z}\big)$ be a stationary stochastic process, and let $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. Then

$$Y_n = g\big((\ldots, X_{n-1}, X_n, X_{n+1}, \ldots)\big),\ n \in \mathbb{Z} \tag{2.12}$$

(with X_n in the zeroth position in the definition of Y_n) is obviously a stationary process as well.

We can view this common situation as a special case of compatible maps as follows. Let $E = G = \mathbb{R}^{\mathbb{Z}}$, $\mathcal{E} = \mathcal{G} = \mathcal{B}^{\mathbb{Z}}$, $m = \mu_X$, the law of the process $\big(X_n,\, n \in \mathbb{Z}\big)$ given by (2.1). Of course, the spaces E and G are identical, and we use the same left shift θ on both. In order to avoid confusion, we will use the notation θ_E when working on the space E, and θ_G when working on the space G.

Define a map $f : E \to G$ by

$$f(\mathbf{x}) = \big(\ldots, g(\theta_E^{-1}\mathbf{x}), g(\mathbf{x}), g(\theta_E(\mathbf{x})), \ldots\big) \in G$$

for $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in E$. Clearly, f is measurable. It is also trivially compatible with the shifts θ_E and θ_G. Note also that the probability measure $m \circ f^{-1}$ is simply the law of the process $\big(Y_n,\, n \in \mathbb{Z}\big)$ given by (2.1).

Therefore, Proposition 2.1.7 applies, and we have proved the following statement.

Corollary 2.1.8. *Let $\big(X_n,\, n \in \mathbb{Z}\big)$ be an ergodic stationary stochastic process, and let $\big(Y_n,\, n \in \mathbb{Z}\big)$ be a stationary process given by (2.12) for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $\big(Y_n,\, n \in \mathbb{Z}\big)$ is ergodic as well.*

Example 2.1.9. Moving average processes are ergodic. Let

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_j\, \varepsilon_{n-j}\,, \quad n \in \mathbb{Z}$$

be an infinite moving average process (1.23). Recall that the noise variables $(\varepsilon_n,\, n = \ldots, -1, 0, 1, 2, \ldots)$ are i.i.d. and (φ_n) are deterministic coefficients. We assume that the series defining the process converges with probability 1.

Define a function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ by

$$g\big((\ldots, x_{-1}.x_0, x_1, \ldots)\big) = \begin{cases} \sum_{j=-\infty}^{\infty} \varphi_j\, x_{-j} & \text{if the sum converges,} \\ 0 & \text{if the sum diverges.} \end{cases}$$

Clearly, g is a measurable function, and

$$X_n = g\big((\ldots, \varepsilon_{n-1}.\varepsilon_n, \varepsilon_{n+1}, \ldots)\big),\ n \in \mathbb{Z}\,.$$

Since the i.i.d. process $(\varepsilon_n,\, n \in \mathbb{Z})$ is ergodic (see Example 2.1.4), it follows from Corollary 2.1.8 that every infinite moving average process is ergodic.

2.2 Mixing and Weak Mixing

We start by introducing another basic ergodic-theoretical notion, that of *mixing*. It applies to measure-preserving maps on probability spaces.

Definition 2.2.1. A nonsingular measure-preserving map ϕ on a probability space $\big(E, \mathcal{E}, m\big)$ is called mixing if for every two sets $A, B \in \mathcal{E}$,

$$\lim_{n\to\infty} m\big(A \cap \phi^{-n}B\big) = m(A)\, m(B)\,.$$

Here ϕ^{-n} is the nth power of the inverse operator ϕ^{-1}. An immediate observation is that mixing is a stronger property than ergodicity. Indeed, let ϕ be a mixing map on a probability space $\big(E, \mathcal{E}, m\big)$, and suppose that ϕ is not ergodic. In that case, there is an invariant set $C \in \mathcal{E}$ with $0 < m(C) < 1$. Taking $A = B = C$, we have

$$\begin{aligned} m\big(A \cap \phi^{-n}B\big) &= m\big(C \cap \phi^{-n}C\big) = m(C) \\ &\neq \big(m(C)\big)^2 = m(A)\, m(B)\,, \end{aligned}$$

contradicting the assumed mixing. Therefore, mixing implies ergodicity. On the other hand, a map can be ergodic without being mixing.

Example 2.2.2. An ergodic but not mixing map.

Consider the left shift θ on the sequence space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$, where μ is the two-point shift-invariant probability measure (2.3). Let $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ be the two points of the support of μ. Note that $\theta\mathbf{x}^{(1)} = \mathbf{x}^{(2)}$ and $\theta\mathbf{x}^{(2)} = \mathbf{x}^{(1)}$. Therefore, every invariant set A contains either both of these points or none, so that $\mu(A) = 1$ or 0. Since all invariant sets are trivial, θ is ergodic.

On the other hand, let

$$A = B = \Big\{(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) : x_0 = 1\Big\}.$$

Then

$$\mu\big(A \cap \theta^{-n}B\big) = \begin{cases} \frac{1}{2} \text{ if } n \text{ is even,} \\ 0 \text{ if } n \text{ is odd} \end{cases} \not\to \frac{1}{4} = \mu(A)\mu(B)\,.$$

Therefore, θ is not mixing.

Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary stochastic process. We assume first that the process is defined by (2.2) on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$ with a shift-invariant μ. We say that the process is mixing if the left shift θ is a mixing map on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$. As in the case of ergodicity, whether or not a stationary process is mixing is determined by its finite-dimensional distributions, regardless of what probability space the process is really defined on. Explicitly, a stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$ is mixing if for every two cylindrical subsets A, B of $\mathbb{R}^{\mathbb{Z}}$,

$$\begin{aligned} &P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in A,\ \big(\ldots, X_{n-1}, X_n, X_{n+1}, \ldots\big) \in B\Big) \qquad (2.13) \\ &\to P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in A\Big) P\Big(\big(\ldots, X_{-1}, X_0, X_1, \ldots\big) \in B\Big) \end{aligned}$$

as $n \to \infty$.

Example 2.2.3. An i.i.d. process is mixing

We showed at the end of Example 2.1.4 that an i.i.d. process $(X_n,\, n \in \mathbb{Z})$ is ergodic. Now we will check that such a process is, in fact, mixing. This fact will also follow from Theorem 2.2.7 below, but a direct argument is instructive. We may assume that the process is defined on the sequence space $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu)$ by (2.2).

Take any sets $A, B \in \mathcal{B}^{\mathbb{Z}}$. As in Example 2.1.4, given $\varepsilon > 0$, we may choose finite-dimensional cylindrical sets A_1 and $B_1 \in \sigma(x_k, \ldots, x_m)$ for some $k \le m$ such that

$$\mu(A \triangle A_1) \le \varepsilon, \quad \mu(B \triangle B_1) \le \varepsilon\,.$$

Note that $\theta^{-n}(B_1) \in \sigma(x_{k+n}, \ldots, x_{m+n})$, so that for $n > m - k$, the sets A_1 and $\theta^{-n}(B_1)$ are generated by disjoint sets of coordinates of a point in the sequence space. Since the measure μ is the law of an i.i.d. sequence, different components are independent under μ, which means that A_1 and $\theta^{-n}(B_1)$ are independent events under μ when $n > m - k$. Therefore, for such n,

$$\begin{aligned}
\left|\mu(A \cap \theta^{-n}B) - \mu(A)\,\mu(B)\right| &\le \left|\mu(A \cap \theta^{-n}B) - \mu(A_1 \cap \theta^{-n}B_1)\right| \\
&+ \left|\mu(A_1 \cap \theta^{-n}B_1) - \mu(A_1)\,\mu(B_1)\right| + \left|\mu(A_1)\,\mu(B_1) - \mu(A)\,\mu(B)\right| \\
&\le 2\mu(A \triangle A_1) + 2\mu(B \triangle B_1) \le 4\varepsilon.
\end{aligned}$$

That is, for every $\varepsilon > 0$,

$$\limsup_{n\to\infty} \left|\mu(A \cap \theta^{-n}B) - \mu(A)\,\mu(B)\right| \le 4\varepsilon,$$

and letting $\varepsilon \to 0$, we conclude that

$$\mu(A \cap \theta^{-n}B) \to \mu(A)\,\mu(B)$$

as $n \to \infty$ for all $A, B \in \mathcal{B}^{\mathbb{Z}}$. Therefore, the shift θ is mixing, and hence so is the i.i.d. process $(X_n,\, n \in \mathbb{Z})$.

Proposition 2.1.7 has a counterpart describing preservation of mixing.

Proposition 2.2.4. *Let $\phi : E \to E$ be a nonsingular measure-preserving map on a probability space $(E, \mathcal{E}, m)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $(G, \mathcal{G})$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is mixing on $(E, \mathcal{E}, m)$, then φ is mixing on $(G, \mathcal{G}, m \circ f^{-1})$.*

Proof. Starting with (2.11), an inductive argument shows that the latter relation extends to

$$\phi^{-n}(f^{-1}(B)) = f^{-1}(\varphi^{-n}(B))$$

for all $n \geq 1$. Therefore, for all sets $A, B \in \mathcal{G}$,

$$m \circ f^{-1}\big(A \cap \varphi^{-n}(B)\big) = m\big(f^{-1}(A) \cap f^{-1}(\varphi^{-n}(B))\big)$$
$$= m\big(f^{-1}(A) \cap \phi^{-n}(f^{-1}(B))\big) \to m\big(f^{-1}(A)\big)\, m\big(f^{-1}(B)\big)$$
$$= m \circ f^{-1}(A)\, m \circ f^{-1}(B)\,,$$

and so φ is mixing on $\big(G, \mathcal{G}, m \circ f^{-1}\big)$. □

As in the case of ergodicity, we immediately obtain the following corollary to Proposition 2.2.4.

Corollary 2.2.5. *Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a mixing stationary stochastic process, and let $\big(Y_n,\ n \in \mathbb{Z}\big)$ be a stationary process given by (2.12) for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $\big(Y_n,\ n \in \mathbb{Z}\big)$ is mixing as well.*

Example 2.2.6. Moving average processes are mixing.

Let $(X_n,\ n \in \mathbb{Z})$ be the infinite moving process of Example 2.1.9. Applying Corollary 2.2.5, we see that the mixing property of the noise variables (ε_n) says that every infinite moving process is mixing as well.

It turns out that the mixing property of a stationary stochastic process is equivalent to weak convergence to independence of the joint distributions of the blocks of observations of the process separated by increasing time intervals.

Theorem 2.2.7. *A stationary process $\big(X_n,\ n \in \mathbb{Z}\big)$ is mixing if and only if for every $k = 1, 2, \ldots,$*

$$\big(X_1, \ldots, X_k, X_{n+1}, \ldots, X_{n+k}\big) \Rightarrow \big(X_1, \ldots, X_k, Y_1, \ldots, Y_k\big) \tag{2.14}$$

as $n \to \infty$, where $(Y_1, \ldots, Y_k)$ is an independent copy of $(X_1, \ldots, X_k)$.

Proof. We start with checking the easier implication, namely the necessity of condition (2.14). Suppose that the process $\big(X_n,\ n \in \mathbb{Z}\big)$ is mixing, and let $k \geq 1$. The weak convergence in (2.14) will follow once we check that for all k-dimensional Borel sets C and D, we have

$$P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big) \tag{2.15}$$
$$\to P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big)$$

as $n \to \infty$ (of course, it is really necessary to check (2.15) for continuity sets of the law of $(X_1, \ldots, X_k)$). This statement, however, is a special case of the statement (2.13).

Suppose now that (2.14) holds. We will prove that the process is mixing by checking the condition in the definition for successively more general sets A and B. We treat finite-dimensional cylindrical sets first. Fix a dimension $k \geq 1$. We prove the statement (2.15). The challenge is, of course, in the fact that by default, weak convergence tells us that this statement holds only for continuity sets C and D, and we need to establish (2.15) for all k-dimensional Borel sets.

Consider first the case in which the sets C and D are "southwest corners" of the form

$$C=\Big\{(x_1,x_2,\ldots,x_k):\ x_i\le a_i \text{ for } i\in I_1,\ \ x_i<a_i \text{ for } i\in\{1,\ldots,k\}\setminus I_1\Big\}, \tag{2.16}$$

$$B=\Big\{(x_1,x_2,\ldots,x_k):\ x_i\le b_i \text{ for } i\in I_2,\ \ x_i<b_i \text{ for } i\in\{1,\ldots,k\}\setminus I_2\Big\},$$

where I_1, I_2 are subsets of $\{1,\ldots,k\}$, and $a_1,\ldots,a_k,b_1,\ldots,b_k$ are real numbers. Note that if all of these $2k$ numbers are continuity points of the distribution of X_1, then the sets C and D are continuity sets of the vector $(X_1,\ldots,X_k,Y_1,\ldots,Y_k)$ on the right-hand side of (2.14), so (2.15) follows from the weak convergence in this case. This case forms the basis of an inductive argument. Specifically, let $m=0,1,\ldots,2k$ be the number of points among $a_1,\ldots,a_k,b_1,\ldots,b_k$ that are not continuity points of the distribution of X_1. We have checked that (2.15) holds if $m=0$, and the induction hypothesis is that (2.15) holds if $m<m_0$ for some $m_0=1,,\ldots,2k$. Suppose now that $m=m_0$, and choose one point out of the m_0 discontinuity points. We suppose that this point is of the type b_i with $i\in I_2$; all other cases are similar. For ease of notation, we will simply use b_1, with $1\in I_2$.

Choose a sequence $\varepsilon_j\downarrow 0$ such that for each j, $b_1+\varepsilon_j$ is a continuity point of the distribution of X_1. By the induction hypothesis,

$$\begin{aligned}
&\limsup_{n\to\infty} P\big((X_1,\ldots,X_k)\in C,\ (X_{n+1},\ldots,X_{n+k})\in D\big)\\
&\le \limsup_{n\to\infty} P\big((X_1,\ldots,X_k)\in C,\ X_{n+1}\le b_1+\varepsilon_j,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\ldots,k\}\setminus I_2\big)\\
&= P\big((X_1,\ldots,X_k)\in C\big)P\big(X_1\le b_1+\varepsilon_j,\ X_i\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_i<b_i \text{ for } i\in\{1,\ldots,k\}\setminus I_2\big)
\end{aligned}$$

for every $j=1,2,\ldots$, and letting $j\to\infty$, we obtain the upper bound

$$\begin{aligned}
&\limsup_{n\to\infty} P\big((X_1,\ldots,X_k)\in C,\ (X_{n+1},\ldots,X_{n+k})\in D\big) \qquad (2.17)\\
&\qquad\le P\big((X_1,\ldots,X_k)\in C\big)\,P\big((X_1,\ldots,X_k)\in D\big)\,.
\end{aligned}$$

In order to obtain a matching lower bound, we write

$$\begin{aligned}
&P\big((X_1,\ldots,X_k)\in C,\ (X_{n+1},\ldots,X_{n+k})\in D\big)\\
&= P\big((X_1,\ldots,X_k)\in C,\ X_{n+1}<b_1,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\ldots,k\}\setminus I_2\big)\\
&+P\big((X_1,\ldots,X_k)\in C,\ X_{n+1}=b_1,\ X_{n+i}\le b_i \text{ for } i\in I_2\setminus\{1\},\\
&\qquad X_{n+i}<b_i \text{ for } i\in\{1,\ldots,k\}\setminus I_2\big):=p_1(n)+p_2(n)\,.
\end{aligned}$$

An argument identical to the one used to obtain the upper bound above (replacing the condition $X_{n+1} < b_1$ by the condition $X_{n+} < b_1 - \varepsilon_j$, where $\varepsilon_j \downarrow 0$ and $b_1 - \varepsilon_j$ is, for each j, a continuity point of the distribution of X_1) shows that

$$\liminf_{n\to\infty} p_1(n) \geq P\big((X_1,\ldots,X_k)\in C\big)P\big(X_1 < b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\ldots,k\} \setminus I_2\big)\,.$$

Furthermore, given $\varepsilon > 0$, we can find $\delta > 0$ such that both $P(b_1 - \delta < X_1 < b_1) \leq \varepsilon$ and $P(b_1 < X_1 < b_1 + \delta) \leq \varepsilon$, and both $b_1 - \delta$ and $b_1 + \delta$ are continuity points of the distribution of X_1. Then

$$p_2(n) \geq P\big((X_1,\ldots,X_k)\in C,\, b_1 - \delta < X_{n+1} < b_1 + \delta,\, X_{n+i} \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_{n+i} < b_i \text{ for } i \in \{1,\ldots,k\} \setminus I_2\big) - 2\varepsilon\,,$$

and by the induction hypothesis, we obtain

$$\begin{aligned}\liminf_{n\to\infty} p_2(n) &\geq P\big((X_1,\ldots,X_k)\in C\big)P\big(b_1-\delta < X_1 < b_1+\delta,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},\\ &\qquad X_i < b_i \text{ for } i \in \{1,\ldots,k\} \setminus I_2\big) - 2\varepsilon\\ &\geq P\big((X_1,\ldots,X_k)\in C\big)P\big(X_1 = b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},\\ &\qquad X_i < b_i \text{ for } i \in \{1,\ldots,k\} \setminus I_2\big) - 2\varepsilon\,.\end{aligned}$$

Letting $\varepsilon \to 0$ we see that

$$\liminf_{n\to\infty} p_2(n) \geq P\big((X_1,\ldots,X_k)\in C\big)P\big(X_1 = b_1,\, X_i \leq b_i \text{ for } i \in I_2 \setminus \{1\},$$
$$X_i < b_i \text{ for } i \in \{1,\ldots,k\} \setminus I_2\big)\,,$$

which establishes a lower bound matching (2.17) and hence completes the inductive argument. Therefore, we have proved that (2.15) holds when the sets C and D are "southwest corners" of the form (2.16).

The next step is to show that (2.15) holds when C and D are "rectangles" of the form

$$C = \Big\{(x_1,x_2,\ldots,x_k) :\, a_i^{(1)} \leq x_i \leq a_i^{(2)} \text{ for } i \in I_{11},\;\; a_i^{(1)} \leq x_i < a_i^{(2)} \text{ for } i \in I_{12},$$
$$a_i^{(1)} < x_i \leq a_i^{(2)} \text{ for } i \in I_{13},\;\; a_i^{(1)} < x_i < a_i^{(2)} \text{ for } i \in I_{14}\Big\}, \tag{2.18}$$

$$B = \Big\{(x_1,x_2,\ldots,x_k) :\, b_i^{(1)} \leq x_i \leq b_i^{(2)} \text{ for } i \in I_{21},\;\; b_i^{(1)} \leq x_i < b_i^{(2)} \text{ for } i \in I_{22},$$
$$b_i^{(1)} < x_i \leq b_i^{(2)} \text{ for } i \in I_{23},\;\; b_i^{(1)} < x_i < b_i^{(2)} \text{ for } i \in I_{24}\Big\}\,,$$

where both $(I_{11}, I_{12}, I_{13}, I_{14})$ and $(I_{21}, I_{22}, I_{23}, I_{24})$ are partitions of $\{1, \ldots, k\}$, and $-\infty \le a_i^{(1)} \le a_i^{(2)} \le \infty$, $i = 1, \ldots, k$, and $b_i^{(1)} \le b_i^{(2)}$, $i = 1, \ldots, k$, are real numbers. For notational ease, we will consider only the case $I_{13} = I_{23} = \{1, \ldots, k\}$, but the other cases are similar.

It is easy to check that

$$P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big) \tag{2.19}$$

$$= \sum_{J_1, J_2} (-1)^{|J_1|+|J_2|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1,$$

$$X_{n+i} \le b_i^{(1)} \text{ for } i \in J_2,\ X_{n+i} \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big)\,,$$

where the sum is taken over all subsets J_1 and J_2 of $\{1, \ldots, k\}$. The right-hand side of (2.19) is a finite sum of probabilities that we have already considered when we proved (2.15) for the "southwest corners." Therefore,

$$\lim_{n\to\infty} P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big)$$

$$= \sum_{J_1, J_2} (-1)^{|J_1|+|J_2|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1\big)$$

$$P\big(X_i \le b_i^{(1)} \text{ for } i \in J_2,\ X_i \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big)$$

$$= \sum_{J_1} (-1)^{|J_1|} P\big(X_i \le a_i^{(1)} \text{ for } i \in J_1,\ X_i \le a_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_1\big)$$

$$\sum_{J_2} (-1)^{|J_2|} P\big(X_i \le b_i^{(1)} \text{ for } i \in J_2,\ X_i \le b_i^{(2)} \text{ for } i \in \{1, \ldots, k\} \setminus J_2\big)$$

$$= P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big)\,,$$

where at the last step we used (2.19) once again. This proves (2.15) in the case that C and D are "rectangles" of the form (2.18).

Next, denote by $\mathcal{U}_k$ the collection of all disjoint finite unions of "rectangles" of the form (2.18). Note that $\mathcal{U}_k$ forms a field in $\mathbb{R}^k$ that generates the Borel σ-field. Since (2.15) holds for the "rectangles," it extends by linearity to the case that C and D are sets in $\mathcal{U}_k$. Furthermore, if C and D are arbitrary k-dimensional Borel sets, then given $\varepsilon > 0$, we can find sets C_1 and D_1 in $\mathcal{U}_k$ such that

$$P\big((X_1, \ldots, X_k) \in C \triangle C_1\big) \le \varepsilon,\ P\big((X_1, \ldots, X_k) \in D \triangle D_1\big) \le \varepsilon\,;$$

see again Corollary 1, p. 169, in Billingsley (1995). Then

$$\Big| P\big((X_1, \ldots, X_k) \in C,\ (X_{n+1}, \ldots, X_{n+k}) \in D\big)$$

$$- P\big((X_1, \ldots, X_k) \in C\big)\, P\big((X_1, \ldots, X_k) \in D\big) \Big|$$

$$
\begin{aligned}
&\le \Big| P\big((X_1,\dots,X_k)\in C_1,\ (X_{n+1},\dots,X_{n+k})\in D_1\big) \\
&\quad - P\big((X_1,\dots,X_k)\in C_1\big)\,P\big((X_1,\dots,X_k)\in D_1\big)\Big| \\
&\quad + \Big| P\big((X_1,\dots,X_k)\in C,\ (X_{n+1},\dots,X_{n+k})\in D\big) \\
&\quad - P\big((X_1,\dots,X_k)\in C_1,\ (X_{n+1},\dots,X_{n+k})\in D_1\big)\Big| \\
&\quad + \Big| P\big((X_1,\dots,X_k)\in C_1\big)\,P\big((X_1,\dots,X_k)\in D_1\big) \\
&\quad - P\big((X_1,\dots,X_k)\in C\big)\,P\big((X_1,\dots,X_k)\in D\big)\Big| \\
&\le \Big| P\big((X_1,\dots,X_k)\in C_1,\ (X_{n+1},\dots,X_{n+k})\in D_1\big) \\
&\quad - P\big((X_1,\dots,X_k)\in C_1\big)\,P\big((X_1,\dots,X_k)\in D_1\big)\Big| + 4\varepsilon\,.
\end{aligned}
$$

Since we have proved that (2.15) holds for sets in $\mathcal{U}_k$, we conclude that

$$
\begin{aligned}
\limsup_{n\to\infty} &\Big| P\big((X_1,\dots,X_k)\in C,\ (X_{n+1},\dots,X_{n+k})\in D\big) \\
&- P\big((X_1,\dots,X_k)\in C\big)\,P\big((X_1,\dots,X_k)\in D\big)\Big| \le 4\varepsilon
\end{aligned}
$$

for every $\varepsilon > 0$, and letting $\varepsilon \to 0$ shows that (2.15) holds for arbitrary k-dimensional Borel sets.

Now that (2.15) has been established in its full generality, mixing of the process $\big(X_n,\ n\in\mathbb{Z}\big)$ follows because the statement (2.13) holds. To show that this is true, one approximates arbitrary sets in $\mathcal{B}^{\mathbb{Z}}$ by finite-dimensional cylindrical sets, as in Example 2.2.3. □

Example 2.2.8. An immediate conclusion of Theorem 2.2.7 is that a stationary Gaussian process $\big(X_n,\ n\in\mathbb{Z}\big)$ is mixing if and only if its correlation function asymptotically vanishes: $\rho_n := \mathrm{Corr}(X_{j+n}, X_j) \to 0$ as $n\to\infty$.

Example 2.2.9. We now take a second look at the tail σ-field $\mathcal{T}$ of a stationary stochastic process. We saw in Example 2.1.4 that if this σ-field is trivial, then the process is ergodic. We will show now that a trivial tail σ-field implies mixing as well. This statement is an immediate corollary of the following characterization: the tail σ-field $\mathcal{T}$ of a stationary process is trivial if and only if for every set A in $\mathcal{B}^{\mathbb{Z}}$,

$$
\lim_{n\to\infty} \sup_{B\in\sigma(x_0,x_1,\dots)} \big|\mu\big(A\cap\theta^{-n}B\big) - \mu(A)\mu(B)\big| = 0\,, \tag{2.20}
$$

where μ is the law of the stationary process on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}})$. This statement makes it possible to view the triviality of the tail σ-field as a kind of uniform mixing and hence akin to the strong mixing properties of Section 2.3.

The sufficiency of (2.20) for triviality of the tail σ-field is clear: if the tail σ-field is not trivial, then there is a tail event $A \in \cap_{n=1}^{\infty}\sigma\big(x_n, x_{n+1}, \ldots\big)$ satisfying $0 < \mu(A) < 1$. This means that for every $n = 1, 2, \ldots$, the set $\theta^n A$ is in $\sigma(x_0, x_1, \ldots)$, so that

$$\sup_{B\in\sigma(x_0,x_1,\ldots)} \big|\mu\big(A \cap \theta^{-n}B\big) - \mu(A)\mu(B)\big| \geq \big|\mu\big(A \cap \theta^{-n}(\theta^n A)\big) - \mu(A)\mu(\theta^n A)\big|$$
$$= \mu(A) - (\mu(A))^2 > 0\,,$$

contradicting (2.20). On the other hand, suppose that the tail σ-field is trivial. For events A and B as in (2.20), define random variables $X = \mathbf{1}_A - \mu(A)$ and $Y_B = \mathbf{1}_{\theta^{-n}B} - \mu(B)$, and note that Y is measurable with respect to the σ-field $\sigma\big(x_n, x_{n+1}, \ldots\big)$. Therefore,

$$\begin{aligned}
&\sup_{B\in\sigma(x_0,x_1,\ldots)} \big|\mu\big(A \cap \theta^{-n}B\big) - \mu(A)\mu(B)\big| \\
&= \sup_{B\in\sigma(x_0,x_1,\ldots)} \big|E(XY_B)\big| = \sup_{B\in\sigma(x_0,x_1,\ldots)} \Big|E\Big(Y_B E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)\Big| \\
&\leq \Big[E\Big(E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)^2\Big]^{1/2} \sup_{B\in\sigma(x_0,x_1,\ldots)} (EY_B^2)^{1/2} \\
&\leq \Big[E\Big(E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big)\Big)^2\Big]^{1/2}.
\end{aligned}$$

Since the σ-fields $\sigma\big(x_n, x_{n+1}, \ldots\big)$ decrease to the trivial σ-field $\mathcal{T}$, it follows that with probability 1,

$$E\big(X\big|\sigma(x_n, x_{n+1}, \ldots)\big) \to E\big(X\big|\mathcal{T}\big) = EX = 0\,;$$

see, e.g., Theorem 35.9 in Billingsley (1995). Since the random variable $|X|$ is, furthermore, bounded by 1, the convergence in (2.20) follows.

Let us consider once again a nonsingular measure-preserving map ϕ on a probability space $\big(E, \mathcal{E}, m\big)$. Notice that the two ergodic-theoretical properties of ϕ we have already discussed, ergodicity and mixing, can be stated as

$$\phi \text{ is } \begin{cases} \text{mixing if } m\big(A \cap \phi^{-n}B\big) - m(A)\,m(B) \to 0,\ A, B \in \mathcal{E}, \\ \text{ergodic if } \frac{1}{n}\sum_{j=0}^{n-1}\big(m\big(A \cap \phi^{-j}B\big) - m(A)\,m(B)\big) \to 0\, A, B \in \mathcal{E}; \end{cases} \tag{2.21}$$

in fact, the first line in (2.21) is just the definition of mixing. Let us check the second line in (2.21). Sufficiency of the condition presented there is clear: if ϕ were not ergodic, then choosing $A = B$ to be an invariant set whose measure takes a value in $(0, 1)$ would provide a counterexample to the condition in (2.21). Let us check the necessity of this condition. Suppose that ϕ is ergodic. By the pointwise ergodic theorem,

$$\frac{1}{n}\sum_{j=0}^{n-1} \mathbf{1}_B(\phi^j x) \to m(B)$$

as $n \to \infty$ for m-almost every $x \in E$. By the bounded convergence theorem, the integral over the set A of the left-hand side above converges to the integral of the right-hand side over the same set, and this gives the condition in (2.21).

Since the usual convergence of a sequence implies its Cesaro convergence, (2.21) provides yet another explanation of the fact that mixing is a stronger property than ergodicity. Furthermore, (2.21) makes it clear that there is an intermediate property of probability measure-preserving maps, weaker than mixing but stronger than ergodicity. We introduce this notion in the following definition.

Definition 2.2.10. A nonsingular measure-preserving map ϕ on a probability space $\big(E, \mathcal{E}, m\big)$ is called weakly mixing if for every two sets $A, B \in \mathcal{E}$,

$$\frac{1}{n}\sum_{j=0}^{n-1} \big|m\big(A \cap \phi^{-j}B\big) - m(A)\, m(B)\big| \to 0\,.$$

If $\big(X_n,\, n \in \mathbb{Z}\big)$ is a stationary stochastic process, then we say that it is weakly mixing if the left shift θ is a weakly mixing map on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$, where μ is the probability measure generated by the process on $\mathbb{R}^{\mathbb{Z}}$.

By definition,

$$\text{mixing} \;\Rightarrow\; \text{weak mixing} \;\Rightarrow\; \text{ergodicity}\,,$$

either for stationary stochastic processes or for measure-preserving maps. However, neither of these two implications can, in general, be reversed. It is easy to construct an example of an ergodic map that is not weakly mixing.

Example 2.2.11. An ergodic but not weakly mixing map.

Consider the two-point left shift-invariant probability measure (2.3) considered in Example 2.2.2. The left shift θ is ergodic on that probability space, but is not mixing. It is not weakly mixing either, since for A and B as in Example 2.2.2, we have

$$\big|m\big(A \cap \theta^{-n}B\big) - m(A)\, m(B)\big| = \frac{1}{4}$$

for each n, and the Cesaro limit of these numbers is equal to $1/4$, not to 0.

We will see in the sequel examples of weakly mixing stationary processes that fail to be mixing.

An alternative point of view on weak mixing is based on the notion of convergence in density. Recall that a subset K of positive integers $\mathbb{N}$ is said to have density zero if

$$\lim_{n\to\infty} \frac{\big|K \cap \{1, 2, \ldots, n\}\big|}{n} = 0\,.$$

A sequence (b_n) in a metric space with a metric d is said to converge in density to b if there is a set K of density zero such that

$$\lim_{n\to\infty,\, n\notin K} b_n = b\,.$$

Explicitly, for every $\varepsilon > 0$, there is $N = N(\varepsilon)$ such that for every $n > N$, $n \notin K$, $d(b_n, b) \le \varepsilon$. The following useful lemma is even more explicit.

Lemma 2.2.12. *A sequence (b_n) converges in density to b if and only if for every $\varepsilon > 0$, the set $K_\varepsilon = \{n : d(b_n, b) > \varepsilon\}$ has density zero.*

Proof. The necessity in the statement is clear, so let us check the sufficiency part. Assume that for every $\varepsilon > 0$, the set K_ε has density zero. The task is to construct a set of density zero away from which the sequence (b_n) converges to b.

We define an increasing sequence of nonnegative integers as follows. Let $N_0 = 0$, and for $k \ge 1$, let

$$N_k = \sup\Big\{n \ge N_{k-1} + 1 : \Big|\{j = 1, \dots, n : d(b_j, b) > \frac{1}{k}\}\Big| > \frac{n}{2^k}\Big\}\,.$$

By the assumption, each N_k is a finite number. Let

$$K = \bigcup_{k=1}^{\infty} \Big\{j : j = N_k + 1, \dots, N_{k+1},\, d(b_j, b) > \frac{1}{k}\Big\}\,.$$

By the definition, if $n \notin K$ and $n > N_k$, then $d(b_n, b) \le \frac{1}{k}$, for $k = 1, 2, \dots$. Therefore, $\lim_{n\to\infty,\, n\notin K} b_n = b$. To check that K has density zero, let k_i be the largest k such that $N_k \le i$, and notice that $k_i \to \infty$ as $i \to \infty$. Choose $m \ge 1$ and let n be so large that $k_n > m$. Then

$$\begin{aligned}
\Big|\big(K \cap \{1, \dots, n\}\big)\Big| &\le N_m + \sum_{k=m}^{k_n-1} \Big|\Big(\Big\{j : j = N_k + 1, \dots, N_{k+1},\, d(b_j, b) > \frac{1}{k}\Big\}\Big)\Big| \\
&\quad + \Big|\Big(\Big\{j : j = N_{k_n} + 1, \dots, n,\, d(b_j, b) > \frac{1}{k_n}\Big\}\Big)\Big| \\
&\le N_m + \sum_{k=m}^{k_n-1} \Big|\Big(\Big\{j : j = 1, \dots, N_{k+1},\, d(b_j, b) > \frac{1}{k}\Big\}\Big)\Big| \\
&\quad + \Big|\Big(\Big\{j : j = 1, \dots, n,\, d(b_j, b) > \frac{1}{k_n}\Big\}\Big)\Big| \\
&\le N_m + \sum_{k=m}^{k_n-1} \frac{N_{k+1}}{2^k} + \frac{n}{2^{k_n}} \le N_m + \sum_{k=m}^{k_n} \frac{n}{2^k} \le N_m + \frac{n}{2^{m-1}}\,.
\end{aligned}$$

Therefore,

$$\limsup_{n\to\infty} \frac{\left|K \cap \{1, 2, \ldots, n\}\right|}{n} \leq \frac{1}{2^{m-1}},$$

and letting $m \to \infty$ shows that K has density zero. □

The following proposition is a simple consequence.

Proposition 2.2.13. *A nonsingular measure-preserving map ϕ on a probability space $(E, \mathcal{E}, m)$ is weakly mixing if and only if for every two sets $A, B \in \mathcal{E}$,*

$$m\left(A \cap \phi^{-n} B\right) \to m(A)\, m(B) \quad \textit{in density.}$$

Proof. The sufficiency for weak mixing of the condition in the proposition is clear. Let us check the necessity. Suppose that the condition in the proposition fails. By Lemma 2.2.12, there is $\varepsilon > 0$ such that the set

$$K_\varepsilon = \left\{n : \left|m\left(A \cap \phi^{-j} B\right) - m(A)\, m(B)\right| > \varepsilon\right\}$$

does not have density zero. Since

$$\frac{1}{n} \sum_{j=0}^{n-1} \left|m\left(A \cap \phi^{-j} B\right) - m(A)\, m(B)\right|$$

$$\geq \frac{1}{n} \sum_{j \in K_\varepsilon \cap \{0,\ldots,n-1\}} \left|m\left(A \cap \phi^{-j} B\right) - m(A)\, m(B)\right| \geq \varepsilon \frac{\left|K_\varepsilon \cap \{1, 2, \ldots, n\}\right|}{n},$$

which does not converge to zero, ϕ cannot be weakly mixing, and the proof is complete. □

Proposition 2.2.4 has an immediate counterpart for weak mixing. The argument is the same; just use Proposition 2.2.13 and replace the usual convergence by convergence away from a set of density zero.

Proposition 2.2.14. *Let $\phi : E \to E$ be a nonsingular measure-preserving map on a probability space $(E, \mathcal{E}, m)$. Let $\varphi : G \to G$ be a one-to-one and onto map on a measurable space $(G, \mathcal{G})$ such that both φ and its inverse are measurable. Let $f : E \to G$ be a measurable map that is compatible with ϕ and φ. If ϕ is weakly mixing on $(E, \mathcal{E}, m)$, then φ is weakly mixing on $(G, \mathcal{G}, m \circ f^{-1})$.*

As in the case of ergodicity and mixing, the following corollary is immediate.

Corollary 2.2.15. *Let $(X_n, n \in \mathbb{Z})$ be a weakly mixing stationary stochastic process, and let $(Y_n, n \in \mathbb{Z})$ be a stationary process given by (2.12) for some measurable function $g : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$. Then $(Y_n, n \in \mathbb{Z})$ is weakly mixing as well.*

Many of the important stochastic processes are constructed as sums of independent, and more elementary, stochastic processes. Therefore, it would be nice to be able to obtain ergodicity of a sum based on the ergodicity of the summands. Unfortunately, this is impossible to do in general, since the sum of two independent stationary ergodic processes does not have to be ergodic; see Exercise 2.6.7. It turns out, however, that strengthening the assumption on one of the two summands from ergodicity to weak mixing is sufficient to guarantee the ergodicity of the sum.

Theorem 2.2.16. *Let* $\big(X_n,\, n \in \mathbb{Z}\big)$ *and* $\big(Y_n,\, n \in \mathbb{Z}\big)$ *be independent stationary stochastic processes. Assume that* $\big(X_n,\, n \in \mathbb{Z}\big)$ *is ergodic, and that* $\big(Y_n,\, n \in \mathbb{Z}\big)$ *is weakly mixing. Then the process* $Z_n = X_n + Y_n$, $n \in \mathbb{Z}$ *is ergodic.*

Proof. Let μ be the probability measure generated by the process $\big(X_n,\, n \in \mathbb{Z}\big)$ on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}\big)$, and let ν be the measure generated by the process $\big(Y_n,\, n \in \mathbb{Z}\big)$. Consider the product probability space, $\big(\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}, \mu \times \nu\big)$. The left shift θ on $\mathbb{R}^{\mathbb{Z}}$ extends naturally to the product shift $\theta \times \theta$ operating on the product space: $(\theta \times \theta)(\mathbf{x}, \mathbf{y}) = (\theta\mathbf{x}, \theta\mathbf{y})$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbb{Z}}$. This is clearly a nonsingular map that preserves the product probability measure. Note that the mapping $f : \mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ defined by $f(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \mathbf{y}$ (coordinatewise addition) is compatible with the maps $\theta \times \theta$ on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}$ and θ on $\mathbb{R}^{\mathbb{Z}}$. Furthermore, the measure $(\mu \times \nu) \circ f^{-1}$ that f induces on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}\big)$ is the law of the process $Z_n = X_n + Y_n$, $n \in \mathbb{Z}$. Therefore, in order to prove ergodicity of the latter process, it is sufficient, by Proposition 2.1.7, to prove that the product left shift $\theta \times \theta$ is itself ergodic.

To this end, we will use the criterion for ergodicity given in (2.21). We need to prove that for every two sets $A, B \in \mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}$,

$$\frac{1}{n}\sum_{j=0}^{n-1}(\mu \times \nu)\big(A \cap (\theta \times \theta)^{-j}B\big) = (\mu \times \nu)(A)\,(\mu \times \nu)(B) \tag{2.22}$$

as $n \to \infty$. We begin with the case that A and B are measurable rectangles of the type $A = A_1 \times A_2$ and $B = B_1 \times B_2$, for A_1, A_2, B_1, B_2 measurable sets in $\mathcal{B}^{\mathbb{Z}}$. In that case, the left-hand side of (2.22) becomes

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu\big(A_1 \cap \theta^{-j}B_1\big)\nu\big(A_2 \cap \theta^{-j}B_2\big)\,. \tag{2.23}$$

Since the left shift θ is weakly mixing on the probability space $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \nu\big)$, we know that

$$\nu\big(A_2 \cap \theta^{-n}B_2\big) \to \nu(A_2)\nu(B_2)$$

in density. Since a set of zero density cannot affect the Cesaro limit of a bounded sequence, the limit of the expression in (2.23) is equal to the limit of the expression

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu\big(A_1 \cap \theta^{-j}B_1\big)\nu(A_2)\nu(B_2)$$

(and the two limits exist at the same time). Since the left shift θ is ergodic on the probability space $\left(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\right)$, this last limit exists and is equal to

$$\mu(A_1)\mu(B_1)\nu(A_2)\nu(B_2) = (\mu \times \nu)(A)\,(\mu \times \nu)(B)\,,$$

and so we have checked that (2.22) holds when A and B are measurable rectangles. Now we can proceed as in the proof of Theorem 2.2.7. Since (2.22) holds for measurable rectangles, it extends, by linearity, to the case that A and B are each a finite disjoint union of measurable rectangles. Since the latter sets form a field generating the product σ-field $\mathcal{B}^{\mathbb{Z}} \times \mathcal{B}^{\mathbb{Z}}$, we can approximate general sets A and B arbitrarily well (with respect to the product measure $\mu \times \nu$) by finite disjoint unions of measurable rectangles, and this shows that (2.22) holds in full generality. □

We now present an expected counterpart of Theorem 2.2.7, which says that weak mixing of a stationary stochastic process is equivalent to weak convergence *in density* to independence of the joint distributions of the blocks of observations separated by long periods of time. Notice that the weak convergence of probability measures on a Euclidian space is metrizable, for example by the Prokhorov metric; see, e.g., Billingsley (1999).

Theorem 2.2.17. *A stationary process* $\left(X_n,\ n \in \mathbb{Z}\right)$ *is weakly mixing if and only if for every* $k = 1, 2, \ldots,$

$$\left(X_1, \ldots, X_k, X_{n+1}, \ldots, X_{n+k}\right) \Rightarrow \left(X_1, \ldots, X_k, Y_1, \ldots, Y_k\right) \tag{2.24}$$

in density, where $(Y_1, \ldots, Y_k)$ *is an independent copy of* $(X_1, \ldots, X_k)$.

Proof. Sufficiency of (2.24) for weak mixing can be proved in exactly the same way as the sufficiency part in Theorem 2.2.7; the latter proof does not differentiate between the usual convergence and convergence in density. We prove now the necessity of (2.24) for weak mixing.

Fix k, and recall that

$$d(\mu_1, \mu_2) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^{2k}} \frac{|\hat{\mu}_1(\boldsymbol{\theta}) - \hat{\mu}_2(\boldsymbol{\theta})|}{1 + \|\boldsymbol{\theta}\|}$$

is a also metric on the space of probability measures on $\mathbb{R}^{2k}$ that metrizes weak convergence. Here $\hat{\mu}$ is the characteristic function of a probability measure μ. If we denote by ν_n the law of the random vector on the left-hand side of (2.24), and by ν the law of the random vector on the right-hand side, then by Lemma 2.2.12, we need to show only that for every $\varepsilon > 0$, the set $K_\varepsilon = \{n : d(\nu_n, \nu) > \varepsilon\}$ has density zero.

Fix, therefore, $\varepsilon > 0$, and choose $N = 1, 2, \ldots$, $\theta_0 > 0$, and $m = 1, 2, \ldots$ such that

$$P(|X_0| > N) \le \frac{\varepsilon}{12k}, \quad \theta_0 \ge \frac{2}{\varepsilon}, \quad \frac{\theta_0\sqrt{2k}}{m} \le \frac{\varepsilon}{6}\,.$$

In the following decomposition, the probability measure μ is either ν_n or ν. Note that in either case, the one-dimensional marginals of μ are the same, and they coincide with the law of X_0. We may assume that the law of X_0 puts no mass on the rational numbers (otherwise, a global scale change of the process would have this property). We write

$$\hat{\mu}(\boldsymbol{\theta}) = \int_{\mathbf{x}\in T_N} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x}) + \sum_{j_1=-m}^{m-1} \cdots \sum_{j_{2k}=-m}^{m-1} \int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x})\,,$$

where

$$T_N = \{\mathbf{x} = (x_1,\ldots,x_{2k}) \in \mathbb{R}^{2k} : |x_i| > N \text{ for some } i = 1,\ldots,2k\}\,,$$
$$I_{N,m}(j_1,\ldots,j_{2k}) = \{\mathbf{x} = (x_1,\ldots,x_{2k}) \in \mathbb{R}^{2k} :$$
$$Nj_i/m < x_i < N(j_i+1)/m,\ i = 1,\ldots,2k\},$$

$j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$. Note that by the choice of N,

$$\left|\int_{\mathbf{x}\in T_N} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x})\right| \le \frac{\varepsilon}{6}. \tag{2.25}$$

Furthermore, for each $j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$, there is a zero-density subset $K_{N,m}(j_1,\ldots,j_{2k})$ of positive integers such that for all n outside of this set,

$$\left|\nu_n\big(I_{N,m}(j_1,\ldots,j_{2k})\big) - \nu\big(I_{N,m}(j_1,\ldots,j_{2k})\big)\right| \le \frac{\varepsilon}{6(2m)^{2k}}\,. \tag{2.26}$$

We define K_ε to be the union of the sets $K_{N,m}(j_1,\ldots,j_{2k})$ over all $j_i = -m,\ldots,m-1$, $i = 1,\ldots,2k$. Clearly, K_ε is a zero-density set. It remains to show that outside of this set, we have $d(\nu_n,\nu) \le \varepsilon$.

For each $j_i = -m,\ldots,m-1,\ i = 1,\ldots,2k$, we can write

$$\int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\mu(d\mathbf{x})$$
$$= \Big[\cos\big(\boldsymbol{\theta},\mathbf{x}^{(1)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k})\big) + i\sin\big(\boldsymbol{\theta},\mathbf{x}^{(2)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k})\big)\Big]\mu\big(I_{N,m}(j_1,\ldots,j_{2k})\big)$$

for some points $\mathbf{x}^{(d)}(\boldsymbol{\theta},\mu,j_1,\ldots,j_{2k}) \in I_{N,m}(j_1,\ldots,j_{2k})$, $d = 1,2$. Applying the resulting decomposition of a characteristic function to the measures ν_n and ν, we obtain for $\boldsymbol{\theta} \in \mathbb{R}^k$ with $\|\boldsymbol{\theta}\| \le \theta_0$ and $n \notin K_\varepsilon$, using (2.25),

$$|\hat{\nu}_n(\boldsymbol{\theta}) - \hat{\nu}(\boldsymbol{\theta})| \le \frac{\varepsilon}{3}$$
$$+ \sum_{j_1=-m}^{m-1} \cdots \sum_{j_{2k}=-m}^{m-1} \left|\int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\nu_n(d\mathbf{x}) - \int_{\mathbf{x}\in I_{N,m}(j_1,\ldots,j_{2k})} e^{i(\boldsymbol{\theta},\mathbf{x})}\,\nu(d\mathbf{x})\right|.$$

Each term in the sum on the right-hand side can be bounded from above by

$$\begin{aligned}&\left|\cos\left(\boldsymbol{\theta}, \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu_n, j_1, \ldots, j_{2k})\right)\nu_n\left(I_{N,m}(j_1, \ldots, j_{2k})\right)\right.\\&\left.-\cos\left(\boldsymbol{\theta}, \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu, j_1, \ldots, j_{2k})\right)\nu\left(I_{N,m}(j_1, \ldots, j_{2k})\right)\right|\end{aligned}$$

plus the corresponding term with the sin function replacing the cos function. Note that each of the two terms can be bounded by

$$\begin{aligned}&\nu\left(I_{N,m}(j_1, \ldots, j_{2k})\right)\|\boldsymbol{\theta}\|\left\|\mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu_n, j_1, \ldots, j_{2k}) - \mathbf{x}^{(1)}(\boldsymbol{\theta}, \nu, j_1, \ldots, j_{2k})\right\|\\&\quad+\left|\nu_n\left(I_{N,m}(j_1, \ldots, j_{2k})\right) - \nu\left(I_{N,m}(j_1, \ldots, j_{2k})\right)\right|\\&\quad\le \nu\left(I_{N,m}(j_1, \ldots, j_{2k})\right)\theta_0\frac{\sqrt{2k}}{m} + \frac{\varepsilon}{6(2m)^{2k}}\\&\quad\le \frac{\varepsilon}{6}\nu\left(I_{N,m}(j_1, \ldots, j_{2k})\right) + \frac{\varepsilon}{6(2m)^{2k}}\end{aligned}$$

by (2.26) and the choice of m. Summarizing, we conclude that

$$|\hat{\nu}_n(\boldsymbol{\theta}) - \hat{\nu}(\boldsymbol{\theta})| \le \varepsilon$$

for all $\boldsymbol{\theta} \in \mathbb{R}^k$ with $\|\boldsymbol{\theta}\| \le \theta_0$ and $n \notin K_\varepsilon$. By the choice of θ_0, this shows that $d(\nu_n, \nu) \le \varepsilon$ for each $n \notin K_\varepsilon$, and the proof is complete. □

Example 2.2.18. Let $\mathbf{X} = \left(X_n,\, n \in \mathbb{Z}\right)$ be a stationary centered Gaussian process. Let μ_X be its spectral measure, i.e., a measure on $(-\pi, \pi]$ such that the covariance function of the process satisfies

$$R_X(n) = \int_{(-\pi,\pi]} e^{inx}\,\mu_X(dx),\ n \in \mathbb{Z}\,;$$

see Theorem 1.2.1. Complementing the characterization of mixing of stationary Gaussian processes in Example 2.2.8, we will show now that $\mathbf{X}$ is weakly mixing if and only if it is ergodic, and that a necessary and sufficient condition for that is that the spectral measure μ_X be atomless.

Since weak mixing implies ergodicity, we need to check only that the presence of an atom in μ_X rules out ergodicity of $\mathbf{X}$, while the absence of atoms implies weak mixing of $\mathbf{X}$.

Suppose that $a \in (-\pi, \pi]$ is an atom of μ_X, i.e., $\mu_X(\{a\}) > 0$. Let μ_1 be a measure on $(-\pi, \pi]$ obtained by removing from μ_X the atoms at $-a$ and a if $a \in (-\pi, \pi)$, or only the atom at a if $a = \pi$. Let μ_2 be the measure on $(-\pi, \pi]$ consisting of these removed atoms. Note that $\mu_X = \mu_1 + \mu_2$, and μ_2 is a nonzero measure. If $\mathbf{X}^{(j)}$ is a stationary centered Gaussian process with spectral measure μ_j, $j = 1, 2$, and if $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent, then $\mathbf{X}$ has the same distribution as

$\mathbf{X}^{(1)} + \mathbf{X}^{(2)}$, so lack of ergodicity of the process $\mathbf{X}^{(2)}$ and Proposition 2.1.6 would imply that the process $\mathbf{X}$ is not ergodic. To see that the process $\mathbf{X}^{(2)}$ is not ergodic, observe that the covariance function of this process has the form

$$\mathrm{Cov}\big(X_0^{(2)}, X_n^{(2)}\big) = \cos(an)\|\mu_2\|, \ n \in \mathbb{Z},$$

so that

$$\big(X_n^{(2)}, \ n \in \mathbb{Z}\big) \stackrel{d}{=} \Big(\|\mu_2\|^{1/2}\big(G_1 \cos(an) + G_2 \sin(an)\big), \ n \in \mathbb{Z}\Big),$$

where G_1 and G_2 are independent standard normal random variables. This shows that for the invariant set

$$A = \big\{\sup_{n\in\mathbb{Z}} |x_n| > 1\big\},$$

we have $P\big(\mathbf{X}^{(2)} \in A\big) \in (0, 1)$, and hence the process $\mathbf{X}^{(2)}$ is not ergodic.

Suppose now that the spectral measure μ_X is atomless. By Fubini's theorem, we conclude that the finite measure on $(-2\pi, 2\pi]$ given by a convolution of μ_X with itself, $F = \mu_X * \mu_X$, does not have atoms at the origin or at 2π. Note further that

$$R_X(n)^2 = \int_{(-2\pi,2\pi]} e^{inx} F(dx), \ n \in \mathbb{Z}.$$

Therefore, for $n \geq 1$,

$$\frac{1}{n}\sum_{j=0}^{n-1} R_X(j)^2 = \int_{(-2\pi,2\pi]} \frac{1}{n}\sum_{j=0}^{n-1} e^{ijx} F(dx).$$

The functions

$$\varphi_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} e^{ijx}, \ \ x \in (-2\pi, 2\pi],$$

are uniformly bounded and converge, as $n \to \infty$, to the function $\mathbf{1}\big(x = 0 \text{ or } x = 2\pi\big)$, which is equal to zero F-a.e. By the bounded convergence theorem, we conclude that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} R_X(j)^2 = 0,$$

and therefore, $R_X(n)$ converges to zero in density as $n \to \infty$. By Theorem 2.2.17, this implies that the process $\mathbf{X}$ is weakly mixing.

2.3 Strong Mixing

In the previous section, we introduced the notions of mixing and weak mixing for stationary stochastic processes. Perhaps unexpectedly, there exist notions of strong mixing of stationary processes that are different from the notion of mixing. These notions of strong mixing are not, strictly speaking, ergodic-theoretical notions in the sense of being properties of nonsingular measure-preserving maps. They are, rather, properties of certain families of σ-fields generated by a stationary process. The ergodic-theoretical notion of mixing of a stationary process turns out to be too weak to be directly applicable in certain limit theorems for stationary processes, and it is natural to view the strong mixing properties as strengthening the notion of mixing in ways that are suitable for different purposes. We discuss several of the common strong mixing notions in this section. An important point to keep in mind is that the notions we present (and a number of notions that we do not present) are collectively known as strong mixing properties. This is in spite of the fact that one of these properties is itself known as the strong mixing property. This is confusing, since often only the difference between the grammatical singular and the grammatical plural clarifies the meaning. Still, this is the existing usage, and we will adhere to this language here.

Let $\big(\Omega, \mathcal{F}, P\big)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. We define

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \big|P(A \cap B) - P(A)P(B)\big| \,. \tag{2.27}$$

Clearly, this is a measure of dependence between the σ-fields $\mathcal{A}$ and $\mathcal{B}$, and $\alpha(\mathcal{A}, \mathcal{B}) = 0$ if and only if the σ-fields $\mathcal{A}$ and $\mathcal{B}$ are independent. The following lemma lists several elementary properties of this measure of dependence.

Lemma 2.3.1. *For all sub-σ-fields $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{F}$,*

$$0 \le \alpha(\mathcal{A}, \mathcal{B}) \le \frac{1}{4} \,.$$

Further,

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{B \in \mathcal{B}} E\big(P\big(B|\mathcal{A}\big) - P(B)\big)_+ = \frac{1}{2} \sup_{B \in \mathcal{B}} E\big|P\big(B|\mathcal{A}\big) - P(B)\big|. \tag{2.28}$$

Proof. The first claim of the lemma follows from the Cauchy–Schwarz inequality: for any two events A and B,

$$\big|P(A \cap B) - P(A)P(B)\big| = |\mathrm{Cov}(\mathbf{1}_A, \mathbf{1}_B)| \le \sqrt{\mathrm{Var}(\mathbf{1}_A)\mathrm{Var}(\mathbf{1}_B)} \le \frac{1}{4} \,.$$

For the second claim, fix $B \in \mathcal{B}$, and note that

$$\sup_{A\in\mathcal{A}} \big|P(A\cap B) - P(A)P(B)\big|$$
$$= \max\Big(\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_+, \sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_-\Big).$$

Further,

$$\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_+ = \sup_{A\in\mathcal{A}} \big(E(\mathbf{1}_A\mathbf{1}_B) - E\mathbf{1}_A E\mathbf{1}_B\big)_+$$
$$= \sup_{A\in\mathcal{A}} \big(E(\mathbf{1}_A P(B|\mathcal{A})) - E\mathbf{1}_A P(B)\big)_+ = \sup_{A\in\mathcal{A}} \Big(E\big(\mathbf{1}_A(P(B|\mathcal{A})) - P(B)\big)\Big)_+$$
$$= E\big(P\big(B|\mathcal{A}\big) - P(B)\big)_+ ,$$

since the last supremum is obviously achieved on the event $A = \{P(B|\mathcal{A}) > P(B)\} \in \mathcal{A}$. Similarly,

$$\sup_{A\in\mathcal{A}} \big(P(A\cap B) - P(A)P(B)\big)_- = E\big(P\big(B|\mathcal{A}\big) - P(B)\big)_- .$$

Since for every random variable X with $EX = 0$ we have $EX_+ = EX_- = E|X|/2$, we obtain (2.28) after optimizing over $B \in \mathcal{B}$. □

Let now $\big(X_n,\, n \in \mathbb{Z}\big)$ be a stationary process defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$. Then

$$\mathcal{F}^0_{-\infty} = \sigma(X_k,\, k \le 0), \quad \mathcal{F}^\infty_n = \sigma(X_k,\, k \ge n),\ n \ge 1, \tag{2.29}$$

are sub-σ-fields of $\mathcal{F}$, and we introduce the *strong mixing coefficient* of the process $\big(X_n,\, n \in \mathbb{Z}\big)$ by

$$\alpha_X(n) := \alpha\big(\mathcal{F}^0_{-\infty}, \mathcal{F}^\infty_n\big),\ n = 1, 2, \dots . \tag{2.30}$$

Definition 2.3.2. A stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is called strongly mixing if $\alpha_X(n) \to 0$ as $n \to \infty$.

A strongly mixing stationary process is alternatively known as α-mixing. In spite of defining, as above, the strong mixing property via certain sub-σ-fields in the probability space on which a process is defined, the presence or absence of this property is determined solely by the finite-dimensional distributions of the process (Exercise 2.6.9).

Example 2.3.3. It turns out that a strongly mixing stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ has a trivial tail σ-field. To see this, let μ be the measure generated by the process on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}\big)$. In order to prove triviality of the tail σ-field, we need to check that (2.20)

holds for every set A in $\mathcal{B}^{\mathbb{Z}}$. If $A \in \sigma\big(\ldots, x_{k-1}, x_k\big)$ for some $k \geq 0$, then $A_1 = \theta^k A \in \sigma\big(\ldots, x_{-1}, x_0\big)$, and so for every $B \in \sigma(x_0, x_1, \ldots)$, by stationarity,

$$\begin{aligned} &\big|\mu\big(A \cap \theta^{-n} B\big) - \mu(A)\mu(B)\big| \\ &= \big|\mu\big(A_1 \cap \theta^{-(n-k)} B\big) - \mu(A_1)\mu(B)\big| \leq \alpha_X(n-k)\,, \end{aligned}$$

and (2.20) holds by the α-mixing of the process. Since sets A as above form a field generating the σ-field $\mathcal{B}^{\mathbb{Z}}$, we can use our usual approximating argument to extend the validity of (2.20) to all A in $\mathcal{B}^{\mathbb{Z}}$. Therefore, the tail σ-field is indeed trivial.

Combining the lessons learned from Examples 2.2.9 and 2.3.3, we immediately obtain the following proposition showing that, as the terminology implies, strong mixing guarantees ergodic-theoretical mixing.

Proposition 2.3.4. *A strongly mixing stationary process is also mixing.*

To introduce the next measure of dependence between two σ-fields, let once again $\big(\Omega, \mathcal{F}, P\big)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. Two natural probability measures on the product space $\big(\Omega \times \Omega, \mathcal{A} \times \mathcal{B}\big)$ are $P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}$ and $P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}$. Here $P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}$ is the probability measure induced on the product space by the measurable map $T_{\mathrm{id}} : \Omega \to \Omega \times \Omega$ with $T_{\mathrm{id}}(\omega) = (\omega, \omega)$, $\omega \in \Omega$, and $P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}$ is the product measure of the restrictions $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$ of P to the σ-fields $\mathcal{A}$ and $\mathcal{B}$ respectively. Note that the two measures $P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}$ and $P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}$ put the same marginal probability measures (equal to $P_{\mathcal{A}}$ and $P_{\mathcal{B}}$) on the two copies of Ω. We define

$$\beta(\mathcal{A}, \mathcal{B}) = \big\| P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}} - P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}} \big\| = \sup_{C \in \mathcal{A} \times \mathcal{B}} \big| P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}(C) - P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}(C) \big|\,, \tag{2.31}$$

the total variation distance between the probability measures $P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}$ and $P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}$.

The following lemma lists several basic properties of the measure of dependence between σ-fields defined in (2.31).

Lemma 2.3.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be sub-σ-fields of $\mathcal{F}$.*

(i) We have

$$0 \leq \beta(\mathcal{A}, \mathcal{B}) \leq 1$$

and

$$\beta(\mathcal{A}, \mathcal{B}) \geq 2\alpha(\mathcal{A}, \mathcal{B})\,. \tag{2.32}$$

(ii) An alternative way of representing $\beta(\mathcal{A}, \mathcal{B})$ is

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup_{\mathcal{I}, \mathcal{J}} \sum_{i=1}^{I} \sum_{j=1}^{J} \big| P(A_i \cap B_j) - P(A_i) P(B_j) \big|\,, \tag{2.33}$$

where the supremum is taken over all finite partitions $\mathcal{I} = \{A_1, \ldots, A_I\}$ *and* $\mathcal{J} = \{B_1, \ldots, B_J\}$ *of* Ω *into* $\mathcal{A}$*-measurable sets and* $\mathcal{B}$*-measurable sets, respectively.*

(iii) $\beta(\mathcal{A}, \mathcal{B}) = 0$ *if and only if the* σ*-fields* $\mathcal{A}$ *and* $\mathcal{B}$ *are independent.*

Proof. The first claim in part (i) is an obvious property of the total variation distance between two probability measures. For the second claim in part (i), note that if $A \in \mathcal{A}$ and $B \in \mathcal{B}$, then $P_{\text{id}}^{\mathcal{A},\mathcal{B}}(A \times B) = P(A \cap B)$. If we define for such events A and B the sets $C_1(A, B) = A \times B$ and $C_2(A, B) = A^c \times B^c$, then these product sets are disjoint events in $\mathcal{A} \times \mathcal{B}$, and

$$\begin{aligned} P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B)\big) &= P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_2(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_2(A,B)\big) \\ &= P(A \cap B) - P(A)P(B)\,. \end{aligned}$$

Therefore,

$$\begin{aligned} \beta(\mathcal{A}, \mathcal{B}) &\geq \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| P_{\text{id}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B) \cup C_2(A,B)\big) - P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big(C_1(A,B) \cup C_2(A,B)\big) \right| \\ &= 2 \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \left| P(A \cap B) - P(A)P(B) \right| = 2\alpha(\mathcal{A}, \mathcal{B})\,, \end{aligned}$$

proving (2.32).

For part (ii), let Q be a probability measure on $\big(\Omega \times \Omega, \mathcal{A} \times \mathcal{B}\big)$ such that both $P_{\text{id}}^{\mathcal{A},\mathcal{B}} \ll Q$ and $P_{\text{ind}}^{\mathcal{A},\mathcal{B}} \ll Q$; an example of such a measure is $Q = \big(P_{\text{id}}^{\mathcal{A},\mathcal{B}} + P_{\text{ind}}^{\mathcal{A},\mathcal{B}}\big)/2$. Setting

$$f_{\text{id}} = \frac{dP_{\text{id}}^{\mathcal{A},\mathcal{B}}}{dQ}\,, \quad f_{\text{ind}} = \frac{dP_{\text{ind}}^{\mathcal{A},\mathcal{B}}}{dQ}\,,$$

we have

$$\begin{aligned} \left\| P_{\text{id}}^{\mathcal{A},\mathcal{B}} - P_{\text{ind}}^{\mathcal{A},\mathcal{B}} \right\| &= \frac{1}{2} \int_{\Omega \times \Omega} |f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)|\, Q(d(\omega_1, \omega_2)) \\ &= \int_{\Omega \times \Omega} \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big)_+\, Q(d(\omega_1, \omega_2)) \\ &= \int_C \big(f_{\text{id}}(\omega_1, \omega_2) - f_{\text{ind}}(\omega_1, \omega_2)\big)\, Q(d(\omega_1, \omega_2))\,, \end{aligned}$$

where

$$C = \{(\omega_1, \omega_2) : f_{\text{id}}(\omega_1, \omega_2) > f_{\text{ind}}(\omega_1, \omega_2)\}\,. \tag{2.34}$$

In particular, for all relevant finite partitions $\mathcal{I} = \{A_1, \ldots, A_I\}$ and $\mathcal{J} = \{B_1, \ldots, B_J\}$ of Ω, we have

$$\begin{aligned}
\beta(\mathcal{A}, \mathcal{B}) &= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \int_{A_i \times B_j} |f_{\mathrm{id}}(\omega_1, \omega_2) - f_{\mathrm{ind}}(\omega_1, \omega_2)| \, Q(d(\omega_1, \omega_2)) \\
&\geq \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| \int_{A_i \times B_j} (f_{\mathrm{id}}(\omega_1, \omega_2) - f_{\mathrm{ind}}(\omega_1, \omega_2)) \, Q(d(\omega_1, \omega_2)) \right| \\
&= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| P_{\mathrm{id}}^{\mathcal{A},\mathcal{B}}(A_i \times B_j) - P_{\mathrm{ind}}^{\mathcal{A},\mathcal{B}}(A_i \times B_j) \right| \\
&= \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| (A_j \cap B_j) - P(A_j) P(B_j) \right| .
\end{aligned}$$

Taking the supremum over all relevant partitions yields

$$\beta(\mathcal{A}, \mathcal{B}) \geq \frac{1}{2} \sup_{\mathcal{I}, \mathcal{J}} \sum_{i=1}^{I} \sum_{j=1}^{J} \left| P(A_i \cap B_j) - P(A_i) P(B_j) \right| .$$

On the other hand, using the fact that the disjoint unions of measurable rectangles of the form $A \times B$ with $A \in \mathcal{A}$, $B \in \mathcal{B}$, form a field generating the product σ-field $\mathcal{A} \times \mathcal{B}$, for every $\varepsilon > 0$ we can select such a disjoint union C_0 satisfying $Q(C \triangle C_0) \leq \varepsilon$, where C is the set in (2.34); see Corollary 1, p. 169, in Billingsley (1995). Given $\delta > 0$, we can select $\varepsilon > 0$ so small that this property of C_0 will imply that

$$\begin{aligned}
&\left| \int_{C} \big(f_{\mathrm{id}}(\omega_1, \omega_2) - f_{\mathrm{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) \right. \\
&\left. - \int_{C_0} \big(f_{\mathrm{id}}(\omega_1, \omega_2) - f_{\mathrm{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) \right| \leq \delta;
\end{aligned}$$

see, e.g., Problem 16.8 in Billingsley (1995). We conclude that for every $\delta > 0$, there are relevant finite partitions $\mathcal{I} = \{A_1, \ldots, A_I\}$ and $\mathcal{J} = \{B_1, \ldots, B_J\}$ of Ω such that

$$\begin{aligned}
\beta(\mathcal{A}, \mathcal{B}) &\leq \int_{C_0} \big(f_{\mathrm{id}}(\omega_1, \omega_2) - f_{\mathrm{ind}}(\omega_1, \omega_2)\big) \, Q(d(\omega_1, \omega_2)) + \delta \\
&\leq \sum_{i=1}^{I} \sum_{j=1}^{J} \big(P(A_i \cap B_j) - P(A_i) P(B_j)\big)_+ + \delta .
\end{aligned}$$

Since for all partitions we have

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\big(P(A_i\cap B_j)-P(A_i)P(B_j)\big)_+=\sum_{i=1}^{I}\sum_{j=1}^{J}\big(P(A_i\cap B_j)-P(A_i)P(B_j)\big)_-,$$

we conclude, after switching to the supremum over the relevant partitions, that for every $\delta>0$,

$$\beta(\mathcal{A},\mathcal{B})\le\frac{1}{2}\sup_{\mathcal{I},\mathcal{J}}\sum_{i=1}^{I}\sum_{j=1}^{J}\big|P(A_i\cap B_j)-P(A_i)P(B_j)\big|+\delta\,.$$

Letting $\delta\to 0$ proves (2.33).

Part (iii) of the lemma is an immediate conclusion from (2.33). □

Let, once again, $\big(X_n,\,n\in\mathbb{Z}\big)$ be a stationary process on a probability space $\big(\Omega,\mathcal{F},P\big)$. Using the notation in (2.29) for the σ-fields generated by the process, we define the *beta mixing coefficient* of the process $\big(X_n,\,n\in\mathbb{Z}\big)$ by

$$\beta_X(n):=\beta\big(\mathcal{F}_{-\infty}^0,\mathcal{F}_n^\infty\big),\ n=1,2,\dots\,. \tag{2.35}$$

Definition 2.3.6. A stationary process $\big(X_n,\,n\in\mathbb{Z}\big)$ is called absolutely regular, or β-mixing, if $\beta_X(n)\to 0$ as $n\to\infty$.

An immediate conclusion of part (i) of Lemma 2.3.5 is the following comparison of the strong mixing property and the absolute regularity property.

Proposition 2.3.7. *An absolutely regular stationary process is also strongly mixing.*

Let, once again, $\big(\Omega,\mathcal{F},P\big)$ be a probability space, and $\mathcal{A}$ and $\mathcal{B}$ two sub-σ-fields of $\mathcal{F}$. We introduce a third notion of dependence between two σ-fields by

$$\phi(\mathcal{A},\mathcal{B})=\sup_{A\in\mathcal{A}:P(A)>0,B\in\mathcal{B}}\big|P(B|A)-P(B)\big|\,. \tag{2.36}$$

An immediate observation is that unlike the measures of dependence introduced in (2.27) and (2.31), this new notion of dependence does not appear to be symmetric, in the sense that there seems to be no reason why $\phi(\mathcal{A},\mathcal{B})$ and $\phi(\mathcal{B},\mathcal{A})$ should coincide. In fact, the two quantities are, in general, different. See, for instance, Problem 2.6.10.

The next lemma summarizes the basic properties of the measure of dependence $\phi(\mathcal{A},\mathcal{B})$.

Lemma 2.3.8. *Let $\mathcal{A}$ and $\mathcal{B}$ be sub-σ-fields of $\mathcal{F}$.*

(i) We have

$$0\le\phi(\mathcal{A},\mathcal{B})\le 1$$

and

$$\phi(\mathcal{A},\mathcal{B}) \geq \beta(\mathcal{A},\mathcal{B}). \tag{2.37}$$

(ii) $\phi(\mathcal{A},\mathcal{B}) = 0$ *if and only if the* σ*-fields* $\mathcal{A}$ *and* $\mathcal{B}$ *are independent.*

Proof. The first statement of part (i) is obvious. For the second statement of part (i) we use the representation of $\beta(\mathcal{A},\mathcal{B})$ given in (2.33). Let $\{A_1,\dots,A_I\}$ and $\{B_1,\dots,B_J\}$ be partitions of Ω into $\mathcal{A}$-measurable sets and $\mathcal{B}$-measurable sets, respectively. Then

$$\begin{aligned}
&\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i\cap B_j)-P(A_i)P(B_j)\right| \\
&=\sum_{i=1}^{I}P(A_i)\sum_{j=1}^{J}\left|P(B_j|A_i)-P(B_j)\right| \\
&=2\sum_{i=1}^{I}P(A_i)\left|P(B_{+,i}|A_i)-P(B_{+,i})\right|,
\end{aligned}$$

where $B_{+,i}$ is the union of all B_j, $j = 1,\dots,J$, such that $P(B_j|A_i) \geq P(B_j)$. Therefore,

$$\begin{aligned}
&\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i\cap B_j)-P(A_i)P(B_j)\right| \\
&\leq 2\phi(\mathcal{A},\mathcal{B})\sum_{i=1}^{I}P(A_i)=2\phi(\mathcal{A},\mathcal{B})
\end{aligned}$$

for all partitions, and the second statement of part (i) follows from (2.33). The claim of part (ii) is obvious.

For a stationary process $\left(X_n,\ n\in\mathbb{Z}\right)$ defined on a probability space $\left(\Omega,\mathcal{F},P\right)$, the *phi mixing coefficient* is defined by

$$\phi_X(n) := \phi\left(\mathcal{F}_{-\infty}^{0},\mathcal{F}_{n}^{\infty}\right),\ n=1,2,\dots, \tag{2.38}$$

where the σ-fields are defined in (2.29). The phi mixing coefficient leads to one more notion of strong mixing for stationary processes.

Definition 2.3.9. A stationary process $\left(X_n,\ n\in\mathbb{Z}\right)$ is called ϕ-mixing if $\phi_X(n)\to 0$ as $n\to\infty$.

By Lemma 2.3.8, we immediately obtain the following statement.

Proposition 2.3.10. *A* ϕ*-mixing stationary process is absolutely regular (and hence also strongly mixing).*

The lack of symmetry of the notion of dependence between σ-fields introduced in (2.36) makes it possible to introduce the "time-reversed" versions of the ϕ mixing coefficient and of the ϕ-mixing property. For a stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$, one simply defines

$$\phi_X^{\mathrm{rev}}(n) := \phi\big(\mathcal{F}_n^{\infty}, \mathcal{F}_{-\infty}^{0}\big),\ n = 1, 2, \dots , \tag{2.39}$$

leading to the following definition.

Definition 2.3.11. A stationary process $\big(X_n,\, n \in \mathbb{Z}\big)$ is called reverse-ϕ-mixing if $\phi_X^{\mathrm{rev}}(n) \to 0$ as $n \to \infty$.

Clearly, an obvious version of Proposition 2.3.10 for reverse-ϕ-mixing holds. We mention that there exist stationary processes that are ϕ-mixing but not reverse-ϕ-mixing (and conversely); see Rosenblatt (1971).

2.4 Conservative and Dissipative Maps

In this section, we return to maps on general σ-finite measure spaces and discuss certain properties of recurrence of such maps. Their probabilistic significance will become apparent in Section 3.6.

Let $\big(E, \mathcal{E}, m\big)$ be a σ-finite measure space. In this section we will deal with a nonsingular map $\phi : E \to E$ that preserves the measure m (a measure-preserving map). A set $W \in \mathcal{E}$ is called *wandering* if the sets $\big(\phi^{-n}(W),\, n = 1, 2, \dots\big)$ are mutually disjoint modulo m (recall that two measurable sets A and B are disjoint modulo m if $m(A \cap B) = 0$). This means that for every point x in a wandering set W, apart from a possible subset of W of measure zero, the trajectory $\phi^n(x)$, $n = 1, 2, \dots$, never reenters the set A.

Every set of measure zero is trivially a wandering set. What sets different maps apart is the existence of nontrivial wandering sets.

Definition 2.4.1. A measure-preserving map ϕ on $\big(E, \mathcal{E}, m\big)$ is conservative if it does not admit a wandering set of positive measure.

Observe that every measure-preserving map on a finite measure space is automatically conservative, since a finite measure space cannot contain infinitely many sets of equal and positive measure that are mutually disjoint modulo m.

Example 2.4.2. Let $E = \mathbb{Z}$ and let m be the counting measure. The right shift $\phi(x) = x + 1$ for $x \in \mathbb{Z}$ is measure-preserving. Since the set $W = \{0\}$ is obviously a wandering set of positive measure, the right shift is not conservative.

The following result shows that in general, a map ϕ has both a "conservative part" and a part that is "purely nonconservative."

Theorem 2.4.3. The Hopf decomposition *Let ϕ be a measure-preserving map on a σ-finite measure space $(E, \mathcal{E}, m)$. Then there is a partition of E into ϕ-invariant sets $\mathcal{C}(\phi)$ and $\mathcal{D}(\phi)$ such that*

(i) there is no wandering set of a positive measure that is a subset of $\mathcal{C}(\phi)$;
(ii) there is a wandering set W such that $\mathcal{D}(\phi) = \cup_{n=-\infty}^{\infty} \phi^n(W)$ modulo m.

The partition $E = \mathcal{C}(\phi) \cup \mathcal{D}(\phi)$ is unique in the sense that if $E = \mathcal{C}_1 \cup \mathcal{D}_1$, where $\mathcal{C}_1$ and $\mathcal{D}_1$ are ϕ-invariant sets satisfying (i) and (ii), then $m\big(\mathcal{C}(\phi) \triangle \mathcal{C}_1\big) = m\big(\mathcal{D}(\phi) \triangle \mathcal{D}_1\big) = 0$.

Proof. We begin with a probability measure P on $(E, \mathcal{E})$ equivalent to the σ-finite measure m. We will construct recursively an increasing sequence of ϕ-invariant sets $(I_n,\ n = 0, 1, 2, \ldots)$ and a sequence of wandering sets $(W_n,\ n = 1, 2, \ldots)$ as follows. Let $I_0 = \emptyset$. For $n = 0, 1, 2, \ldots$, let

$$\alpha_n = \sup\{P(W),\ W \subset I_n^c,\ W \text{ wandering}\}\,.$$

Let $W_{n+1} \subset I_n^c$ be such that $P(W_{n+1}) \geq \alpha_n/2$ (if $\alpha_n = 0$, we can always choose $W_{n+1} = \emptyset$), and let

$$I_{n+1} = I_n \cup \bigcup_{k=-\infty}^{\infty} \phi^k(W_{n+1})\,.$$

Since P is a probability measure, we see that $\alpha_n \to 0$ as $n \to \infty$. Set $W = \cup_{n=1}^{\infty} W_n$. It is clear that W is a wandering set.

Let $\mathcal{D}(\phi) = \cup_{k=-\infty}^{\infty} \phi^k(W)$. Then $\mathcal{D}(\phi)$ is a ϕ-invariant set, and so is its complement, which we denote by $\mathcal{C}(\phi)$. By construction, $\mathcal{C}(\phi) \subset I_n^c$ for each $n = 1, 2, \ldots$, and hence so is every wandering subset V of $\mathcal{C}(\phi)$. This implies that $P(V) \leq \alpha_n$ for each n, and since $\alpha_n \to 0$, we conclude that $P(V) = 0$. Since P is equivalent to m, we also have $m(V) = 0$.

To prove uniqueness, suppose, for example, that $m\big(\mathcal{C}_1 \setminus \mathcal{C}(\phi)\big) > 0$. Then for some n, $m\big(\mathcal{C}_1 \cap \phi^n(W)\big) > 0$. Since a subset of a wandering set is a wandering set, this contradicts the fact that $\mathcal{C}_1$ possesses property (i). □

The Hopf decomposition $E = \mathcal{C}(\phi) \cup \mathcal{D}(\phi)$ is called the decomposition of E into the conservative part of E and the dissipative part of E with respect to ϕ. It is usual to say simply that $\mathcal{C}(\phi)$ and $\mathcal{D}(\phi)$ are the conservative part of ϕ and the dissipative part of ϕ respectively. We see immediately that ϕ is conservative if and only if its dissipative part vanishes (modulo m). The picture is completed by the following definition.

Definition 2.4.4. A measure-preserving map ϕ on $(E, \mathcal{E}, m)$ is dissipative if $\mathcal{C}(\phi) = \emptyset$ modulo m.

Since an ergodic map ϕ cannot have two disjoint invariant sets of positive measure, each ergodic map is either conservative or dissipative. For the right shift on the integers in Example 2.4.2, it is immediate that $\mathcal{D}(\phi) = \mathbb{Z}$, and so the right shift is dissipative.

A useful and intuitive criterion for distinguishing between a conservative map and a dissipative map, or, more generally, between the conservative and dissipative parts of a map, is through the difference in the behavior of certain sums.

Theorem 2.4.5. *Let ϕ be a measure-preserving map on a σ-finite measure space $(E, \mathcal{E}, m)$. For every function $f \in L_1(m)$ such that $f > 0$ m-a.e., we have*

$$\left\{ x \in E : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) = \infty \right\} = \mathcal{C}(\phi) \ \textit{modulo } m. \tag{2.40}$$

Proof. We first show that the set on the left-hand side of (2.40) is, modulo m, a subset of $\mathcal{C}(\phi)$. By Theorem 2.4.3, it is enough to prove that for every wandering set W,

$$m\left(x \in W : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) = \infty \right) = 0\,. \tag{2.41}$$

However, for $k = 1, 2, \ldots$, since ϕ is measure-preserving and W is wandering,

$$\begin{aligned} &\int_W \sum_{n=1}^{k} f \circ \phi^n \, dm = \sum_{n=1}^{k} \int_E \mathbf{1}_W f \circ \phi^{k-n+1} \, dm \\ &= \sum_{n=1}^{k} \int_E \mathbf{1}_W \circ \phi^n f \circ \phi^{k+1} \, dm = \int_E \left(\sum_{n=1}^{k} \mathbf{1}_W \circ \phi^n \right) f \circ \phi^{k+1} \, dm \\ &\le \int_E f \circ \phi^{k+1} \, dm = \int_E f \, dm\,. \end{aligned}$$

By the monotone convergence theorem,

$$\int_W \sum_{n=1}^{\infty} f \circ \phi^n \, dm \le \int_E f \, dm < \infty\,,$$

which clearly implies (2.41).

In order to prove the second inclusion in (2.40), suppose that to the contrary,

$$m\left(x \in \mathcal{C}(\phi) : \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) < \infty \right) > 0\,.$$

Let $\varepsilon > 0$ be such that

$$m\left(x \in \mathcal{C}(\phi) : \phi^j(x) > \varepsilon \text{ for some } j \in \mathbb{Z},\ \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) < \infty \right) > 0\,. \tag{2.42}$$

Let $\mathcal{C}_\varepsilon$ be the set of positive measure in (2.42). Define W to be the set of those points x in $\mathcal{C}_\varepsilon$ such that

$$\sup\{j \in \mathbb{Z} : f(\phi^j(x)) > \varepsilon\} = 0\,.$$

Then W is a set of positive measure, and it is clearly a wandering set. It is also a subset of $\mathcal{C}(\phi)$, which is impossible by the definition of the conservative part of ϕ. This contradiction proves the second inclusion in (2.40). □

We finish this section with a refinement of the Hopf decomposition that will prove very useful to us in the sequel in a discussion of the memory of stationary non-Gaussian infinitely divisible processes.

Definition 2.4.6. A measure-preserving map ϕ on $(E, \mathcal{E}, m)$ is positive if there is a finite measure $\tilde{m}$ on $(E, \mathcal{E})$ that is preserved by ϕ and is equivalent to the measure m.

The following statement is immediate.

Proposition 2.4.7. *A positive map ϕ does not admit a wandering set of positive measure and hence is conservative.*

Proof. Indeed, if W is a wandering set with $m(W) > 0$, then $\tilde{m}(W) > 0$ as well. Since the sets $(\phi^{-n}(W),\ n = 1, 2, \ldots)$ are disjoint modulo m, they are also disjoint modulo $\tilde{m}$, so that

$$\tilde{m}\left(\bigcup_{n=1}^{\infty} \phi^{-n}(W)\right) = \sum_{n=1}^{\infty} \tilde{m}(\phi^{-n}(W)) = \infty,$$

since ϕ preserves $\tilde{m}$. This contradicts the fact that $\tilde{m}$ is a finite measure.

The fact that a positive map ϕ is also conservative follows now from Theorem 2.4.3, since if ϕ had a nontrivial dissipative part, it would admit a wandering set of positive measure. □

As in the case of the Hopf decomposition, in general a map ϕ has both a "positive part" and a part that is "purely nonpositive."

Theorem 2.4.8. The positive–null decomposition

Let ϕ be a measure-preserving map on a σ-finite measure space $(E, \mathcal{E}, m)$. Then there is a partition of E into ϕ-invariant sets $\mathcal{P}(\phi)$ and $\mathcal{N}(\phi)$ such that

(i) ϕ is positive on $\mathcal{P}(\phi)$;
(ii) no ϕ-invariant measurable set $A \subseteq \mathcal{N}(\phi)$ satisfies $0 < m(A) < \infty$.

The partition $E = \mathcal{P}(\phi) \cup \mathcal{N}(\phi)$ is unique in the sense that if $E = \mathcal{P}_1 \cup \mathcal{N}_1$, where $\mathcal{P}_1$ and $\mathcal{N}_1$ are ϕ-invariant sets satisfying (i) and (ii), then $m(\mathcal{P}(\phi) \triangle \mathcal{P}_1) = m(\mathcal{N}(\phi) \triangle \mathcal{N}_1) = 0$.

Proof. We proceed by "exhaustion," as in the proof of Theorem 2.4.3. We again begin with a probability measure P on $(E, \mathcal{E})$ equivalent to the σ-finite measure m and construct recursively an increasing sequence of ϕ-invariant sets $(I_n,\ n = 0, 1, 2, \ldots)$ and a sequence of sets $(F_n,\ n = 1, 2, \ldots)$ as follows. Let $I_0 = \emptyset$. For $n = 0, 1, 2, \ldots$, let

$$\alpha_n = \sup\{P(F),\ F \subset I_n^c,\ F\ \phi\text{-invariant and } m(F) < \infty\}\,.$$

Let $F_{n+1} \subset I_n^c$ be such that $P(F_{n+1}) \geq \alpha_n/2$ (use $F_{n+1} = \emptyset$ if $\alpha_n = 0$), and let $I_{n+1} = I_n \cup F_{n+1}$. Finally, let $\mathcal{P}(\phi) = \cup_{n=1}^{\infty} F_n$. By construction, $\mathcal{P}(\phi)$ is an invariant set, and we denote its complement by $\mathcal{N}(\phi)$. We now check that these sets satisfy (i) and (ii) of the theorem. If $m(\mathcal{P}(\theta)) = 0$, then ϕ is trivially positive on $\mathcal{P}(\phi)$, since we can take $\tilde{m}$ to be the null measure. If $m(\mathcal{P}(\theta)) > 0$, then we can construct a nontrivial ϕ-invariant finite measure on $\mathcal{P}(\phi)$, equivalent to m on that set, by setting

$$\frac{d\tilde{m}}{dm} = 2^{-n} \text{ on } F_n \text{ if } m(F_n) > 0,\ n = 1, 2, \ldots\,.$$

Therefore, ϕ is positive on $\mathcal{P}(\phi)$. The property (ii) holds by the construction of the set $\mathcal{P}(\phi)$.

It remains to prove the uniqueness of a decomposition. Suppose, once again, that $m\big(\mathcal{C}_1 \setminus \mathcal{C}(\phi)\big) > 0$. Then there is a probability measure m_1 supported by $\mathcal{C}_1 \setminus \mathcal{C}(\phi)$, equivalent to m on that set and invariant under ϕ. The function $g = dm_1/dm$ is ϕ-invariant, finite, and positive on $\mathcal{C}_1 \setminus \mathcal{C}(\phi)$ modulo m, and integrates to 1 with respect to m. Therefore, there are $0 < \varepsilon_1 \leq \varepsilon_2 < \infty$ such that

$$\int_{\mathcal{C}_1 \setminus \mathcal{C}(\phi)} g(x)\mathbf{1}\big(\varepsilon_1 \leq g(x) \leq \varepsilon_2\big)\, m(dx) > 0\,.$$

The set

$$\{x \in \mathcal{C}_1 \setminus \mathcal{C}(\phi) :\ \varepsilon_1 \leq g(x) \leq \varepsilon_2\}$$

is a ϕ-invariant set of finite positive measure m. Since this set is also a subset of $\mathcal{N}(\phi)$, we obtain a contradiction with the fact that $\mathcal{N}(\phi)$ satisfies (ii). □

We call a positive–null decomposition $E = \mathcal{P}(\phi) \cup \mathcal{N}(\phi)$ the decomposition of E into the positive part of E and the null part of E with respect to ϕ. With terminology similar to that in the Hopf decomposition, we say that $\mathcal{P}(\phi)$ and $\mathcal{N}(\phi)$ are the positive part of ϕ and the null part of ϕ respectively. Then ϕ is positive if and only if its null part vanishes (modulo m). We observe that it follows from Proposition 2.4.7 that

$$\mathcal{P}(\phi) \subseteq \mathcal{C}(\phi),\ \mathcal{D}(\phi) \subseteq \mathcal{N}(\phi)\,, \tag{2.43}$$

and the Hopf decomposition can be combined with the positive–null decomposition into a three-way decomposition of E: let $\mathcal{CN}(\phi) = \mathcal{C}(\phi) \cap \mathcal{N}(\phi)$. Then we decompose

$$E = \mathcal{P}(\phi) \cup \mathcal{CN}(\phi) \cup \mathcal{D}(\phi) \tag{2.44}$$

uniquely into ϕ-invariant sets that inherit their properties from the Hopf decomposition and the positive–null decomposition. Finally, we also introduce the following definition.

Definition 2.4.9. A measure-preserving map ϕ on $\big(E, \mathcal{E}, m\big)$ is null if $\mathcal{P}(\phi) = \emptyset$ modulo m.

Our next goal is to understand better the structure of the positive–null decomposition. To this end, we introduce a new notion. A set $W \in \mathcal{E}$ is called *weakly wandering* if there is a sequence $n_k \to \infty$ such that the sets $\big(\phi^{-n_k}(W),\, k = 0, 1, 2, \ldots\big)$, are disjoint modulo m (with $n_0 = 0$). Clearly, every wandering set is weakly wandering as well. Furthermore, Proposition 2.4.7 immediately extends, with the same proof, to the following.

Proposition 2.4.10. *A positive map ϕ does not admit a weakly wandering set of positive measure.*

The following theorem, which we prove only partially, completely clarifies the connection between weakly wandering sets and the positive–null decomposition. It also presents a "weighted version" of the criterion for distinguishing the parts in the Hopf decomposition, presented in Theorem 2.40. The weighted version allows one to distinguish the parts in the positive–null decomposition. Denote by $\mathcal{W}$ the set of sequences $(w_n,\, n = 1, 2, \ldots)$ with the following properties:

$$w_n > 0,\ w_{n+1} \le w_n,\ n = 1, 2, \ldots,\ \sum_{n=1}^{\infty} w_n = \infty\,. \tag{2.45}$$

Theorem 2.4.11. *Let ϕ be a measure-preserving map on a σ-finite measure space $\big(E, \mathcal{E}, m\big)$.*

(i) There exist a set W and a sequence $n_k \to \infty$ such that the sets $\big(\phi^{-n_k}(W),\, k = 0, 1, 2, \ldots\big)$ are disjoint modulo m and $\mathcal{N}(\phi) = \cup_{k=1}^{\infty} \phi^{-n_k}(W)$ modulo m (the set W is automatically weakly wandering).

(ii) For every sequence $(w_n,\, n = 1, 2, \ldots)$ in $\mathcal{W}$ and every function $f \in L_1(m)$ such that $f > 0$ m-a.e,

$$\sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) = \infty \text{ } m\text{-a.e. on } \mathcal{P}(\phi). \tag{2.46}$$

Further, there is a sequence $(w_n, n = 1, 2, \ldots)$ *in* $\mathcal{W}$ *such that for every function* $f \in L_1(m)$ *such that* $f > 0$ *m-a.e., we have*

$$\left\{ x \in E : \sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) < \infty \right\} = \mathcal{N}(\phi) \ \textit{modulo } m. \tag{2.47}$$

Proof. Part (i) of the theorem is in Jones and Krengel (1974). For the first statement of part (ii), fix a sequence in $\mathcal{W}$. We may assume without loss of generality that $w_1 = 1$. Let $p_n = w_n - w_{n+1}$, $n = 1, 2, \ldots$, and $p_\infty = \lim_{n\to\infty} w_n$, so that (p_n) are probabilities on $\mathbb{N} \cup \{+\infty\}$. By the definition of the set $\mathcal{W}$, it follows that this probability distribution has an infinite mean. Then

$$\sum_{n=1}^{\infty} w_n f\big(\phi^n(x)\big) = p_\infty \sum_{n=1}^{\infty} f\big(\phi^n(x)\big) + \sum_{j=1}^{\infty} p_j \sum_{n=1}^{j} f\big(\phi^n(x)\big)\,. \tag{2.48}$$

If $m\big(\mathcal{P}(\phi)\big) > 0$, then there exists a probability measure $\tilde{m}$ supported by $\mathcal{P}(\phi)$ and equivalent to m on that set. By the pointwise ergodic theorem,

$$\frac{1}{n} \sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to g_f(x) \ \text{ as } n \to \infty$$

for P-almost every $x \in \mathcal{P}(\phi)$, hence for m-almost every $x \in \mathcal{P}(\phi)$. Furthermore, the limit is the conditional expectation of a positive random variable, hence is itself a.s. positive. Now (2.46) follows from (2.48) and the fact that the probabilities (p_n) have an infinite mean.

The second statement of part (ii) is in Theorem 3 in Krengel (1967). □

The following example should clarify the distinction between a wandering set and a weakly wandering set.

Example 2.4.12. Let $0 < p < 1$. The transition rule $p_{i,i+1} = 1 - p_{i,i-1} = p$ for $i \in \mathbb{Z}$ defines a Markov chain on $\mathbb{Z}$, the simple random walk that takes a step to the right with probability p. This Markov chain is transient if $p \neq 1/2$, and it is null recurrent if $p = 1/2$. In any case, it has an infinite invariant measure, which is the counting measure on $\mathbb{Z}$. We use the law of this random walk to construct a σ-finite measure m on $E = \mathbb{Z}^{\mathbb{Z}}$ by

$$m(A) = \sum_{i\in\mathbb{Z}} P_i\big(\text{a realization of the random walk is in } A\big)\,. \tag{2.49}$$

The probability on the right-hand side of (2.49) is computed according to the law of the Markov chain that visits state i at time zero. In the special case of a simple random walk, this law can be constructed very simply: let $(W_n^+, n = 0, 1, 2, \ldots)$ and $(W_n^-, n = 0, 1, 2, \ldots)$ be two independent simple random walks starting at zero, the

first one stepping to the right with probability p, and the second one stepping to the right with probability $1-p$. Then the stochastic process

$$\left(\ldots, i + W_2^-, i + W_1^-, i, i + W_1^+, i + W_2^+, \ldots\right)$$

has the law P_i on E. Since the counting measure is an invariant measure for the random walk, the measure m defined by (2.49) is invariant under the left shift on E.

Suppose first that $p \neq 1/2$, and consider the set

$$A = \left\{\mathbf{x} \in \mathbb{Z}^{\mathbb{Z}} : x_0 = 0,\, x_n \neq 0 \text{ for } n > 0\right\}. \tag{2.50}$$

By construction, the set A in (2.50) is a wandering set, irrespective of the value of p. We claim that for $p \neq 1/2$, we have

$$\bigcup_{n=-\infty}^{\infty} \phi^n(A) = E \text{ modulo } m. \tag{2.51}$$

According to Theorem 2.4.3, this would show that the left shift ϕ is dissipative in this case. To see that (2.51) holds, note that

$$\left(\bigcup_{n=-\infty}^{\infty} \phi^n(A)\right)^c = \left\{\mathbf{x} :\in \mathbb{Z} :\, x_n \neq 0 \text{ for all } n \in \mathbb{Z}\right\}$$
$$\cup \left\{\mathbf{x} :\in \mathbb{Z} :\, x_n = 0 \text{ for arbitrary large } n \in \mathbb{Z}\right\}.$$

The first set above has zero measure with respect to m, because a simple random walk converges a.s. to the two different infinities as the times goes to $\pm\infty$, and does so without skipping steps. Therefore, regardless of the initial state, it will a.s. visit state zero. The second set above has zero measure with respect to m, because a simple random walk with $p \neq 1/2$ is transient, and so it visits any given state only finitely many times. This proves (2.51).

This argument for the relation (2.51) fails in the symmetric case $p = 1/2$, since in that case, the random walk is recurrent. This actually implies that the left shift ϕ is conservative with respect to m, and hence does not admit any wandering set of positive measure. In order to see that ϕ is now conservative, consider a positive measurable function on $\mathbb{Z}$ given by $f(\mathbf{x}) = 2^{-|x_0|}$. It is clear that $f \in L_1(m)$. However,

$$\sum_{n=1}^{\infty} f\left(\phi^n(\mathbf{x})\right) \geq \sum_{n=1}^{\infty} \mathbf{1}(x_n = 0) = \infty$$

m-a.e., since the simple symmetric random walk is recurrent and hence visits state zero infinitely many times. By Theorem 2.4.5, $\mathcal{C}(\theta) = E$ modulo m, and ϕ is conservative.

It turns out, however, that in the case $p = 1/2$, the left shift admits quite "large" weakly wandering sets, and we will construct a family of such sets. Let $n_1, n_2, \ldots$ and $b_0, b_1, b_2, \ldots$ be two strictly increasing sequences of positive integers. Let

$$B = \left\{ \mathbf{x} \in \mathbb{Z}^{\mathbb{Z}} : x_0 \in (-b_0, b_0),\ x_{-n_k} \in (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k),\ k \geq 1 \right\}. \tag{2.52}$$

Observe that for each $k \geq 1$ and $\mathbf{x} \in \phi^{-n_k}(B)$, we have $x_0 \in (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k)$, so the sets $\left(\phi^{-n_k}(B),\ k = 1, 2, \ldots\right)$ are disjoint, and hence the set B is weakly wandering. We now show that one can choose the sequences $n_1, n_2, \ldots$ and $b_0, b_1, b_2, \ldots$ in such a way as to make this set "large." To this end, we will simply note that by the elementary central limit theorem, for a simple symmetric random walk $(W_n,\ n = \ldots, -1, 0, 1, 2, \ldots)$, the law of $n^{-1/2} W_n$ converges weakly, as $n \to \infty$, to the standard normal law, whence for every $x > 0$,

$$P_0(|W_n| \leq x) \to 0 \text{ as } n \to \infty. \tag{2.53}$$

Let $\varepsilon > 0$ be a small number. We will choose b_0 depending on ε momentarily, but let us keep b_0 fixed for now. By (2.53), we can select n_1 so large that

$$P_i(|W_{n_1}| < b_0) \leq \varepsilon/8 \text{ for every } |i| < b_0,$$

and then we choose $b_1 > b_0$ so large that

$$P_i(|W_{n_1}| \geq b_1) \leq \varepsilon/8 \text{ for every } |i| < b_0.$$

We proceed inductively. Once $n_1, \ldots, n_k$ and $b_0, b_1, \ldots, b_k$ have been chosen, we first use (2.53) to select $n_{k+1} > n_k$ so large that

$$P_i(|W_{n_{k+1}-n_k}| < b_k) \leq 2^{-(k+3)}\varepsilon \text{ for every } |i| < b_k,$$

and then we select $b_{k+1} > b_k$ so large that

$$P_i(|W_{n_{k+1}-n_k}| \geq b_{k+1}) \leq 2^{-(k+3)}\varepsilon \text{ for every } |i| < b_k.$$

In order to convince ourselves that we have obtained a "large" set B in (2.52), we construct a probability measure ν on $\mathbb{Z}^{\mathbb{Z}}$, equivalent to m, by selecting strictly positive probabilities (p_i) on $\mathbb{Z}$, and defining, analogously to (2.49),

$$\nu(A) = \sum_{i \in \mathbb{Z}} p_i P_i\left(\text{a realization of the random walk is in } A\right). \tag{2.54}$$

Select b_0 so large that

$$\sum_{|i| \geq b_0} p_i \leq \varepsilon/2,$$

and note that

$$\nu(B^c) \le \varepsilon/2 + \sum_{|i|<b_0} p_i P_i \left(\bigcup_{k=1}^{\infty} \Big\{ |W_{n_{k-1}}| < b_{k-1},\ W_{n_k} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big\} \right)$$

(with $n_0 = 0$). We now use the Markov property to see that by construction, for every $k = 1, 2, \ldots,$

$$\begin{aligned} &P_i\Big(|W_{n_{k-1}}| < b_{k-1},\ W_{n_k} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big) \\ &= \sum_{|j|<b_{k-1}} P_i\big(|W_{n_{k-1}}| = j\big) P_j\Big(W_{n_k - n_{k-1}} \notin (-b_k, -b_{k-1}] \cup [b_{k-1}, b_k) \Big) \\ &\le 2 \cdot 2^{-(k+2)} \varepsilon = 2^{-(k+1)} \varepsilon \,, \end{aligned}$$

so that $\nu(B^c) \le \varepsilon$. Selecting $\varepsilon > 0$ small, we can hence obtain a "large" weakly wandering set B.

Notice that we have actually proved that the map ϕ is a null map. Indeed, if we had $m\big(\mathcal{P}(\phi)\big) > 0$, then we would also have $\nu\big(\mathcal{P}(\phi)\big) > 0$ for the probability measure ν in (2.54). This would imply that for $\varepsilon > 0$ small enough, $\nu\big(\mathcal{P}(\phi) \cap B\big) > 0$, hence also $m\big(\mathcal{P}(\phi) \cap B\big) > 0$. Since a subset of weakly wandering set is itself weakly wandering, we obtain a contradiction with Proposition 2.4.10.

2.5 Comments on Chapter 2

Comments on Section 2.1

There are many books on ergodic theory. The notions of ergodicity and mixing are widely used in probability. Among the sources that have been used in probability are Cornfeld et al. (1982) and Walters (1982).

Comments on Section 2.3

A very useful survey on strong mixing conditions is in Bradley (2005).

Comments on Section 2.4

Much of the material on the positive–null decomposition presented in this section is in Section 3.6 of Aaronson (1997) and Section 3.4 of Krengel (1985), and many of the results are due to U. Krengel and his coworkers, e.g., Krengel (1967) and Jones and Krengel (1974).

The phenomenon exhibited in Example 2.4.12 goes much further than the case of a simple random walk. For example, it is shown in Harris and Robbins (1953) that every recurrent real-valued Markov chain with an invariant measure generates similarly a shift-invariant measure on $\mathbb{R}^{\mathbb{Z}}$ with respect to which the left shift is conservative.

2.6 Exercises to Chapter 2

Exercise 2.6.1. *Let $(E, \mathcal{E}, m)$ be a σ-finite measure space, and $\phi : E \to E$ a measurable map. Show that the collection $\mathcal{I}$ of all ϕ-invariant sets $\mathcal{E}$ is a σ-field.*

Exercise 2.6.2. *Let $\phi : E \to E$ be onto and one-to-one. Check that ϕ is nonsingular if and only if its inverse ϕ^{-1} is nonsingular.*

Exercise 2.6.3. *(i) Prove that measurable sets satisfying either $m(A) = 0$ or $m(A^c) = 0$ are invariant sets for every nonsingular map ϕ.*

(ii) Suppose that ϕ is nonsingular. Prove that the σ-field of ϕ-invariant sets coincides with the σ-field of ϕ^{-1}-invariant sets. Does this remain true if ϕ is one-to-one and onto, but not necessarily nonsingular? Conclude that a nonsingular map is ergodic if and only if the inverse map ϕ^{-1} is ergodic.

Exercise 2.6.4. *(i) Let ϕ be a measurable map, and $f : E \to \mathbb{R}$ a function measurable with respect to the σ-field $\mathcal{I}$ of ϕ-invariant sets. Show that f is ϕ-invariant in the sense that $f(\phi(x)) = f(x)$ for m-almost every x.*

(ii) Suppose that ϕ is a nonsingular and ergodic map, and f is as in part (i). Show that f is constant in the sense that there is $a \in \mathbb{R}$ such that $f(x) = a$ for m-almost every x.

Exercise 2.6.5. *Recall that the conditional expectation can be defined for nonnegative random variables without a finite mean; see Billingsley (1995). Prove that the ergodic theorem in the form of (2.7) holds for a nonnegative measurable function f even if the integrability assumption (2.6) fails.*

Exercise 2.6.6. *Show that a map ϕ is mixing if and only if the inverse map ϕ^{-1} is mixing.*

Exercise 2.6.7. *Show by example that the sum of two independent stationary ergodic processes may not be ergodic. **Hint**: What is the stationary process corresponding to the measure in (2.3)? Show that on the other hand, the sum of two independent stationary mixing processes must be mixing.*

Exercise 2.6.8. *Weak convergence is not kind to various ergodic-theoretical notions.*

(i) Give an example of a family of nonergodic stationary processes that converge weakly to a mixing process (no need to work hard: a trivial example suffices).

(ii) Give an example of a family of mixing processes that converge weakly to a nonergodic process (an autoregressive process of order 1 with Gaussian innovations can provide an easy example).

Exercise 2.6.9. *Show that the presence or absence of strong mixing is determined by the finite-dimensional distributions of a stationary process.*

Exercise 2.6.10. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* A, B *two events in* $\mathcal{F}$ *such that* $P(A \cap B) = .2, P(A \setminus B) = .3$, $P(B \setminus A) = .4$. *Let* $\mathcal{A}$ *be the* σ*-field generated by the event A, and* $\mathcal{B}$ *the* σ*-field generated by the event B. Calculate* $\phi(\mathcal{A}, \mathcal{B})$ *and* $\phi(\mathcal{B}, \mathcal{A})$ *and check that they are different.*

Exercise 2.6.11. *Let* $g : \mathbb{R} \to \mathbb{R}$ *be a one-to-one function such that both g and* g^{-1} *are measurable. Show that a stationary process* $(X_n,\ n \in \mathbb{Z})$ *is* α*- (*β*-,* ϕ*-) mixing if and only if the process* $(g(X_n),\ n \in \mathbb{Z})$ *is* α*- (*β*-,* ϕ*-) mixing.*

Exercise 2.6.12. *Theorem 2.4.5 requires the function to be strictly positive m-a.e. We can relax this assumption somewhat. Let* $B \subset \mathcal{C}(\phi)$ *be such that* $m(B) > 0$. *Prove that*

$$\sum_{n=1}^{\infty} \mathbf{1}_B\big(\phi^n(x)\big) = \infty \ \ m\text{-a.e. on } B.$$

Exercise 2.6.13. *Let* ϕ *be a null map on (an automatically infinite)* σ*-finite measure space* $(E, \mathcal{E}, m)$, *and* $f \in L_1(m)$. *Prove that*

$$\frac{1}{n} \sum_{j=0}^{n-1} f\big(\phi^j(x)\big) \to 0 \ \ \text{as } n \to \infty$$

for m-almost every $x \in E$.

Exercise 2.6.14. *A nonnegative sequence* $(g_n,\ n = 1, 2, \ldots)$ *is called subadditive if* $g_{n+m} \le g_n + g_m$ *for all* $n, m \ge 1$. *Prove that for every subadditive sequence, the limit* $\lim_{n\to\infty} g_n/n$ *exists and is equal to* $\inf_{n\ge1} g_n/n$.

Chapter 3
Infinitely Divisible Processes

3.1 Infinitely Divisible Random Variables, Vectors, and Processes

Infinitely divisible stochastic processes form a broad family whose structure is reasonably well understood. A stochastic process $\big(X(t),\, t \in T\big)$ is said to be infinitely divisible if for every $n = 1, 2, \ldots$, there is a stochastic process $\big(Y(t),\, t \in T\big)$ such that

$$\big(X(t),\, t \in T\big) \stackrel{d}{=} \left(\sum_{j=1}^{n} Y_j(t),\, t \in T\right),$$

where $\big(Y_j(t),\, t \in T\big)$, $j = 1, \ldots, n$, are i.i.d. copies of $\big(Y(t),\, t \in T\big)$. It is also common to say that the law of a stochastic process (rather than the process itself) is infinitely divisible, and we will use both pieces of terminology interchangeably.

The basic ingredient of an infinitely divisible stochastic process is a one-dimensional infinitely divisible random variable, to which an infinitely divisible stochastic process reduces if the parameter space T is a singleton. The most powerful tool for studying infinitely divisible random variables is their *Lévy–Khinchine representation*. Specifically, a random variable X is infinitely divisible if and only if there is a uniquely determined triple (σ^2, μ, b) such that the characteristic function of X can be written in the form

$$Ee^{i\theta X} = \exp\left\{-\frac{1}{2}\theta^2\sigma^2 + \int_{-\infty}^{\infty}\Big(e^{i\theta x} - 1 - i\theta[\![x]\!]\Big)\,\mu(dx) + i\theta b\right\} \tag{3.1}$$

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_3

for $\theta \in \mathbb{R}$, where for $x \in \mathbb{R}$, its truncation $[\![x]\!]$ is defined by

$$[\![x]\!] = \begin{cases} x & \text{if } |x| \le 1, \\ -1 & \text{if } x < -1, \\ 1 & \text{if } x > 1. \end{cases} \tag{3.2}$$

The triple (σ^2, μ, b) is called the *characteristic triple* of the infinitely divisible random variable X, and it consists of $\sigma^2 \ge 0$, a measure μ on $\mathbb{R}$ satisfying $\mu\big(\{0\}\big) = 0$ and

$$\int_{\mathbb{R}} \big(1 \wedge x^2\big)\, \mu(dx) < \infty\,,$$

and finally, $b \in \mathbb{R}$. See Sato (1999).

The basic examples of infinitely divisible random variables are the normal random variable and the compound Poisson random variable of the following example.

Example 3.1.1. If N is a mean λ Poisson random variable independent of an i.i.d. sequence $Y_1, Y_2, \ldots$ with a common distribution F_Y, then

$$X = \sum_{i=1}^{N} Y_i \tag{3.3}$$

is an infinitely divisible random vector because its characteristic function is of the form (3.1) with $\sigma^2 = 0$, $\mu = \lambda F_Y$, and $b = \lambda E[\![Y_1]\!]$. An infinitely divisible random variable with a representation of the form (3.1.1) is said to be *compound Poisson*. Note that a compound Poisson random variable has a finite measure μ in its characteristic triple.

Other examples of one-dimensional infinitely divisible random variables include the gamma random variable and the geometric (and, more generally, the negative binomial) random variable, as can be verified by identifying, in each of these examples, the characteristic triple. See Exercise 3.8.1.

The entry σ^2 in the characteristic triple of an infinitely divisible random variable X is the variance of the Gaussian component of X; if $\sigma^2 = 0$, we say that the infinitely divisible random variable has no Gaussian component. The entry μ in the characteristic triple of an infinitely divisible random variable X is the *Lévy measure* of X; it describes the Poisson component of the infinitely divisible random variable; if $\mu = 0$, the infinitely divisible random variable has no Poisson component and hence is a Gaussian random variable. The entry b in the characteristic triple of an infinitely divisible random variable X is sometimes referred to as the shift component of X; this, however, has to be taken with a grain of salt, since b interacts with the truncation $[\![\cdot]\!]$ in the representation (3.1) of the

characteristic function. In certain special cases, the characteristic function of an infinitely divisible random variable has slightly different representations, in which the role of a "shift" parameter is clearer.

Example 3.1.2. A Lévy motion

A Lévy motion (also known as a Lévy process) is continuous in a probability stochastic process with stationary and independent increments, $\big(X(t),\, t \in \mathbb{R}\big)$, such that $X(0) = 0$ a.s. Note that it is also common to define a Lévy process only on the positive half-line, as $\big(X(t),\, t \geq 0\big)$, satisfying the same requirements. The definition on the entire real line is more natural and convenient for our purposes. The finite-dimensional distributions of a Lévy process are completely determined by its one-dimensional marginal distribution at time 1, which is necessarily infinitely divisible (in one dimension). In fact, there is a one-to-one correspondence between the laws of Lévy processes and the laws of one-dimensional infinitely divisible random variables that are the values of the Lévy processes at time 1. See Sato (1999).

It is easy to see that a Lévy process is an infinitely divisible stochastic process. Indeed, let $n = 1, 2, \ldots$. Since $X = X(1)$ is a one-dimensional infinitely divisible random variable, there is an infinitely divisible random variable Y such that $X \stackrel{d}{=} Y_1 + \ldots + Y_n$, where $Y_1, \ldots, Y_n$ are i.i.d. copies of Y. Let $\big(Y(t),\, t \in \mathbb{R}\big)$ be a Lévy process such that $Y(1) \stackrel{d}{=} Y$, and let $\big(Y_j(t),\, t \in \mathbb{R}\big)$, $j = 1, \ldots, n$, be i.i.d. copies of $\big(Y(t),\, t \in \mathbb{R}\big)$. Then the stochastic process $\big(\sum_{j=1}^{d} Y_j(t),\, t \in \mathbb{R}\big)$ is continuous in probability, has stationary and independent increments, and vanishes at time 0. Therefore, it is a Lévy process. Since

$$\sum_{j=1}^{d} Y_j(1) \stackrel{d}{=} \sum_{j=1}^{d} Y_j \stackrel{d}{=} X \stackrel{d}{=} X(1)$$

by construction, we conclude that

$$\left(\sum_{j=1}^{d} Y_j(t),\, t \in \mathbb{R}\right) \stackrel{d}{=} \big(X(t),\, t \in \mathbb{R}\big)\,.$$

Therefore, $\big(X(t),\, t \in \mathbb{R}\big)$ is an infinitely divisible stochastic process.

A Brownian motion is a particular Lévy process for which $X(1)$ has a normal distribution.

An infinitely divisible stochastic process corresponding to a finite parameter set T is a (finite-dimensional) infinitely divisible random vector. The distribution of an infinitely divisible random vector, say $\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})$, is once again uniquely determined by a characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$, where this time, Σ is a $d \times d$ nonnegative definite matrix, and μ a measure on $\mathbb{R}^d$ such that $\mu\big(\{\mathbf{0}\}\big) = 0$ and

$$\int_{\mathbb{R}^d} \big(1 \wedge \|\mathbf{x}\|^2\big)\, \mu(d\mathbf{x}) < \infty\,.$$

Finally, $\mathbf{b} \in \mathbb{R}^d$. The characteristic function of an infinitely divisible random vector, say $\mathbf{X}$, with a characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$ is given by

$$Ee^{i(\boldsymbol{\theta},\mathbf{X})} = \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T\Sigma\boldsymbol{\theta} + \int_{\mathbb{R}^d}\Big(e^{i(\boldsymbol{\theta},\mathbf{x})} - 1 - i\big(\boldsymbol{\theta}, [\![\mathbf{x}]\!]\big)\Big)\,\mu(d\mathbf{x}) + i(\boldsymbol{\theta},\mathbf{b})\right\} \tag{3.4}$$

for $\boldsymbol{\theta} \in \mathbb{R}^d$, where the truncation of a vector is defined componentwise: for $\mathbf{x} = \big(x^{(1)}, \ldots, x^{(d)}\big)$,

$$[\![\mathbf{x}]\!] = \big([\![x^{(1)}]\!], \ldots, [\![x^{(d)}]\!]\big)\,.$$

The role of the entries in the characteristic triple of an infinitely divisible random vector $\mathbf{X}$ is similar to their role in the one-dimensional case. The entry Σ is the covariance matrix of the Gaussian component of $\mathbf{X}$, the measure μ describes the Poisson component of $\mathbf{X}$, and the vector $\mathbf{b}$ can be (imprecisely) thought of as the shift vector of $\mathbf{X}$.

The following characterization of an infinitely divisible stochastic process is almost immediate; see Exercise 3.8.5.

Proposition 3.1.3. *A stochastic process $\big(X(t),\, t \in T\big)$ is infinitely divisible if and only if all of its finite-dimensional distributions are infinitely divisible.*

The most useful description of infinitely divisible processes is through their own characteristic triples. Such triples are similar in nature to the characteristic triples of one-dimensional infinitely divisible random variables and of infinitely divisible random vectors, but this time, they "live on the appropriate function spaces." In the description below, we follow Rosiński (2007).

For a parameter space T, let $\mathbb{R}^T$ be the collection of all real-valued functions on T. We transform $\mathbb{R}^T$ into a measurable space by endowing it with the cylindrical σ-field. A measure μ on $\mathbb{R}^T$ is said to be a Lévy measure if the two conditions stated below hold. The first condition is as follows:

Condition 3.1.4.

$$\int_{\mathbb{R}^T} [\![x]\!](t)^2\,\mu(d\mathbf{x}) < \infty \quad \text{for every } t \in T. \tag{3.5}$$

As in the finite-dimensional case, the truncation of a function in (3.5) is understood componentwise: for $\mathbf{x} = \big(x(t),\, t \in T\big) \in \mathbb{R}^T$, $[\![\mathbf{x}]\!] \in \mathbb{R}^T$ is defined by

$$[\![x]\!](t) = [\![x(t)]\!],\ t \in T\,.$$

The second condition will be stated separately in two cases: when the parameter space T is countable (or finite), and when it is uncountable.

Condition 3.1.5 (Countable space T).

$$\mu\Big(\mathbf{x} \in \mathbb{R}^T :\, x(t) = 0 \ \text{ for all } t \in T\Big) = 0\,. \tag{3.6}$$

Condition 3.1.6 (Uncountable space T). For every countable subset T_1 of T, such that

$$\mu\Big(\mathbf{x} \in \mathbb{R}^T : x(t) = 0 \text{ for all } t \in T_1\Big) > 0, \tag{3.7}$$

there is $t_0 \in T_1^c$ such that

$$\mu\Big(\mathbf{x} \in \mathbb{R}^T : x(t) = 0 \text{ for all } t \in T_1,\ x(t_0) \neq 0\Big) > 0\,. \tag{3.8}$$

To state the theorem characterizing infinitely divisible stochastic processes, we introduce the following notation: let

$$\mathbb{R}^{(T)} = \Big\{\mathbf{x} \in \mathbb{R}^T : x(t) = 0 \text{ for all but finitely many } t \in T\Big\}$$

be the collection of functions on T vanishing outside of a finite set. Note the obvious fact that if $\mathbf{x} \in \mathbb{R}^{(T)}$ and $\mathbf{y} \in \mathbb{R}^T$, then the sum $\sum_{t\in T} x(t)y(t)$ makes perfect sense because the (potentially uncountable) number of terms in the sum is really only finite, corresponding to the set of nonzero coordinates of $\mathbf{x}$. Similarly, if $\mathbf{a} \in \mathbb{R}^{T\times T}$, then the double sum $\sum_{s\in T}\sum_{t\in T} a(s,t)x(s)x(t)$ makes perfect sense as well.

Theorem 3.1.7. *A stochastic process $\big(X(t),\ t \in T\big)$ is infinitely divisible if and only if there exists a uniquely determined triple $\big(\Sigma, \mu, \mathbf{b}\big)$ such that for every $\boldsymbol{\theta} \in \mathbb{R}^{(T)}$,*

$$E\exp\left\{i\sum_{t\in T}\theta(t)X(t)\right\} \tag{3.9}$$

$$= \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T\Sigma\boldsymbol{\theta} + \int_{\mathbb{R}^T}\Big(e^{i(\boldsymbol{\theta},\mathbf{x})} - 1 - i\big(\boldsymbol{\theta}, [\![\mathbf{x}]\!]\big)\Big)\,\mu(d\mathbf{x}) + i(\boldsymbol{\theta},\mathbf{b})\right\}$$

with the notation

$$\boldsymbol{\theta}^T\Sigma\boldsymbol{\theta} = \sum_{s\in T}\sum_{t\in T}\Sigma(s,t)\theta(s)\theta(t),\quad (\boldsymbol{\theta},\mathbf{y}) = \sum_{t\in T}\theta(t)y(t),\ \mathbf{y}\in\mathbb{R}^T\,.$$

In (3.9):

- $\Sigma = \big(\Sigma(s,t),\ s,t \in T\big)$ *is a nonnegative definite function on T;*
- μ *is a Lévy measure on T;*
- $\mathbf{b} \in \mathbb{R}^T$.

See Rosiński (2007) for a proof. The sufficiency part of Theorem 3.1.7 is clear: if a stochastic process $\big(X(t),\ t \in T\big)$ satisfies (3.9) for some triple $\big(\Sigma, \mu, \mathbf{b}\big)$, as in the theorem, then for every finite subset $\{t_1, \ldots, t_d\}$ of T, the d-dimensional random vector $\big(X_(t_1), \ldots, X(t_d)\big)$ has the characteristic function

$$E\exp\left\{\sum_{j=1,\dots,d}\theta_j X(t_j)\right\} = \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T\Sigma_{t_1,\dots,t_d}\boldsymbol{\theta}\right.$$

$$\left.+\int_{\mathbb{R}^{\{t_1,\dots,t_d\}}}\left(e^{i(\boldsymbol{\theta},\mathbf{x})}-1-i\big(\boldsymbol{\theta},[\![\mathbf{x}]\!]\big)\right)\mu_{t_1,\dots,t_d}(d\mathbf{x})+i(\boldsymbol{\theta},\mathbf{b}_{t_1,\dots,t_d})\right\}$$

for $\boldsymbol{\theta} = (\theta_1,\dots,\theta_d) \in \mathbb{R}^d$, with $\Sigma_{t_1,\dots,t_d}$ the restriction of Σ from $T\times T$ to $\{t_1,\dots,t_d\}\times\{t_1,\dots,t_d\}$, $\mathbf{b}_{t_1,\dots,t_d}$ the restriction of $\mathbf{b}$ from T to $\{t_1,\dots,t_d\}$, and $\mu_{t_1,\dots,t_d}$ a measure on $\mathbb{R}^{\{t_1,\dots,t_d\}}$ defined by

$$\mu_{t_1,\dots,t_d}(B) = \mu\Big(\mathbf{x}\in\mathbb{R}^T:\ \big(x(t_1),\dots,x(t_d)\big)\in B\setminus\{0\}\Big)$$

for Borel sets $B\subset\mathbb{R}^{\{t_1,\dots,t_d\}}$. Therefore, $\big(X(t_1),\dots,X(t_d)\big)$ is a d-dimensional infinitely divisible random vector, and so all finite-dimensional distributions of the process $\big(X(t),\ t\in T\big)$ are infinitely divisible. By Proposition 3.1.3, one concludes that $\big(X(t),\ t\in T\big)$ is an infinitely divisible process.

As in the finite-dimensional case, the triple $\big(\Sigma,\mu,\mathbf{b}\big)$ appearing in Theorem 3.1.7 is referred to as the characteristic triple of the infinitely divisible process, and the representation (3.9) its Lévy–Khinchine representation.

Remark 3.1.8. Conditions 3.1.5 and 3.1.6 on a Lévy measure on T ensure that the measure does not put "redundant" weight on functions with "many" zero coordinates. This guarantees (by Theorem 3.1.7) that an infinitely divisible stochastic process has a unique Lévy measure.

Let, on the other hand, $\big(X(t),\ t\in T\big)$ be an infinitely divisible stochastic process and suppose that $\big(\Sigma,\mu,\mathbf{b}\big)$ is a triple as in Theorem 3.1.7, except that the measure μ is assumed to satisfy only Condition 3.1.4 but not necessarily conditions 3.1.5 and 3.1.6. In such a case we will say that μ is a *weak Lévy measure* of the process $\big(X(t),\ t\in T\big)$, and $\big(\Sigma,\mu,\mathbf{b}\big)$ is a *weak characteristic triple*.

"Redundant weight on zeros" of a weak Lévy measure does not play any role in the representation (3.9) of the characteristic function of the process. For example, suppose that T is uncountable and there exists a countable set T_1 such that (3.7) holds but (3.8) fails for every $t_0\in T_1^c$. We can then remove some "redundant zeros" by defining a measure μ_1 on $\mathbb{R}^T$ by

$$\mu_1(A) = \mu\Big(A\cap\Big\{\mathbf{x}\in\mathbb{R}^T:\ x(t)\neq 0\ \text{ for some } t\in T_1\Big\}\Big),\quad A \text{ a cylindrical set.}$$

The triple $\big(\Sigma,\mu_1,\mathbf{b}\big)$ clearly still satisfies (3.9). Therefore, a weak Lévy measure is not unique.

In some cases, we will find it convenient to work with weak Lévy measures of infinitely divisible stochastic processes. Apart from lack of uniqueness, their role in describing the structure of infinitely divisible processes is identical to that of "true" Lévy measures.

Example 3.1.9. A Gaussian process is a stochastic process $\big(X(t),\, t \in T\big)$ whose finite-dimensional distributions are multivariate normal distributions. These are infinitely divisible stochastic processes with characteristic triples in which the Lévy measure μ vanishes. That is, the characteristic function of a Gaussian process can be written in the form

$$E \exp\left\{\sum_{t\in T} \theta(t)X(t)\right\} = \exp\left\{-\frac{1}{2}\boldsymbol{\theta}^T \Sigma \boldsymbol{\theta} + i(\boldsymbol{\theta}, \mathbf{b})\right\} \tag{3.10}$$

for $\boldsymbol{\theta} \in \mathbb{R}^{(T)}$. Then Σ is the covariance function of the process, and $\mathbf{b}$ is its mean function.

The following proposition lists elementary properties of the characteristic triples of infinitely divisible stochastic processes. We leave the proof for Exercise 3.8.10.

Proposition 3.1.10. *(i) If $\big(X(t),\, t \in T\big)$ is an infinitely divisible stochastic process with a characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$ and $a \in \mathbb{R} \setminus \{0\}$, then $\big(aX(t),\, t \in T\big)$ is an infinitely divisible stochastic process with the characteristic triple $\big(\Sigma_a, \mu_a, \mathbf{b}_a\big)$, where*

$$\Sigma_a = a^2\Sigma, \quad \mu_a(\cdot) = \mu\big(a^{-1}\cdot\big), \quad \mathbf{b}_a = a\mathbf{b} + \int_{\mathbb{R}^T} \big(a[\![\mathbf{x}]\!] - [\![a\mathbf{x}]\!]\big)\, \mu(dx)\,.$$

(ii) If $\big(X_1(t),\, t \in T\big)$ and $\big(X_2(t),\, t \in T\big)$ are independent infinitely divisible stochastic processes with respective characteristic triples $\big(\Sigma_i, \mu_i, \mathbf{b}_i\big)$, $i = 1, 2$, then $\big(X_1(t) + X_2(t),\, t \in T\big)$ is an infinitely divisible stochastic process with the characteristic triple $\big(\Sigma_1 + \Sigma_2,\, \mu_1 + \mu_2,\, \mathbf{b}_1 + \mathbf{b}_2\big)$.

Example 3.1.11. A stochastic process $\big(X(t),\, t \in T\big)$ is called α-stable if it has the following property: for every $n \geq 1$, there is a (nonrandom) function $(c_n(t),\, t \in T)$ such that

$$\big(X(t),\, t \in T\big) \stackrel{d}{=} \left(n^{-1/\alpha} \sum_{j=1}^{n} X_j(t) + c_n(t),\, t \in T\right), \tag{3.11}$$

where $\big(X_j(t),\, t \in T\big)$, $j = 1, \ldots, n$, are i.i.d. copies of $\big(X(t),\, t \in T\big)$. It is called *strictly α-stable* if $c_n(t) \equiv 0$ for every $n \geq 1$. It is called *symmetric α-stable* (often abbreviated to SαS) if it is α-stable and symmetric. It is clear that a symmetric α-stable process is also strictly α-stable.

The definition immediately implies that every α-stable process is infinitely divisible. If $\big(\Sigma, \mu, \mathbf{b}\big)$ is the characteristic triple of the process $\big(X(t),\, t \in T\big)$, then, by Proposition 3.1.10, the stochastic process on the right-hand side of (3.11) is infinitely divisible with the characteristic triple $\big(n^{1-2/\alpha}\Sigma,\, n\mu(n^{1/\alpha}\cdot), \mathbf{b}_n\big)$ for

some function $\mathbf{b}_n \in \mathbb{R}^T$. It follows from Theorem 3.1.7 that a stochastic process $\big(X(t),\, t \in T\big)$ is α-stable if and only if it is infinitely divisible and its characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$ satisfies

$$\Sigma = n^{1-2/\alpha}\Sigma \;\text{ and }\; \mu = n\mu(n^{1/\alpha}\cdot) \tag{3.12}$$

for every $n = 1, 2, \ldots$.

It follows immediately from (3.12) that $\Sigma = 0$ unless $\alpha = 2$. Further, suppose that $\mu \neq 0$. Then there exists $t_0 \in T$ such that

$$h(a) = \mu\Big(\mathbf{x} \in \mathbb{R}^T :\, |x(t_0)| > a\Big) > 0$$

for some $a > 0$. Clearly, h is a nonincreasing function. Since it follows from (3.12) that $h\big(an^{1/\alpha}\big) = n^{-1}h(a)$, we conclude that necessarily $\alpha > 0$, and further, this relation means that $\lim_{a\to\infty} a^\alpha h(a) = 1$. Since

$$\int_{\mathbb{R}^T} [\![\mathbf{x}]\!](t_0)^2\, \mu(d\mathbf{x}) \geq \int_0^1 yh(y)\, dy\,,$$

Condition 3.1.4 on a Lévy measure implies that $\alpha < 2$.

That is, a nondeterministic α-stable process can exist only for $0 < \alpha \leq 2$. If α=2, then the Lévy measure of the process vanishes, and the process is the Gaussian process of Example 3.1.9.

If $0 < \alpha < 2$, then the function Σ in the characteristic triple of the process must vanish, while (3.12) implies that the Lévy measure of the process must have the scaling property

$$\mu(c\cdot) = c^{-\alpha}\mu \;\text{ for each } c > 0. \tag{3.13}$$

See Exercise 3.8.11.

Conversely, every infinitely divisible process with $\Sigma = 0$ and Lévy measure that scales as in (3.13) is α-stable. One usually constructs α-stable processes as stochastic integrals as in Section 3.3.

We summarize the discussion of this section by recording the following immediate but important corollary of Theorem 3.1.7.

Corollary 3.1.12. *A stochastic process $\big(X(t),\, t \in T\big)$ is infinitely divisible if and only if it has a unique decomposition in law*

$$\big(X(t),\, t \in T\big) \stackrel{d}{=} \big(G(t),\, t \in T\big) + \big(Y(t),\, t \in T\big)\,,$$

where $\big(G(t),\, t \in T\big)$ and $\big(Y(t),\, t \in T\big)$ are independent stochastic processes, with $\big(G(t),\, t \in T\big)$ a centered Gaussian process and $\big(Y(t),\, t \in T\big)$ has a characteristic function of the form

$$E\exp\left\{\sum_{t\in T}\theta(t)Y(t)\right\} = \exp\left\{\int_{\mathbb{R}^T}\left(e^{i(\boldsymbol{\theta},\mathbf{x})}-1-i\big(\boldsymbol{\theta},[\![\mathbf{x}]\!]\big)\right)\mu(d\mathbf{x})+i(\boldsymbol{\theta},\mathbf{b})\right\}$$

for $\boldsymbol{\theta}\in\mathbb{R}^{(T)}$*, where* μ *is a Lévy measure on* T*, and* $\mathbf{b}$ *is a real-valued function on* T*.*

If $\big(X(t),\, t\in T\big)$ has a characteristic function given by (3.9), then Σ is the covariance function of the Gaussian process $\big(G(t),\, t\in T\big)$ (the Gaussian component of $\big(X(t),\, t\in T\big)$), and the rest of the characteristic triple $\big(\Sigma,\mu,\mathbf{b}\big)$ determines the law of the process $\big(Y(t),\, t\in T\big)$. The latter process is an infinitely divisible process without a Gaussian component (the Poisson component of $\big(X(t),\, t\in T\big)$ plus a deterministic "shift").

3.2 Infinitely Divisible Random Measures

An infinitely divisible random measure is the single most important infinitely divisible stochastic process, and it often serves as the basic ingredient in constructing other infinitely divisible stochastic processes. In order to define an infinitely divisible random measure, we start with a measurable space $(S,\mathcal{S})$. We need three measures as ingredients:

- a σ-finite measure γ on S;
- a measure ν on $S\times\big(\mathbb{R}\setminus\{0\}\big)$ such that the measure

$$m_0(B) := \int_B\int_{\mathbb{R}\setminus\{0\}}[\![x]\!]^2\,\nu(ds,dx),\ B\in\mathcal{S} \tag{3.14}$$

  is σ-finite;
- a σ-finite signed measure β on S (see Section 10.3).

Let $\mathcal{S}_0$ denote the collection of sets B in $\mathcal{S}$ satisfying

$$m(B) := \gamma(B)+\|\beta\|(B)+m_0(B)<\infty\,. \tag{3.15}$$

For sets $B_1,B_2\in\mathcal{S}_0$, define

$$\Sigma\big(B_1,B_2\big)=\gamma\big(B_1\cap B_2\big)\,. \tag{3.16}$$

Note that for every $d=1,2,\ldots,a_1,\ldots,a_d\in\mathbb{R}$, and $B_1,\ldots,B_d\in\mathcal{S}_0$,

$$\sum_{j=1}^d\sum_{k=1}^d a_ja_k\Sigma\big(B_j,B_k\big)=\int_S\left(\sum_{j=1}^d a_j\mathbf{1}\big(s\in B_j\big)\right)^2\gamma(ds)\ge 0\,.$$

Therefore, Σ is a nonnegative definite function on $\mathcal{S}_0$.

Next, let $\Phi : S \times \big(\mathbb{R} \setminus \{0\}\big) \to \mathbb{R}^{\mathcal{S}_0}$ be a map defined by

$$\Phi(s,x)(B) = x\mathbf{1}(s \in B),\ B \in \mathcal{S}_0$$

for $s \in S$ and $x \in \mathbb{R} \setminus \{0\}$. Clearly, Φ is a measurable map, and it defines a measure on $\mathbb{R}^{\mathcal{S}_0}$ by

$$\mu = \nu \circ \Phi^{-1}\,. \tag{3.17}$$

We claim that μ is a Lévy measure on $\mathcal{S}_0$. Note, first of all, that for every $B \in \mathcal{S}_0$,

$$\int_{\mathbb{R}^{\mathcal{S}_0}} [\![\mathbf{x}]\!](B)^2\, \mu(d\mathbf{x}) = \int_B \int_{\mathbb{R}\setminus\{0\}} [\![x]\!]^2\, \nu(ds,dx) < \infty$$

by the definition (3.15) of the collection $\mathcal{S}_0$. Therefore, the measure μ satisfies Condition 3.1.4. If the collection $\mathcal{S}_0$ is countable, we also need to check Condition 3.1.5. Since the measure m in (3.15) is σ-finite, the countable collection $\mathcal{S}_0$ covers S. Then

$$\mu\Big(\mathbf{x} \in \mathbb{R}^{\mathcal{S}_0} :\ x(B) = 0 \ \text{ for all } B \in \mathcal{S}_0\Big) = \nu\Big((s,x) :\ s \notin \cup_{B\in\mathcal{S}_0} B\Big) = \nu\big(\emptyset\big) = 0\,,$$

and so Condition 3.1.5 holds. If the collection $\mathcal{S}_0$ is uncountable, we need to check Condition 3.1.6. Suppose that $B_j \in \mathcal{S}_0$, $j = 1, 2, \ldots$, are such that

$$\mu\Big(\mathbf{x} \in \mathbb{R}^{\mathcal{S}_0} :\ x(B_j) = 0 \ \text{ for all } j = 1, 2, \ldots\Big) > 0\,.$$

Then

$$\nu\Big((s,x) :\ s \notin \cup_{j=1,2,\ldots} B_j\Big) > 0\,,$$

which is equivalent to saying that for the measure m_0 in (3.14),

$$m_0\Big(\big(\cup_{j=1,2,\ldots} B_j\big)^c\Big) > 0\,.$$

Since the measures m_0 and m are σ-finite, there is a set B in $\mathcal{S}_0$ such that $B \subset \big(\cup_{j=1,2,\ldots} B_j\big)^c$ and $m_0(B) > 0$. Then also

$$\nu\Big((s,x) :\ s \notin \cup_{j=1,2,\ldots} B_j,\ s \in B\Big) > 0\,,$$

which means that

$$\mu\Big(\mathbf{x} \in \mathbb{R}^{\mathcal{S}_0} :\ x(B_j) = 0 \ \text{ for all } j = 1, 2, \ldots,\ x(B) > 0\Big) > 0\,.$$

Therefore, Condition 3.1.6 holds, and the measure μ in (3.17) is a Lévy measure on $\mathcal{S}_0$.

Finally, define $\mathbf{b} \in R^{\mathcal{S}_0}$ by

$$b(B) = \beta(B),\ B \in \mathcal{S}_0\,. \tag{3.18}$$

The infinitely divisible stochastic process $M = \big(M(B), B \in \mathcal{S}_0\big)$ with the characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$ defined by (3.16), (3.17) and (3.18) is called an *infinitely divisible random measure on* $(S, \mathcal{S})$ with Gaussian variance measure γ, Lévy measure ν, and shift measure β.

The basic properties of an infinitely divisible random measure are described in the following proposition.

Proposition 3.2.1. *(i) For every $B \in \mathcal{S}_0$, $M(B)$ is an infinitely divisible random variable with a characteristic triplet (σ_B^2, μ_B, b_B), where*

$$\sigma_B^2 = \gamma(B),\ \mu_B(\cdot) = \nu(B \times \cdot),\ \beta_B = \beta(B)\,.$$

(ii) An infinitely divisible random measure is independently scattered. That is,

$$M(B_1), \ldots, M(B_d) \text{ are independent for every collection of disjoint sets } B_1, \ldots, B_d \in \mathcal{S}_0. \tag{3.19}$$

(iii) An infinitely divisible random measure is σ-additive. That is, for each choice of disjoint sets $(B_j) \subset \mathcal{S}_0$ such that $\cup_j B_j \in \mathcal{S}_0$,

$$M\Big(\cup_j B_j\Big) = \sum_j M(B_j) \ \text{ a.s.} \tag{3.20}$$

(note that the exceptional set in (3.20) depends, in general, on the choice of $(B_j) \subset \mathcal{S}_0$).

Proof. Part (i) follows from the construction of the stochastic process $M = \big(M(B), B \in \mathcal{S}_0\big)$ and Theorem 3.1.7. For part (ii), we use the fact that the components of an infinitely divisible random vector are independent if and only if every pair of the components is independent (see Sato (1999)). Therefore, it is enough to prove part (i) in the case $d = 2$, and then the statement follows by noticing that for disjoint sets $B_1, B_2 \in \mathcal{S}_0$, the Gaussian component of the infinitely divisible random vector $\big(N(B_1), N(B_2)\big)$ has vanishing covariance, while the bivariate Lévy measure is concentrated on the axes in the plane. This implies the independence of $N(B_1)$ and $N(B_2)$; see Sato (1999).

For part (iii) of the proposition, it is enough to prove that

$$\Delta_n := \sum_{j=1}^{n} M(B_j) - M\Big(\cup_{j=1}^{\infty} B_j\Big) \to 0 \tag{3.21}$$

in probability as $n \to \infty$, which will, in turn, follow if we check that the characteristic function of the difference Δ_n on the left-hand side converges to the constant function. By (3.9), for $\theta \in \mathbb{R}$,

$$Ee^{i\theta\Delta_n} = \exp\left\{-\frac{1}{2}\theta^2\sigma_n^2 + I_n(\theta) + i\theta b_n\right\} ,$$

where by (3.16),

$$\sigma_n^2 = \sum_{j=1}^{n} \gamma(B_j) + \gamma\left(\cup_{j=1}^{\infty} B_j\right) - 2\sum_{j=1}^{n} \gamma\left[B_j \cap \left(\cup_{k=1}^{\infty} B_k\right)\right] = \gamma\left(\cup_{j=n+1}^{\infty} B_j\right) \to 0$$

as $n \to \infty$ because $\cup_j B_j \in \mathcal{S}_0$. Further, by (3.18),

$$b_n = \sum_{j=1}^{n} \beta(B_j) - \beta\left(\cup_{j=1}^{\infty} B_j\right) = -\beta\left(\cup_{j=n+1}^{\infty} B_j\right) \to 0$$

as $n \to \infty$, also because $\cup_j B_j \in \mathcal{S}_0$. Finally, by (3.17),

$$I_n(\theta) = \int_{\cup_{j=n+1}^{\infty} B_j} \int_{\mathbb{R}\setminus\{0\}} \left(e^{-i\theta x} - 1 + i\theta[\![x]\!]\right) \nu(ds, dx) .$$

Since

$$\left|e^{i\theta x} - 1 - i\theta[\![x]\!]\right| \le K(\theta)[\![x]\!]^2$$

for some $K(\theta) \in (0, \infty)$, we conclude that

$$\left|I_n(\theta)\right| \le K(\theta) \int_{\cup_{j=n+1}^{\infty} B_j} \int_{\mathbb{R}\setminus\{0\}} [\![x]\!]^2 \, \nu(ds, dx) \to 0$$

as $n \to \infty$, once again because $\cup_j B_j \in \mathcal{S}_0$, whence the σ-additivity of the random measure. □

The Lévy measure of an infinitely divisible random measure can also be represented in a disintegrated form, given in the following proposition. This representation has the advantage of providing additional intuition into the structure of infinitely divisible random measures.

Proposition 3.2.2. *There exists a family of one-dimensional Lévy measures* $\big(\rho(s,\cdot),\ s \in S\big)$ *that is measurable in the sense that for every Borel set* $A \in \mathbb{R} \setminus \{0\}$, *the function* $\big(\rho(s,A), s \in S\big)$ *is measurable, and such that for each such* A *and* $B \in \mathcal{S}_0$,

$$\nu\big(B \times A\big) = \int_B \rho(s,A)\, m(ds) . \tag{3.22}$$

The family $\big(\rho(s,\cdot), s \in S\big)$ *is unique in the sense that if* $\big(\tilde{\rho}(s,\cdot), s \in S\big)$ *is another measurable family satisfying* (3.22)*, then* $\tilde{\rho}(s,\cdot) = \rho(s,\cdot)$ *for all* $s \in S$ *apart from a set of m-measure zero.*

Proof. Let (S_n) be disjoint sets in $\mathcal{S}_0$ such that $m_0(S_n) > 0$ for each n and

$$m_0\Big(\big(\cup_n S_n\big)^c\Big) = 0\,. \tag{3.23}$$

We will show that for each n, there is a measurable family of one-dimensional finite measures $\big(\rho_1(s,\cdot), s \in S_n\big)$ such that for every Borel set $A \in \mathbb{R} \setminus \{0\}$ and measurable subset B_n of S_n,

$$\int_{B_n}\int_A [\![x]\!]^2 \nu(ds,dx) = \int_{B_n} \rho_1(s,A)\, m_0(ds)\,. \tag{3.24}$$

Then for each $s \in S_n$,

$$\frac{d\rho_2(s,\cdot)}{d\rho_1(s,\cdot)}(x) = [\![x]\!]^{-2},\ x \in \mathbb{R} \setminus \{0\}$$

defines a σ-finite measure on $\mathbb{R} \setminus \{0\}$, which is a one-dimensional Lévy measure because $\rho_1(s,\cdot)$ is a finite measure. Combining these one-dimensional Lévy measures into a single family $\big(\rho_2(s,\cdot), s \in S\big)$ produces a measurable family of one-dimensional Lévy measures such that for every Borel set $A \in \mathbb{R} \setminus \{0\}$ and $B \in \mathcal{S}_0$,

$$\int_B \rho_2(s,A)\, m_0(ds) = \sum_n \int_{B\cap S_n} \rho_2(s,A)\, m_0(ds)$$

$$= \sum_n \left(\int_{B\cap S_n}\int_A [\![x]\!]^{-2}\rho_1(s,dx)\right) m_0(ds) = \sum_n \nu\big((B\cap S_n)\times A\big) = \nu(B\times A)$$

by (3.23), and now (3.22) follows by setting

$$\rho(s,\cdot) = \rho_2(s,\cdot)\frac{dm_0}{dm}(s)\ \ s \in S\,.$$

We now prove (3.24). By scaling if necessary, we may assume that $m_0(S_n) = 1$. Consider the probability space $\big(S_n \times (\mathbb{R}\setminus\{0\}), P_n\big)$, where we are using the product σ-field on $S_n \times (\mathbb{R}\setminus\{0\})$, and $P_n(ds,dx) = [\![x]\!]^2\nu(ds,dx)$. On this probability space, consider the random variable $X = X(s,x) = x$, $(s,x) \in S_n \times (\mathbb{R}\setminus\{0\})$, and the σ-field $\mathcal{G} = \big\{B \times (\mathbb{R}\setminus\{0\}), B \text{ a Borel subset of } S_n\big\}$. Let $\rho_0\big(A,\ (s,x)\big)$, $A \in \mathbb{R}\setminus\{0\}$, $(s,x) \in S_n \times (\mathbb{R}\setminus\{0\})$ be a conditional distribution of X given $\mathcal{G}$ (see Section 33 in Billingsley (1995)). Since for each A, $\rho_0\big(A,\ (s,x)\big)$, considered as a function on the probability space, is $\mathcal{G}$ measurable, it must be a function of s alone. Setting $\rho_1(A,s) = \rho_0\big(A,\ (s,x)\big)$ produces a required family $\big(\rho_1(s,\cdot), s \in S_n\big)$.

It remains to prove uniqueness. Let ρ and $\tilde{\rho}$ be two measurable families satisfying (3.22). Let (A_n) be an enumeration of intervals with rational endpoints that are subsets of $\mathbb{R} \setminus \{0\}$. For a fixed n, and A_n in place of A in (3.22), the latter relation says that both $\big(\rho(s, A_n),\ s \in S\big)$ and $\big(\tilde{\rho}(s, A_n),\ s \in S\big)$ are the Radon–Nikodym derivatives of the σ-finite measure $\nu(\cdot \times A_n)$ on S with respect to the σ-finite measure m. Therefore, they coincide outside of a set of m-measure zero. The union of these null sets taken over all n is a null set, and for s outside of this null set one has $\tilde{\rho}(s, A_n) = \rho(s, A_n)$ for each n. Since the sets A_n form a π-system generating the Borel σ-field on $\mathbb{R} \setminus \{0\}$, we conclude that $\tilde{\rho}(s, \cdot) = \rho(s, \cdot)$ for s outside of the above null set. □

It is clear from the definition of the measure m in (3.15) that both $\gamma \ll m$ and $\beta \ll m$, so we can define

$$\sigma(s) = \left(\frac{d\gamma}{dm}(s)\right)^{1/2} \tag{3.25}$$

and

$$b(s) = \frac{d\beta}{dm}(s) \tag{3.26}$$

for $s \in S$. The representations (3.22), (3.25), and (3.26) lead to a different way of describing an infinitely divisible random measure. Specifically, we will say that the corresponding infinitely divisible stochastic process $M = \big(M(B),\ B \in \mathcal{S}_0\big)$ with the characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$ defined by (3.16), (3.17) and (3.18) is an infinitely divisible random measure with control measure m, local Gaussian variance $(\sigma^2(s),\ s \in S)$, local Lévy measures $\big(\rho(s, \cdot), s \in S\big)$, and local shifts $(b(s),\ s \in S)$. This terminology is intuitively appealing if one thinks of the *local characteristics* $\big(\sigma^2(s), \rho(s, \cdot), b(s)\big)$ as describing an infinitely divisible random variable representing the "value of the random measure around the point $s \in S$." In other words, such an infinitely divisible random measure is an independently scattered and σ-additive random set function $M = \big(M(B),\ B \in \mathcal{S}_0\big)$ such that for every $B \in \mathcal{S}_0$, $M(B)$ is an infinitely divisible random variable with the characteristic triple

$$\sigma^2 = \int_B \sigma^2(s)\, m(ds), \quad \mu(\cdot) = \int_B \rho(s, \cdot)\, m(ds), \quad b = \int_B b(s)\, m(ds). \tag{3.27}$$

Example 3.2.3. A Lévy motion as a random measure

A Lévy process of Example 3.1.2 can be thought of as an infinitely divisible random measure in the following way. Let $S = \mathbb{R}$ (or a subinterval of $\mathbb{R}$), and let $\mathcal{S}$ denote the Borel σ-field. Let $\sigma^2 \geq 0$, $b \in \mathbb{R}$, and let ρ be a one-dimensional Lévy measure. Consider an infinitely divisible random measure M on S with the Lebesgue

control measure, constant local Gaussian variance $\sigma^2(s) = \sigma^2$, constant local Lévy measure $\rho(s, \cdot) = \rho(\cdot)$, and constant local shift $b(s) = b$, $s \in S$. In this case, $\mathcal{S}_0$ consists of all Borel sets of a finite Lebesgue measure, so we can define

$$X(t) = M\big((0,t]\big) \text{ for } t \geq 0 \text{ and } X(t) = -M\big((t,0]\big) \text{ for } t < 0.$$

It is easy to check using Proposition 3.2.1 that the process $(X(t),\ t \in \mathbb{R})$ thus constructed is continuous in probability and has stationary and independent increments. Hence, it is a Lévy motion. It is also elementary that for every $t > 0$, $X(t)$ is an infinitely divisible random variable with the characteristic triple $(t\sigma^2, t\rho, tb)$. In particular,

$$Ee^{i\theta X(t)} = \big(Ee^{i\theta X(1)}\big)^t \text{ for } t > 0.$$

Therefore, (the increments of) a Lévy process can be identified with a particular infinitely divisible random measure, and the law of a Lévy process is determined by the one-dimensional distribution of its value at time 1, or any other fixed positive time.

Example 3.2.4. A Gaussian random measure

An infinitely divisible random measure with vanishing local Lévy measures is a Gaussian random measure in the sense that the stochastic process $M = \big(M(B), B \in \mathcal{S}_0\big)$ is Gaussian. The most common sort of Gaussian random measures are the centered Gaussian random measures, where the local shifts $(b(s),\ s \in S)$ also vanish. In the latter case, there is no longer any utility in distinguishing the control measure m from the Gaussian variance measure γ, and a centered Gaussian random measure is a centered Gaussian process $M = \big(M(B),\ B \in \mathcal{S}_0\big)$ with covariance function given by

$$\mathrm{Cov}\big(M(B_1), M(B_2)\big) = \gamma(B_1 \cap B_2) = m(B_1 \cap B_2). \tag{3.28}$$

Example 3.2.5. A Poisson random measure

A Poisson random measure is an infinitely divisible random measure for which the local Gaussian variances vanish, $\sigma^2(s) = 0$, the local shifts are equal to unity, $b(s) = 1$, while $\rho(s, \cdot) = \delta_1(\cdot)$, for m-almost every $s \in S$. That is, all local Lévy measures are unit masses at the unity. In this case, the control measure m is also called the *mean measure* of the Poisson random measure M. Indeed, it follows from (3.27) that for every $B \in \mathcal{S}_0$, $M(B)$ is a Poisson random variable with parameter $m(B)$; hence $EM(B) = m(B)$ in this case.

Example 3.2.6. An α-stable random measure

Let $0 < \alpha < 2$. An infinitely divisible random measure with vanishing local Gaussian variance $\sigma^2(s) = 0$ for m-almost every $s \in S$ and local Lévy measures satisfying

$$\rho(s,\, dx) = \Big[w_+(s)x^{-(1+\alpha)}\mathbf{1}(x > 0) + w_-(s)|x|^{-(1+\alpha)}\mathbf{1}(x < 0)\Big]\, dx,\ s \in S\,, \tag{3.29}$$

where $w_+, w_- : S \to [0,\infty)$ are measurable functions, is an α-stable random measure. It is α-stable in the sense that the infinitely divisible process $M = \big(M(B),\, B \in \mathcal{S}_0\big)$ is α-stable, as in Example 3.1.11.

To see that this is the case, we need to check that the Lévy measure (3.17) satisfies the scaling condition (3.13). Let A be a measurable set in $\mathbb{R}^{\mathcal{S}_0}$. For $c > 0$,

$$\begin{aligned}
\mu(cA) &= \int_S \left(\int_{\mathbb{R}\setminus\{0\}} \mathbf{1}\big(x\mathbf{1}(s \in \cdot) \in cA\big)\, \rho(s, dx) \right) m(ds) \\
&= \int_S \left(w_+(s) \int_0^\infty \mathbf{1}\big(xc^{-1}\mathbf{1}(s \in \cdot) \in A\big) x^{-(1+\alpha)}\, dx \right. \\
&\left. + w_-(s) \int_{-\infty}^0 \mathbf{1}\big(xc^{-1}\mathbf{1}(s \in \cdot) \in A\big) |x|^{-(1+\alpha)}\, dx \right) m(ds) \\
&= \int_S \left(w_+(s) c^{-\alpha} \int_0^\infty \mathbf{1}\big(y\mathbf{1}(s \in \cdot) \in A\big) y^{-(1+\alpha)}\, dy \right. \\
&\left. + w_-(s) c^{-\alpha} \int_{-\infty}^0 \mathbf{1}\big(y\mathbf{1}(s \in \cdot) \in A\big) |y|^{-(1+\alpha)}\, dy \right) m(ds) = c^{-\alpha}\mu(A)\,,
\end{aligned}$$

as required.

If the weights in (3.29) satisfy $w_+(s) = w_-(s)$ for m-almost every $s \in S$, and the local shifts vanish, $b(s) = 0$ for m-almost every $s \in S$, then the infinitely divisible random measure M is symmetric α-stable; see Example 3.1.11 and Exercise 3.8.12. In this case, it is common to incorporate the weights $w(s) = w_+(s) - w_-(s)$, $s \in S$, into the control measure m. In fact, for an SαS random measure, it is usual to take as the control measure and the local Lévy measures

$$\tilde{m}(ds) = \frac{1}{\alpha C\alpha} w(s)\, m(ds), \quad \tilde{\rho}(s,\, dx) = \alpha C_\alpha |x|^{-(1+\alpha)}\, dx,\ s \in S,\ x \in \mathbb{R} \setminus \{0\}\,, \tag{3.30}$$

where

$$C_\alpha = \begin{cases} (\Gamma(1-\alpha)\cos(\pi\alpha/2))^{-1} & \text{if } \alpha \neq 1,, \\ 2/\pi & \text{if } \alpha = 1. \end{cases} \tag{3.31}$$

In this notation, for every set $B \in \mathcal{S}_0$, one has

$$Ee^{i\theta M(B)} = e^{-\tilde{m}(B)|\theta|^\alpha},\ \theta \in \mathbb{R}\,.$$

See Samorodnitsky and Taqqu (1994). In this book, in order to avoid ambiguity, we will call the measure given in (3.30) the *modified control measure* of an SαS random measure.

We conclude this section with an obvious but useful property of infinitely divisible random measures.

Proposition 3.2.7. *Let $M_1, \ldots, M_k$ be independent infinitely divisible random measures on $(S, \mathcal{S})$ with Gaussian variance measures $\gamma_1, \ldots, \gamma_k$, Lévy measures $\nu_1, \ldots, \nu_k$, and shift measures $\beta_1, \ldots, \beta_k$. Let $m_1, \ldots, m_k$ be the corresponding control measures, and let $\mathcal{S}_{0j} = \{B \in \mathcal{S} : m_j(B) < \infty\}$, $j = 1, \ldots, k$. Then*

$$M(B) = M_1(B) + \ldots + M_k(B), \ B \in \bigcap_{j=1}^{k} \mathcal{S}_{0j}$$

is an infinitely divisible random measure on $(S, \mathcal{S})$ with Gaussian variance measure $\gamma_1 + \ldots + \gamma_k$, Lévy measure $\nu_1 + \ldots + \nu_k$, and shift measure $\beta_1 + \ldots + \beta_k$.

Example 3.2.8. The sum of k independent centered Gaussian random measures on $(S, \mathcal{S})$ with control measures $m_1, \ldots, m_k$ is a centered Gaussian random measure on $(S, \mathcal{S})$ with control measure $m_1 + \ldots + m_k$.

Example 3.2.9. Let $0 < \alpha < 2$. The sum of k independent SαS random measures on $(S, \mathcal{S})$ with modified control measures $m_1, \ldots, m_k$ is an SαS random measure on $(S, \mathcal{S})$ with modified control measure $m_1 + \ldots + m_k$.

3.3 Infinitely Divisible Processes as Stochastic Integrals

The infinitely divisible random measures of Section 3.2 are important mainly because it is possible to construct a great variety of stochastic models by integrating suitable functions with respect to these random measures. In particular, it turns out that a very large number of infinitely divisible stochastic processes can be represented in the form

$$X(t) = \int_S f(t, s)\, M(ds), \quad t \in T, \tag{3.32}$$

where M is an infinitely divisible random measure, and $\big(f(t, \cdot),\ t \in T\big)$ is a family of nonrandom measurable functions. It turns out that many properties of infinitely divisible stochastic processes can be conveniently described when a process is given in the form of a stochastic integral as in (3.32).

Let $(S, \mathcal{S})$ be a measurable space, and M an infinitely divisible random measure on this space with control measure m, local Gaussian variance $(\sigma^2(s),\ s \in S)$, local Lévy measures $\big(\rho(s, \cdot), s \in S\big)$, and local shifts $(b(s),\ s \in S)$. For a simple function $f : S \to \mathbb{R}$ of the form

$$f(s) = \sum_{j=1}^{k} f_j \mathbf{1}(s \in B_j),\ s \in S, \tag{3.33}$$

with $k \geq 1, f_1, \ldots, f_k \in \mathbb{R}$, and $B_1, \ldots, B_k$ disjoint sets in $\mathcal{S}_0$, we define the integral of f with respect to M by

$$I(f) = \int_S f(s)\, M(ds) = \sum_{j=1}^{k} f_j M(B_j)\,.$$

It is easy to see using the additivity of an infinitely divisible random measure (part (ii) of Proposition 3.2.1) that this definition is consistent, in the sense that it will change only on a set of measure zero if one uses a different representation of the same simple function of the form (3.33).

The following proposition lists several important properties of the integrals of simple functions of the form (3.33).

Proposition 3.3.1. *(i) For a simple function f of the form (3.33), the integral $I(f)$ is an infinitely divisible random variable with the characteristic function given by*

$$Ee^{i\theta I(f)} = \exp\left\{ -\frac{1}{2}\theta^2 \int_S f(s)^2 \sigma^2(s)\, m(ds) \right. \tag{3.34}$$

$$\left. + \int_S \left(\int_{-\infty}^{\infty} \left(e^{i\theta f(s)x} - 1 - i\theta f(s)[\![x]\!] \right) \rho(s, dx) \right) m(ds) + i\theta \int_S f(s) b(s)\, m(ds) \right\}\,.$$

In particular, the characteristic triple $\left(\sigma^2(f), \mu(f), b(f)\right)$ of the integral $I(f)$ is as follows: the Gaussian variance is given by

$$\sigma^2(f) = \int_S f(s)^2 \sigma^2(s)\, m(ds)\,, \tag{3.35}$$

the Lévy measure is given by

$$\mu(f) = \nu_f \circ T_f^{-1}\,, \tag{3.36}$$

where the measure ν_f on $S \times \left(\mathbb{R} \setminus \{0\}\right)$ is defined by

$$\nu_f(A) = \nu\left(A \cap \left\{(s,x) : f(s) \neq 0\right\}\right),\ A \text{ measurable},$$

ν is the Lévy measure of the random measure M given by (3.22), and $T_f : S \times \left(\mathbb{R} \setminus \{0\}\right) \to \mathbb{R}$ is given by $T_f(s,x) = xf(s)$, $s \in S$, $x \in \mathbb{R}$. Finally, the "shift" parameter is given by

$$b(f) = \int_S \left(f(s)b(s) + \int_{-\infty}^{\infty} \left([\![f(s)x]\!] - f(s)[\![x]\!] \right) \rho(s, dx) \right) m(ds)\,. \tag{3.37}$$

(ii) If f and g are simple functions of the form (3.33), and a, b are real constants, then $af + bg$ is also a simple function of the form (3.33), and the integral has the linearity property

$$I(af + bg) = aI(f) + bI(g) \ \ a.s. \tag{3.38}$$

Proof. The expression for the characteristic function (3.34) follows from the following calculation, which uses the fact that an infinitely divisible random measure is independently scattered (part (i) of Proposition 3.2.1) and the expressions (3.27) for the characteristic triple of $M(B)$ for a set $B \in \mathcal{S}_0$:

$$Ee^{i\theta I(f)} = \prod_{j=1}^{k} Ee^{i\theta f_j M(B_j)} = \prod_{j=1}^{k} \exp\Big\{ -\frac{1}{2}\theta^2 f_j^2 \int_{B_j} \sigma^2(s)\, m(ds)$$

$$+ \int_{B_j} \left(\int_{-\infty}^{\infty} \left(e^{i\theta f_j x} - 1 - i\theta f_j [\![x]\!] \right) \rho(s, dx) \right) m(ds) + i\theta f_j \int_{B_j} b(s)\, m(ds) \Big\} ,$$

which coincides with the expression on the right-hand side of (3.34) for a function f of the form (3.33). The expressions (3.35), (3.36), and (3.37) follow by comparing (3.33) with the canonical representation (3.1); the latter requires insertion inside the truncation of $f(s)$ in the middle term on the right-hand side of (3.33). Hence we have a somewhat complicated expression for the "shift" parameter (3.37).

Since the value of the integral of a simple function of the form (3.33) changes only on an event of probability zero if one changes a representation of the function, part (ii) of the proposition is immediate.

In order to extend the definition of the integral from simple functions of the form (3.33) to a more general class of measurable functions on S, we need to introduce some notation. Let $L_0\big(\Omega, \mathcal{F}, P\big)$ be the space of all random variables defined on a probability space $\big(\Omega, \mathcal{F}, P\big)$, equipped with the metric $d_0(X, Y) = \|X - Y\|_0$, where $\|X\|_0 = E \min(|X|, 1)$. Then $L_0\big(\Omega, \mathcal{F}, P\big)$ is a complete metric space, and convergence in that space is equivalent to convergence in probability. For a simple function f of the form (3.33), we set

$$\|f\|_M = \sup\Big\{ \|I(g)\|_0 : \ g \text{ a simple function of the form (3.33)}, \\ |g(s)| \le |f(s)|, \text{ all } s \in S. \Big\} \tag{3.39}$$

We call a measurable function $f : S \to \mathbb{R}$ integrable with respect to the infinitely divisible random measure M if there is a sequence (f_n) of simple functions of the form (3.33) such that

$$f_n \to f \ \ m\text{-almost everywhere as } n \to \infty \text{ and } \lim_{k,n\to\infty} \|f_k - f_n\|_M = 0 . \tag{3.40}$$

The last requirement implies that the sequence of the integrals $\big(I(f_n)\big)$ is fundamental in $L_0\big(\Omega, \mathcal{F}, P\big)$, and since the latter space is complete, we can define

$$I(f) = \int_S f(s)\, M(ds) = \lim_{n\to\infty} I(f_n)\,, \tag{3.41}$$

defined in the sense of convergence in probability.

We denote by $L_0(M)$ the class of functions integrable with respect to an infinitely divisible random measure M. The next theorem, presented without a proof, describes explicitly the class $L_0(M)$ of integrable functions and the basic properties of their integrals. See Rosiński (2007) and Kwapień and Woyczyński (1992) for a proof.

Theorem 3.3.2. *(i) The class $L_0(M)$ of integrable functions is a linear vector space of functions, and a measurable function $f : S \to \mathbb{R}$ belongs to $L_0(M)$ if and only if the following three conditions hold:*

$$\int_S f(s)^2\sigma^2(s)\, m(ds) < \infty\,, \tag{3.42}$$

$$\int_S \left(\int_{-\infty}^{\infty} [\![f(s)x]\!]^2\, \rho(s, dx)\right) m(ds) < \infty\,, \tag{3.43}$$

$$\int_S \left| f(s)b(s) + \int_{-\infty}^{\infty} \Big([\![f(s)x]\!] - f(s)[\![x]\!]\Big)\, \rho(s, dx)\right| m(ds) < \infty\,. \tag{3.44}$$

(ii) For every $f \in L_0(M)$, the integral $I(f)$ is an infinitely divisible random variable with the characteristic function given by (3.34) and the characteristic triple $\big(\sigma^2(f), \mu(f), b(f)\big)$ given by (3.35), (3.36), and (3.37), respectively.

(iii) The integral is linear: for every $f, g \in L_0(M)$ and $a, b \in \mathbb{R}$, the equality (3.38) holds.

(iv) Let (f_n) be a sequence of measurable functions such that $|f_n| \le g$ m-a.e. for every $n \ge 1$, where $g \in L_0(M)$. Suppose that as $n \to \infty$, $f_n \to f$ m-a.e. for some measurable function f. Then $f \in L_0(M)$ and $I(f_n) \to I(f)$ in probability as $n \to \infty$.

Remark 3.3.3. Intuitively, condition (3.42) for integrability in part (i) of Theorem 3.3.2 requires the function to be integrable with respect to the Gaussian part of the random measure M. Condition (3.43) deals with the integrability of the function with respect to the Poisson part of the infinitely divisible random measure; it makes sure that the function does not "rearrange the Poisson jumps" of the random measure in such a way that they "break" the properties required of a Lévy measure. The requirement (3.44) is, in a sense, the trickiest. It is related to the centering of the Poisson points in a neighborhood of the origin, and its complicated form is due to the fact that the function f may "move" the points of the random measure in and out of a neighborhood of the origin.

Remark 3.3.4. If $f : S \to \mathbb{R}$ is a measurable function and $B \in \mathcal{S}$, it is common to use the notation $\int_B f(s)\, M(ds)$ for the integral $\int_S f(s)\mathbf{1}(s \in B)\, M(ds)$ as long as $f\mathbf{1}(\cdot \in B) \in L_0(M)$.

The definition (3.40) of an integrable function uses a somewhat involved "size functional" $\|\cdot\|_M$ in (3.39). Instead, one could have used the functional $\|\cdot\|_0$ directly and still defined the integral as in (3.41). The definition we are using, even though more restrictive, has the appealing property that for measurable functions f, g,

$$\text{if } f \in L_0(M) \text{ and } |g| \le |f| \text{ } m\text{-a.e. on } S, \text{ then } g \in L_0(M). \tag{3.45}$$

See Exercise 3.8.15. This property would not automatically follow under the broader definition, not the least due to a different treatment of the Poisson points of the random measure in and out of a neighborhood of the origin. See Remark 3.3.3 above.

In certain cases, however, the two definitions become equivalent, and the conditions for integrability in part (i) of Theorem 3.3.2 simplify. Suppose, for example, that an infinitely divisible random measure M is *symmetric*. That is,

$$\rho(s, \cdot) \text{ is a symmetric measure and } b(s) = 0 \text{ for } m\text{-a.e. } s \in S. \tag{3.46}$$

It follows from (3.27) that if M is a symmetric infinitely divisible random measure, then for every set $B \in \mathcal{S}_0$, $M(B)$ is an infinitely divisible random variable, with a symmetric Lévy measure and a vanishing "shift" parameter. This implies that $M(B)$ is a symmetric random variable, whence the adjective "symmetric" applied to M.

Proposition 3.3.5. *Let M be a symmetric infinitely divisible random measure. A measurable function f is integrable with respect to M if and only if there is a sequence (f_n) of simple functions of the form (3.33) such that*

$$f_n \to f \text{ } m\text{-almost everywhere as } n \to \infty \text{ and } \lim_{k,n\to\infty} \|f_k - f_n\|_0 = 0\,. \tag{3.47}$$

Further, $f \in L_0(M)$ if and only if (3.42) and (3.43) hold.

Proof. The only part of the statement that requires proof is that in the symmetric case, (3.47) implies (3.40). Let f and g be simple functions of the form (3.33) such that $|g(s)| \le |f(s)|$ for all $s \in S$. We can find disjoint sets $B_1, \ldots, B_k$ in $\mathcal{S}_0$ and real numbers $f_1, \ldots, f_k$ and $g_1, \ldots, g_k$ with $|g_j| \le |f_j|$ for each $j = 1, \ldots, k$ such that

$$f(s) = \sum_{j=1}^{k} f_j \mathbf{1}(s \in B_j), \quad g(s) = \sum_{j=1}^{k} g_j \mathbf{1}(s \in B_j), \ s \in S\,.$$

Because of the symmetry, we may use the contraction inequality in Theorem 10.7.5 to obtain

$$\|I(g)\|_0 = E\min\big(|I(g)|,1\big) = \int_0^1 P\big(|I(g)| > t\big)\,dt$$

$$= \int_0^1 P\left(\left|\sum_{j=1}^k g_j M(B_j)\right| > t\right) dt \le 2\int_0^1 P\left(\left|\sum_{j=1}^k f_j M(B_j)\right| > t\right) dt$$

$$= 2\int_0^1 P\big(|I(f)| > t\big)\,dt = 2E\min\big(|I(f)|,1\big) = 2\|I(f)\|_0\,.$$

We conclude that

$$\|I(f)\|_M \le 2\|I(f)\|_0$$

for every simple function of the form (3.33), and so (3.47) implies (3.40). □

Example 3.3.6. The integral with respect to a centered Gaussian measure

Let M be a centered Gaussian measure on S with control measure m; see Example 3.2.4. Applying Theorem 3.3.2 or Proposition 3.3.5, we see that $L_0(M) = L^2(m)$, and for $f \in L^2(m)$, $I(f)$ is a zero-mean normal random variable with variance $\|f\|^2_{L^2(m)}$.

Example 3.3.7. Functions integrable with respect to a Poisson random measure

Let M be a Poisson random measure on S with mean measure m; see Example 3.2.5. By Theorem 3.3.2, it is immediate that $f \in L_0(M)$ if and only if

$$\int_S \min(1, |f(s)|)\, m(ds) < \infty\,. \tag{3.48}$$

Example 3.3.8. The integral with respect to a symmetric α-stable measure

Let M be a symmetric α-stable random measure on S, $0 < \alpha < 2$, with control measure m and weights $w(s)$, $s \in S$; see Example 3.2.6. By Theorem 3.3.2 or Proposition 3.3.5, a measurable function f is integrable with respect to M if and only if (3.43) holds. Since

$$\int_S \left(\int_{-\infty}^{\infty} [\![f(s)x]\!]^2\, \rho(s,dx)\right) m(ds) = \int_S \left(\int_{-\infty}^{\infty} [\![f(s)x]\!]^2\, |x|^{-(1+\alpha)}\,dx\right) w(s)\, m(ds)$$

$$= \frac{4}{\alpha(2-\alpha)} \int_S |f(s)|^\alpha w(s)\, m(ds)\,,$$

we conclude that $L_0(M) = L^\alpha(\tilde{m})$, where $\tilde{m}$ is defined in (3.30). Furthermore, for $f \in L^\alpha(\tilde{m})$, the characteristic function of the integral $I(f)$ satisfies

$$
\begin{aligned}
Ee^{i\theta I(f)} &= \exp\left\{ \int_S \left(\int_{-\infty}^{\infty} \left(e^{i\theta f(s)x} - 1 - i\theta f(s)[\![x]\!] \right) \rho(s, dx) \right) m(ds) \right\} \\
&= \exp\left\{ -2 \int_S \left(\int_0^{\infty} \left(1 - \cos\big(\theta f(s)x\big) \right) |x|^{-(1+\alpha)}\, dx \right) w(s)\, m(ds) \right. \\
&= \exp\left\{ -2 \int_0^{\infty} \big(1 - cosy\big)\, y^{-(1+\alpha)}\, dy\, |\theta|^{\alpha} \int_S |f(s)|^{\alpha} w(s)\, m(ds) \right\} \\
&= \exp\Big\{ -|\theta|^{\alpha} \int_S |f(s)|^{\alpha}\, \tilde{m}(ds) \Big\}, \qquad (3.49)
\end{aligned}
$$

and so $I(f)$ is an SαS random variable. See Samorodnitsky and Taqqu (1994).

Integrals with respect to infinitely divisible random measures behave in the expected way with respect to the addition of the measures as well. In the situation of Proposition 3.2.7, we have the following additivity statement.

Proposition 3.3.9. *Let $M_1, \ldots, M_k$ be independent infinitely divisible random measures on $(S, \mathcal{S})$ and $f \in L_0(M_j)$ for $j = 1, \ldots, k$. Then $f \in L_0\big(M_1 + \ldots + M_k\big)$ and*

$$
\int_S f(s) \big(M_1 + \ldots + M_k\big)(ds) = \sum_{j=1}^{k} \int_S f(s)\, M_j(ds) \quad a.s.
$$

Proof. The fact that $f \in L_0\big(M_1 + \ldots + M_k\big)$ follows immediately from Theorem 3.3.2 and Proposition 3.2.7. Next, for $n \geq 1$ define

$$
f_n(s) = \begin{cases} \frac{j}{2^n} \quad \text{if } \frac{j}{2^n} \leq f(s) < \frac{j+1}{2^n},\ j = -n2^n, \ldots, n2^2 - 1, n2^n, \\ -n \text{ if } f(s) < -n, \\ n \quad \text{if } f(s) \geq n. \end{cases} \qquad (3.50)
$$

It follows by Theorem 3.3.2 that (f_n) is a sequence of simple functions of the type (3.33) for the infinitely divisible random measure $M_1 + \ldots + M_k$, and then, by Proposition 3.2.7, also for each infinitely divisible random measure M_j, $j = 1, \ldots, k$. Applying part (iv) of Theorem 3.3.2, we see that

$$
\int_S f_n(s) \big(M_1 + \ldots + M_k\big)(ds) \to \int_S f(s) \big(M_1 + \ldots + M_k\big)(ds),
$$

$$
\int_S f_n(s)\, M_j(ds) \to \int_S f(s)\, M_j(ds),\ j = 1, \ldots, k
$$

in probability as $n \to \infty$. Since for each $n \geq 1$ we have

$$
\int_S f_n(s) \big(M_1 + \ldots + M_k\big)(ds) = \sum_{j=1}^{k} \int_S f_n(s)\, M_j(ds)\,,
$$

the claim of the proposition follows. □

Not only is the integral with respect to an infinitely divisible random measure M of an integrable function an infinitely divisible random variable, but also the family of the integrals of a collection of integrable functions is an infinitely divisible stochastic process. This fact, announced at the beginning of this section, is the single most important property of stochastic integrals with respect to infinitely divisible random measures. It is contained in Theorem 3.3.10 below.

Theorem 3.3.10. *Let M be an infinitely divisible random measure on a measurable space $(S, \mathcal{S})$, with control measure m, local Gaussian variance $(\sigma^2(s),\ s \in S)$, local Lévy measures $\big(\rho(s,\cdot), s \in S\big)$, and local shifts $(b(s),\ s \in S)$. Let $f(t,\cdot) \in L_0(M)$ for each $t \in T$. Then the stochastic process $X(t) = I\big(f(t,\cdot)\big)$, $t \in T$, defined in (3.32) is infinitely divisible, with a weak characteristic triple $\big(\Sigma_{\mathbf{X}}, \mu_{\mathbf{X}}, \mathbf{b}_{\mathbf{X}}\big)$ given by*

$$\Sigma_{\mathbf{X}}(t_1,t_2) = \int_S f(t_1,s)f(t_2,s)\sigma^2(s)\,m(ds),\ t_1,t_2 \in T\,, \tag{3.51}$$

$$\mu_{\mathbf{X}} = \nu \circ H^{-1}\,, \tag{3.52}$$

where ν is the Lévy measure of the random measure M given by (3.22), and $H : S \times \mathbb{R} \to \mathbb{R}^T$ is given by $H(s,x) = \big(xf(t,s),\ t \in T\big)$ for $s \in S$, $x \in \mathbb{R}$. Finally,

$$b_{\mathbf{X}}(t) = \int_S \left(f(t,s)b(s) + \int_{-\infty}^{\infty} \Big([\![f(t,s)x]\!] - f(t,s)[\![x]\!] \Big)\, \rho(s,dx) \right) m(ds),\ t \in T\,. \tag{3.53}$$

Proof. We begin by noticing that by (3.42) and (3.44), the functions $\Sigma_{\mathbf{X}}$ and $\mathbf{b}_{\mathbf{X}}$ are well defined, since each $f(t,\cdot)$ is in $L_0(M)$. Next, let $t_1, \ldots, t_d \in T$, and $\theta_1, \ldots, \theta_d \in \mathbb{R}$. By the linearity of the integral (Theorem 3.3.2),

$$E \exp\left\{ \sum_{j=1}^{d} \theta_j X(t_j) \right\} = E \exp\left\{ \sum_{j=1}^{d} \theta_j I\big(f(t_j,\cdot)\big) \right\} = E \exp\left\{ I\left(\sum_{j=1}^{d} \theta_j f(t_j,\cdot) \right) \right\}.$$

Next, we use part (ii) of Theorem 3.3.2 with $f = \sum_{j=1}^{d} \theta_j f(t_j,\cdot)$ to conclude that in the notation (3.51), (3.52), and (3.53), this is further equal to

$$\exp\left\{ -\frac{1}{2} \int_S \left(\sum_{j=1}^{d} \theta_j f(t_j,s) \right)^2 \sigma^2(s)\, m(ds) \right.$$

$$+ \int_S \left(\int_{-\infty}^{\infty} \Big(e^{ix \sum_{j=1}^{d} \theta_j f(t_j,s)} - 1 - i \sum_{j=1}^{d} \theta_j f(t_j,s) [\![x]\!] \Big)\, \rho(s,dx) \right) m(ds)$$

$$\left. + i \int_S \sum_{j=1}^{d} \theta_j f(t_j,s) b(s)\, m(ds) \right\}$$

$$= \exp\left\{ -\frac{1}{2} \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \Sigma_{\mathbf{X}}\big(t_{j_1}, t_{j_2}\big)\theta_{j_1}\theta_{j_2} \right.$$

$$\left. + \int_{\mathbb{R}^T} \Big(e^{i\sum_{j=1}^{d} \theta_j x(t_j)} - 1 - i \sum_{j=1}^{d} \theta_j [\![x]\!](t_j) \Big) \mu_{\mathbf{X}}(d\mathbf{x}) + i \sum_{j=1}^{d} \theta_j b_{\mathbf{X}}(t_j) \right\} .$$

According to Theorem 3.1.7 and Remark 3.1.8, our statement will follow once we check that the measure $\mu_{\mathbf{X}}$ in (3.52) is a weak Lévy measure on T. However, for every $t \in T$,

$$\int_{\mathbb{R}^T} [\![x]\!](t)^2 \, \mu_{\mathbf{X}}(d\mathbf{x}) = \int_S \left(\int_{-\infty}^{\infty} [\![f(t,s)x]\!]^2 \, \rho(s, dx) \right) m(ds) < \infty$$

by (3.44), since $f(t, \cdot) \in L_0(M)$. Therefore, $\mu_{\mathbf{X}}$ satisfies Condition 3.1.4. □

In many concrete cases, it is easy to check that the measure $\mu_{\mathbf{X}}$ in (3.52) is a Lévy measure on $\mathbb{R}^T$, and not only a weak Lévy measure; see Exercise 3.8.18.

Combining the statement of Theorem 3.3.10 with Exercise 3.8.3, we immediately obtain the following important corollary.

Corollary 3.3.11. (i) *Let M be an infinitely divisible random measure on a measurable space $(S, \mathcal{S})$, with control measure m, local Gaussian variance $(\sigma^2(s),\ s \in S)$, local Lévy measures $\big(\rho(s, \cdot), s \in S\big)$, and local shifts $(b(s),\ s \in S)$. Let $f, g \in L_0(M)$. Then the integrals $I(f)$ and $I(g)$ are independent if and only if the following two conditions hold:*

$$\int_S f(s)g(s)\sigma^2(s)\, m(ds) = 0\,,$$

$$xf(s)g(s) = 0 \ \textit{for } \nu\textit{-a.e. } (s,x).$$

(ii) *Let $f(t, \cdot) \in L_0(M)$, $g(t, \cdot) \in L_0(M)$ for each $t \in T$. Then the stochastic processes $X(t) = I\big(f(t, \cdot)\big)$, $t \in T$, and $Y(t) = I\big(g(t, \cdot)\big)$, $t \in T$, are independent if and only if $f(t_1, \cdot)$ and $g(t_2, \cdot)$ satisfy the two assumptions of part (i) for all $t_1, t_2 \in T$ (i.e., if $X(t_1)$ and $Y(t_2)$ are independent for all $t_1, t_2 \in T$).*

When an infinitely divisible stochastic process is represented as a stochastic integral with respect to an infinitely divisible random measure, as in Theorem 3.3.10, it is sometimes more convenient to work with the process and study its properties via the integral representation, and not via its Lévy–Khinchine representation (3.9). Fortunately, it turns out that *every* infinitely divisible stochastic process with a σ-finite Lévy measure has an integral representation.

Theorem 3.3.12. *Let $\big(X(t),\ t \in T\big)$ be an infinitely divisible process with a σ-finite Lévy measure. Then there exist a measurable space $(S, \mathcal{S})$ and an infinitely divisible random measure on that space, with some control measure m and some local characteristics $\big(\sigma^2(s), \rho(s, \cdot), b(s)\big)$, and functions $f(t, \cdot) \in L_0(M)$, $t \in T$, such that*

$$\big(X(t),\, t \in T\big) \stackrel{d}{=} \left(\int_S f(t,s)\, M(ds),\ \ t \in T \right). \tag{3.54}$$

Proof. Let $\big(\Sigma_{\mathbf{X}}, \mu_{\mathbf{X}}, \mathbf{b}_{\mathbf{X}}\big)$ be the characteristic triple of the infinitely divisible process $\big(X(t),\, t \in T\big)$. We choose $S = \mathbb{R}^T$, and $\mathcal{S}$ the cylindrical σ-field on $\mathbb{R}^T$. Let

$$\big(X(t),\, t \in T\big) \stackrel{d}{=} \big(G(t),\, t \in T\big) + \big(Y(t),\, t \in T\big)$$

be the decomposition of the infinitely divisible stochastic process into a sum of two independent processes, a centered Gaussian process and an infinitely divisible process without a Gaussian component given in Corollary 3.1.12. We defined the ingredients of an infinitely divisible random measure on $(S, \mathcal{S})$ as follows.

Let γ be the (probability) law of the Gaussian process $\big(G(t),\, t \in T\big)$ on $\mathbb{R}^T$, and define a measure on $\mathbb{R}^T \times \big(\mathbb{R} \setminus \{0\}\big)$ by $\nu = \mu_{\mathbf{X}} \times \delta_1$. Finally, let $\beta = \delta_{\mathbf{a}}$, where $\mathbf{a}$ is a function on T defined by

$$a(t) = b_{\mathbf{X}}(t) - \int_{\mathbb{R}^T} \big([\![x(t)]\!] - x(t) \big)\, \mu_{\mathbf{X}}(d\mathbf{x}),\ \ t \in T\,.$$

Let now M be an infinitely divisible random measure on $(S, \mathcal{S})$ defined by the measures γ, ν, and β, and let m be the corresponding control measure, and $\big(\sigma^2(s), \rho(s,\cdot), b(s)\big)$ the corresponding local characteristics. For $\mathbf{x} = \big(x(t),\, t \in T\big) \in S$ and $t \in T$, we define $f(t,\mathbf{x}) = x(t)$. That is, f is the evaluation function on $\mathbb{R}^T$. By the definition of the cylindrical σ-field, $f(t,\cdot) : \mathbb{R}^T \to \mathbb{R}$ is a measurable function for every $t \in T$. Next we check that $f(t,\cdot) \in L_0(M)$ for each $t \in T$ by checking the three conditions in part (i) of Theorem 3.3.2.

First of all,

$$\int_S f(t,\mathbf{x})^2 \sigma^2(\mathbf{x})\, m(d\mathbf{x}) = \int_{\mathbb{R}^t} x(t)^2\, \gamma(d\mathbf{x}) = EG(t)^2 < \infty,$$

since a Gaussian process has a finite variance. This verifies condition (3.42). Next,

$$\int_S \left(\int_{-\infty}^{\infty} [\![f(t,\mathbf{x}) x]\!]^2\, \rho(\mathbf{x}, dx) \right) m(d\mathbf{x}) = \int_{\mathbb{R}^t} [\![x(t)]\!]^2\, \mu_{\mathbf{X}}(d\mathbf{x}) < \infty,$$

since $\mu_{\mathbf{X}}$ is a Lévy measure on $\mathbb{R}^T$. This verifies condition (3.43). Finally,

$$\int_S \left| f(t,\mathbf{x}) b(\mathbf{x}) + \int_{-\infty}^{\infty} \Big([\![f(t,\mathbf{x}) x]\!] - f(t,\mathbf{x}) [\![x]\!] \Big)\, \rho(\mathbf{x}, dx) \right| m(d\mathbf{x})$$

$$\leq \int_{\mathbb{R}^t} |x(t)|\, \beta(d\mathbf{x}) + \int_{\mathbb{R}^T} \big| [\![x(t)]\!] - x(t) \big|\, \mu_{\mathbf{X}}(d\mathbf{x}) = |a(t)| + \int_{\mathbb{R}^T} \big| [\![x(t)]\!] - x(t) \big|\, \mu_{\mathbf{X}}(d\mathbf{x}) < \infty,$$

because once again, $\mu_{\mathbf{X}}$ is a Lévy measure on $\mathbb{R}^T$. This verifies condition (3.44), and so $f(t,\cdot) \in L_0(M)$ for each $t \in T$.

The only step remaining to complete the proof of the theorem is to check that the (weak) characteristic triple of the infinitely divisible stochastic process defined as the integral on the right-hand side of (3.54) (and given in Theorem 3.3.10) coincides with the characteristic triple of the process $\big(X(t),\ t \in T\big)$. First of all, for $t_1, t_2 \in T$,

$$\int_S f(t_1, \mathbf{x}) f(t_2, \mathbf{x}) \sigma^2(\mathbf{x})\, m(d\mathbf{x}) = \int_{\mathbb{R}^t} x(t_1) x(t_2)\, \gamma(d\mathbf{x}) = EG(t_1)G(t_2)\,,$$

and so the Gaussian components in the two triples coincide. Trivially, $\nu \circ H^{-1} = \mu_{\mathbf{X}}$, which means that the Lévy measures in the two triples coincide as well. Finally, for $t \in T$,

$$\begin{aligned}\int_S \left(f(t, \mathbf{x}) b(\mathbf{x}) + \int_{-\infty}^{\infty} \Big([\![f(t, \mathbf{x}) x]\!] - f(t, \mathbf{x}) [\![x]\!] \Big)\, \rho(\mathbf{x}, dx) \right) m(d\mathbf{x}) \\ = a(t) - \int_{\mathbb{R}^T} \big([\![x(t)]\!] - x(t) \big)\, \mu_{\mathbf{X}}(d\mathbf{x}) = b_{\mathbf{X}}(t)\,,\end{aligned}$$

which checks that the last components in the two triples are the same, and the proof is complete. □

Example 3.3.13. Centered Gaussian processes as integrals with respect to a centered Gaussian measure Let M be a centered Gaussian measure on S with control measure m, and T a parameter space. Suppose that for each $t \in T$, a function $f(t, \cdot) \in L^2(m)$ is given. Generalizing Example 3.3.6, we observe that by Theorem 3.3.10, $X(t) = I\big(f(t, \cdot)\big)$, $t \in T$, is a centered Gaussian process with covariance function given by (3.51).

Example 3.3.14. Symmetric α-stable processes as integrals with respect to a symmetric α-stable measure Let M be an SαS random measure on S with a modified control measure $\tilde{m}$, and T a parameter space. Suppose now that for each $t \in T$, a function $f(t, \cdot) \in L^\alpha(\tilde{m})$, $0 < \alpha < 2$, is given. We can generalize Example 3.3.8 and define a stochastic process by $X(t) = I\big(f(t, \cdot)\big)$, $t \in T$. By Theorem 3.3.10, this process is an SαS process with characteristic function satisfying

$$E \exp\left\{ i \sum_{j=1}^{k} \theta_j X(t_j) \right\} = \exp\left\{ - \int_S \left| \sum_{j=1}^{k} \theta_j f\big(t_j, s\big) \right|^{\alpha} \tilde{m}(ds) \right\} \tag{3.55}$$

for every $k = 1, 2, \ldots$, $t_1, \ldots, t_k$, in T and real numbers $\theta_1, \ldots, \theta_k$. We leave the verification to Exercise 3.8.20.

Using stochastic integrals with respect to infinitely divisible random measures, we can define new infinitely divisible random measures, as described in the following proposition.

Proposition 3.3.15. *Let M be an infinitely divisible random measure on $(S, \mathcal{S})$ with a Gaussian variance measure γ, a Lévy measure ν, and a shift measure β, and $f : S \to \mathbb{R}$ a measurable function. Let*

$$\tilde{\mathcal{S}}_0 = \left\{ B \in \mathcal{S} : f\mathbf{1}(\cdot \in B) \in L_0(M) \right\}.$$

Then

$$M_f(B) = \int_B f(s)\, M(ds),\ B \in \tilde{\mathcal{S}}_0$$

is an infinitely divisible random measure on $(S, \mathcal{S})$ with a Gaussian variance measure $\tilde{\gamma}$, a Lévy measure $\tilde{\nu}$, and a shift measure $\tilde{\beta}$, where

$$\tilde{\gamma} \ll \gamma \ \text{ with } \ \frac{d\tilde{\gamma}}{d\gamma}(s) = f(s)^2,$$

the measure $\tilde{\nu}$ is the restriction to $S \times (\mathbb{R} \setminus \{0\})$ of the measure $\nu \circ F^{-1}$ with

$$F : S \times (\mathbb{R} \setminus \{0\}) \to S \times \mathbb{R} \ \text{ given by } \ F(s,x) = (s, xf(s)),$$

and $\tilde{\beta} = \tilde{\beta}_1 + \tilde{\beta}_2$, where

$$\tilde{\beta}_1 \ll \beta \ \text{ with } \ \frac{d\tilde{\beta}_1}{d\beta}(s) = f(s)$$

and $\tilde{\beta}_2$ is a signed measure on S defined by

$$\tilde{\beta}_2(B) = \int_B \int_{-\infty}^{\infty} \left([\![f(s)x]\!] - f(s)[\![x]\!] \right) \nu(ds, dx),\ B \in \tilde{\mathcal{S}}_0.$$

Remark 3.3.16. In order to simplify the notation in defining the measure $\tilde{\nu}_2$ in the proposition, we used the letter ν to denote the obvious extension of the measure ν to $S \times \mathbb{R}$. Note that the proposition simply says that for every set $B \in \tilde{\mathcal{S}}_0$, we have

$$\tilde{\gamma}(B) = \int_B f(s)^2\, \gamma(ds) \tag{3.56}$$

and

$$\tilde{\beta}(B) = \int_B f(s)\, \beta(ds) + \int_B \int_{-\infty}^{\infty} \left([\![f(s)x]\!] - f(s)[\![x]\!] \right) \nu(ds, dx). \tag{3.57}$$

Proof of Proposition 3.3.15. Let m be the control measure of the infinitely divisible random measure M and let $\left(\sigma^2(s),\ \rho(s,\cdot),\ b(s)\right)$, $s \in S$, be its local characteristics. By Theorem 3.3.10, $\left(M_f(B),\ B \in \tilde{\mathcal{S}}_0\right)$ is an infinitely divisible process with a weak characteristic triple $\left(\tilde{\Sigma},\ \tilde{\mu},\ \tilde{\mathbf{b}}\right)$ given as follows. For any sets $B_1,\ B_2 \in \tilde{\mathcal{S}}_0$,

$$\tilde{\Sigma}\big(M_f(B_1),\, M_f(B_2)\big) = \int_S \big(f(s)\mathbf{1}(s \in B_1)\big)\big(f(s)\mathbf{1}(s \in B_2)\big)\sigma^2(s)\, m(ds)$$
$$= \int_{B_1 \cap B_2} f(s)^2 \sigma^2(s)\, m(ds) = \int_{B_1 \cap B_2} f(s)^2\, \gamma(ds) = \tilde{\gamma}(B)\,.$$

Further, if $H : S \times \mathbb{R} \to \mathbb{R}^{\tilde{\mathcal{S}}_0}$ is given by $H(s,x) = \big(xf(s)\mathbf{1}(s \in B),\, B \in \tilde{\mathcal{S}}_0\big)$ for $s \in S$, $x \in \mathbb{R}$, and if Φ is as in (3.17), then

$$\tilde{\mu} = \nu \circ H^{-1} = \tilde{\nu} \circ \Phi^{-1}\,.$$

Finally, for every set $B \in \tilde{\mathcal{S}}_0$,

$$\tilde{b}(B) = \int_S \Bigg(f(s)\mathbf{1}(s \in B)b(s)$$
$$+ \int_{-\infty}^{\infty} \Big([\![f(s)\mathbf{1}(s \in B)x]\!] - f(s)\mathbf{1}(s \in B)[\![x]\!] \Big)\, \rho(s, dx) \Bigg)\, m(ds)$$
$$= \tilde{\beta}_1(B) + \tilde{\beta}_2(B).$$

Since the measure

$$\tilde{m}_0(B) := \int_B \int_{\mathbb{R}\setminus\{0\}} [\![x]\!]^2\, \tilde{\nu}(ds, dx) = \int_B \int_{\mathbb{R}\setminus\{0\}} [\![xf(s)]\!]^2\, \tilde{\nu}(ds, dx),\ B \in \mathcal{S}$$

is σ-finite because the measure m_0 in (3.14) is, the statement of the proposition will follow once we check that

$$\tilde{\mathcal{S}}_0 = \Big\{ B :\, \tilde{\gamma}(B) + \|\tilde{\beta}\|(B) + \tilde{m}_0(B) < \infty \Big\}.$$

This last statement, however, follows immediately from the description of $L_0(M)$ in Theorem 3.3.2. □

Combining Proposition 3.3.15 with examples 3.3.13 and 3.3.14 leads to the following useful corollary.

Corollary 3.3.17. *(i) Let M be a centered Gaussian measure on S with control measure m, and $f : S \to \mathbb{R}$ a measurable function. Let*

$$\tilde{\mathcal{S}}_0 = \Big\{ B \in \mathcal{S} :\, \int_B f(s)^2\, m(ds) < \infty \Big\}\,.$$

Then $\big(M_f(B),\, B \in \tilde{\mathcal{S}}_0\big)$ is a centered Gaussian measure on S with control measure $\tilde{m} \ll m$ and $d\tilde{m}/dm = f^2$.

(ii) Let M me an SαS random measure on S with a modified control measure m, and $f : S \to \mathbb{R}$ a measurable function. Let

$$\tilde{\mathcal{S}}_0 = \left\{ B \in \mathcal{S} : \int_B |f(s)|^\alpha \, m(ds) < \infty \right\}.$$

Then $\left(M_f(B),\, B \in \tilde{\mathcal{S}}_0\right)$ is an SαS random measure on S with modified control measure $\tilde{m} \ll m$ and $d\tilde{m}/dm = |f|^\alpha$.

Intuitively, one can view the relation between infinitely divisible random measures M and M_f in Proposition 3.3.15 as the relation of absolute continuity, with f serving as the "derivative." With this point of view, the following proposition is completely expected.

Proposition 3.3.18. *Let M, f, and M_f be as in Proposition 3.3.15. Let $g : S \to \mathbb{R}$ be a measurable function. Then $g \in L_0(M_f)$ if and only if $fg \in L_0(M)$, and in that case,*

$$\int_S g(s)\, M_f(ds) = \int_S f(s) g(s)\, M(ds) \text{ a.s.}$$

Proof. The fact that $g \in L_0(M_f)$ if and only if $fg \in L_0(M)$ follows from the description of the two L_0 spaces in Theorem 3.3.2.

Let now $g \in L_0(M_f)$. For $n \geq 1$, let g_n be the simple function defined on the basis of g as in (3.50).

By part (iv) of Theorem 3.3.2, we know that

$$\int_S g_n(s)\, M_f(ds) \to \int_S g(s)\, M_f(ds), \quad \int_S f(s) g_n(s)\, M(ds) \to \int_S f(s) g(s)\, M(ds)$$

in probability as $n \to \infty$. However, each g_n is a simple function of the form $g_n(s) = \sum_{j=1}^k a_j \mathbf{1}(s \in B_j)$ for some $k \geq 1$, real numbers (a_j), and measurable sets (B_j) such that $\mathbf{1}(\cdot \in B_j) \in L_0(M_f)$ and $f\mathbf{1}(\cdot \in B_j) \in L_0(M)$ for each j. Therefore,

$$\int_S g_n(s)\, M_f(ds) = \sum_{j=1}^k a_j M_f(B_j) = \sum_{j=1}^k a_j \int_{B_j} f(s)\, M(ds)$$

$$= \int_S f(s) \sum_{j=1}^k a_j \mathbf{1}(s \in B_j)\, M(ds) = \int_S f(s) g_n(s)\, M(ds),$$

and the claim follows. □

3.4 Series Representations

An infinitely divisible stochastic process without a Gaussian component has only a Poisson component. Such a process has representations, often explicit, as an infinite series involving the standard Poisson arrivals as well as an independent of them sequence of i.i.d. random vectors. Representations of this kind are often very helpful in understanding the structure of infinitely divisible processes and their properties. In this section, we present an introduction to such representations. A full discussion can be found in Rosiński and Samorodnitsky (2016).

We begin by introducing the notation that is standard in series representations of infinitely divisible stochastic processes. Let $\Gamma_1, \Gamma_2, \ldots$ be the ordered points of a unit-rate Poisson process on $(0, \infty)$. In queuing applications, it is common to use the terminology "arrival times" or "Poisson arrivals," and we will use these terms interchangeably. Note that

$$\Gamma_n = e_1 + \ldots + e_n, \ n = 1, 2, \ldots ,$$

where (e_j) is a sequence of i.i.d. standard exponential random variables.

Let T be an arbitrary parameter space, and let $(E, \mathcal{E})$ be a measurable space. Given a family of measurable functions $H(t; \cdot) : (0, \infty) \times E \to \mathbb{R}$, $t \in T$, we define a stochastic process

$$X(t) = \sum_{n=1}^{\infty} H(t; \Gamma_n, Y_n), \ t \in T , \tag{3.58}$$

where (Y_n) is a sequence of i.i.d. E-valued random variables independent of the Poisson arrivals (Γ_n). Of course, the functions $H(t; \cdot)$ must satisfy certain assumptions to make sure that the series in (3.58) converge. If the series converges, then the process $(X(t), \ t \in T)$ is automatically infinitely divisible, and it is possible to represent its characteristic triple through the functions $H(t; \cdot)$ and the law of a generic representative Y of the sequence (Y_n). In this case, one views the right-hand side of (3.58) as a series representation of the infinitely divisible stochastic process on the left-hand side. A given infinitely divisible stochastic process has, generally, multiple series representations.

The following theorem gives sufficient conditions for convergence in (3.58).

Theorem 3.4.1. *Let $H : (0, \infty) \times E \to \mathbb{R}$ be a measurable function. If*

$$\int_0^{\infty} E\big([\![H(x, Y)]\!]\big)^2 \, dx < \infty \ \text{ and } \ \int_0^{\infty} \big|E\big([\![H(x, Y)]\!]\big)\big| \, dx < \infty ,$$

then the series

$$X = \sum_{n=1}^{\infty} H(\Gamma_n, Y_n) \tag{3.59}$$

converges a.s. Further, X is infinitely divisible, with a characteristic triplet (σ^2, μ, b), where $\sigma = 0$, the Lévy measure is given by

$$\mu(B) = \int_0^\infty P\big(H(x, Y) \in B \setminus \{0\}\big)\, dx, \ B \text{ Borel}\,, \tag{3.60}$$

and

$$b = \int_0^\infty E\big([\![H(x, Y)]\!]\big)\, dx\,. \tag{3.61}$$

Proof. In order to prove the a.s. convergence of the series (3.59), we would like to use the three series theorem, Theorem 10.7.6. However, the latter applies to series of independent random variables, while the series in (3.59) is not, generally, a series of independent random variables due to the dependence introduced by the Poisson arrival sequence. To overcome this difficulty, we first condition on that sequence. Formally, we may assume that the underlying probability space is a product space, $\big(\Omega, \mathcal{F}, P\big) = \big(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2\big)$, and the Poisson arrival sequence (Γ_n) lives on Ω_1, while the i.i.d. sequence (Y_n) lives on Ω_2. By Fubini's theorem, the a.s. convergence of the series (3.59) will follow from its P_2-a.s. convergence for P_1-a.e. $\omega_1 \in \Omega_1$, which we now proceed to prove. Note that for a fixed $\omega_1 \in \Omega_1$, the series (3.59) is a series of independent random variables, so the three series theorem applies.

We begin with the first series in Theorem 10.7.6. Choosing $c = 1/2$, we use the Markov inequality to write

$$\begin{aligned}\sum_{n=1}^\infty P_2\big(|H(\Gamma_n(\omega_1), Y)| > 1/2\big) &= \sum_{n=1}^\infty P_2\big(|[\![H(\Gamma_n(\omega_1), Y)]\!]| > 1/2\big) \\ &\le 4 \sum_{n=1}^\infty E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]^2\big)\,.\end{aligned}$$

In order to prove that the last sum is finite P_1-a.s., it is enough to check that its expectation is finite. However, by Fubini's theorem, since the nth Poisson arrival has the gamma distribution with n degrees of freedom, we have

$$\begin{aligned}E_1 \sum_{n=1}^\infty E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]^2\big) &= \sum_{n=1}^\infty \int_0^\infty E_2\big([\![H(x, Y)]\!]^2\big) \frac{x^{n-1}}{(n-1)!} e^{-x}\, dx \\ &= \int_0^\infty E\big([\![H(x, Y)]\!]^2\big)\, dx < \infty,\end{aligned}$$

by an assumption of the theorem.

For the second series in Theorem 10.7.6, also with $c = 1/2$, we have

$$\sum_{n=1}^{\infty} E_2\Big(H(\Gamma_n(\omega_1), Y)\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| \le 1/2\big)\Big)$$

$$= \sum_{n=1}^{\infty} E_2\Big([\![H(\Gamma_n(\omega_1), Y)]\!]\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| \le 1/2\big)\Big)$$

$$= \sum_{n=1}^{\infty}\Big[E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]\big) - E_2\Big([\![H(\Gamma_n(\omega_1), Y)]\!]\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| > 1/2\big)\Big)\Big].$$

Note that

$$\sum_{n=1}^{\infty}\Big|E_2\Big([\![H(\Gamma_n(\omega_1), Y)]\!]\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| > 1/2\big)\Big)\Big|$$

$$\le \sum_{n=1}^{\infty} P_2\big(|H(\Gamma_n(\omega_1), Y)| > 1/2\big) < \infty$$

on a subset of Ω_1 of P_1-probability 1, as we have already proved. Therefore, it is enough to prove that

$$\sum_{n=1}^{\infty}\big|E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]\big)\big| < \infty$$

on a subset of Ω_1 of P_1-probability 1 as well. Taking as before an expectation with respect to P_1, we obtain by the second assumption of the theorem,

$$E_1\sum_{n=1}^{\infty}\big|E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]\big)\big| = \sum_{n=1}^{\infty}\int_0^{\infty}\big|E_2\big([\![H(x, Y)]\!]\big)\big|\frac{x^{n-1}}{(n-1)!}e^{-x}\,dx$$

$$= \int_0^{\infty}\big|E\big([\![H(x, Y)]\!]\big)\big|\,dx < \infty\,,$$

as required. Finally, convergence of the third series in Theorem 10.7.6 follows from what we have already proved, since

$$\sum_{n=1}^{\infty}\mathrm{Var}_2\Big(H(\Gamma_n(\omega_1), Y)\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| \le 1/2\big)\Big)$$

$$\le \sum_{n=1}^{\infty} E_2\Big[\Big(H(\Gamma_n(\omega_1), Y)\mathbf{1}\big(|H(\Gamma_n(\omega_1), Y)| \le 1/2\big)\Big)^2\Big]$$

$$\le \sum_{n=1}^{\infty} E_2\big([\![H(\Gamma_n(\omega_1), Y)]\!]^2\big) < \infty$$

on a subset of Ω_1 of P_1-probability 1.

Now that we have proved that the series in (3.59) converges with probability 1, we can write

$$X = \lim_{w\to\infty} \sum_{n:\, \Gamma_n \le w} H(\Gamma_n, Y_n)$$

in the sense of almost sure, hence also distributional, convergence. Let X_w be the random variable under the limit. We can write, in distribution,

$$X_w = \sum_{k=1}^{K_w} H(U_k, Y_k)\,,$$

where (U_n) is an i.i.d. sequence of random variables uniformly distributed between 0 and w, K_w has the Poisson distribution with mean w, and (U_n), (Y_n) and K are independent; see Exercise 3.8.19. That is, X_w is a compound Poisson random variable, so that

$$\begin{aligned} E^{i\theta X_w} = \exp\Bigg\{ \int_0^w \int_E \Big(e^{i\theta H(x,s)} - 1 - i\theta [\![H(x,s)]\!] \Big)\, F_Y(ds)\, dx \\ + i\theta \int_0^w \int_E [\![H(x,s)]\!]\, F_Y(ds)\, dx \Bigg\}\ ; \end{aligned}$$

see Example 3.1.1. Here F_Y is the law of Y_1 on E. As $w \to \infty$, the characteristic function on the left-hand side converges to the characteristic function of X, while the assumptions of the theorem guarantee that the expression on the right-hand side converges to

$$\begin{aligned} &\exp\Bigg\{ \int_0^\infty \int_E \Big(e^{i\theta H(x,s)} - 1 - i\theta [\![H(x,s)]\!] \Big)\, F_Y(ds)\, dx \\ &\qquad + i\theta \int_0^\infty \int_E [\![H(x,s)]\!]\, F_Y(ds)\, dx \Bigg\} \\ &= \exp\Bigg\{ \int_{-\infty}^\infty \Big(e^{i\theta z} - 1 - i\theta [\![z]\!] \Big)\, \mu(dz) + i\theta b \Bigg\} \end{aligned}$$

with μ and b given by (3.60) and (3.61) respectively. This completes the proof. □

The following is a consequence of Theorem 3.4.1. It can be proved using linearity in the same manner as in the proof of Theorem 3.3.10.

Corollary 3.4.2. *Let $H(t;\cdot) : (0,\infty) \times E \to \mathbb{R}$, $t \in T$, be a family of measurable functions. Assume that for every $t \in T$,*

$$\int_0^\infty E\big([\![H(t;x,Y)]\!] \big)^2\, dx < \infty \quad \text{and} \quad \int_0^\infty \big| E\big([\![H(t;x,Y)]\!] \big) \big|\, dx < \infty\,. \tag{3.62}$$

Then the $(X(t),\ t\ \in\ T)$ *in* (3.58) *is a well-defined infinitely divisible stochastic process with a weak characteristic triple* $\left(\Sigma_{\mathbf{X}}, \mu_{\mathbf{X}}, \mathbf{b}_{\mathbf{X}}\right)$*, where* $\Sigma_{\mathbf{X}}$ *vanishes, while*

$$\mu_{\mathbf{X}} = (\lambda \times F_Y) \circ H^{-1}$$

and

$$b_{\mathbf{X}}(t) = \int_0^\infty E\big([\![H(t;x,Y)]\!]\big)\, dx,\ t \in T\,.$$

Here F_Y *is the law of* Y_1 *on* E*, and* $H:\ (0,\infty)\times E \to \mathbb{R}^T$ *is defined by the collection* $H(t;\cdot):\ (0,\infty)\times E \to \mathbb{R}$*,* $t \in T$*.*

One of the very useful types of series representations of infinitely divisible processes is obtained as series representations of processes given as stochastic integrals with respect to infinitely divisible random measures. We now present one such representation. We restrict ourselves to the integrals with respect to *symmetric* infinitely divisible random measures.

Let M be a symmetric infinitely divisible random measure on S without a Gaussian component, with control measure m. Recall that this means that the local Gaussian variances vanish and that $b(s) = 0$ and $\rho(s,\cdot)$ is a symmetric measure for m-a.e. $s \in S$. Suppose that $(f(t,\cdot),\ t \in T)$ is a family of measurable functions on T such that

$$\int_S \int_{\mathbb{R}} [\![f(t;s)x]\!]^2\, \rho(s,dx)\, m(ds) < \infty \ \text{ for each } t \in T. \tag{3.63}$$

By Theorem 3.3.2, each function $f(t,\cdot)$ is in the space $L_0(M)$ of integrable functions; hence the stochastic process

$$X(t) = \int_S f(t,s)\, M(ds),\ t \in T\,, \tag{3.64}$$

is a well-defined symmetric infinitely divisible process. Its series representation can be obtained as follows. Let γ be a probability measure on S equivalent to the control measure m. Then

$$r(s) = \frac{dm}{d\gamma}(s),\ s \in S\,, \tag{3.65}$$

is a strictly positive (on a set of full measure m) function, and

$$\rho_r(s,\cdot) = r(s)\rho(s,\cdot) \tag{3.66}$$

is, for each $s\ \in\ S$, a symmetric one-dimensional Lévy measure. We define its generalized inverse by

$$G(x,s) = \inf\big\{y > 0:\ \rho_r\big(s,(y,\infty)\big) \le x/2\big\},\ x > 0\,. \tag{3.67}$$

Theorem 3.4.3. *Let (ε_n) be a sequence of i.i.d. Rademacher random variables (i.e., random variables taking values ± 1 with probability $1/2$ each), and let (Y_n) be a sequence of i.i.d. S-valued random variables with a common law γ. Let (Γ_n) be a sequence of standard Poisson arrivals on $(0,\infty)$. All three sequences are assumed to be independent. Then*

$$Y(t) = \sum_{n=1}^{\infty} \varepsilon_n G(\Gamma_n, Y_n) f(t, Y_n),\ t \in T\,, \tag{3.68}$$

is a well-defined stochastic process equal in its finite-dimensional distributions to the process $(X(t),\ t \in T)$ in (3.64).

Proof. We begin by checking the conditions (3.4.2) with $E = \{-1,1\} \times S$, $H(t;x,(y,s)) = y\,G(x,s)f(t,s)$, $t \in T$, $x > 0$, $y \in \{-1,1\}$, and $s \in S$. Since $E\big([\![H(t;x,(\varepsilon_1,Y_1)]\!]\big) = 0$ by symmetry for every $t \in T$ and $x > 0$, only the first condition in (3.4.2) needs to be verified. Note that for a fixed $s \in S$,

$$\lambda\big(\{x > 0 :\ G(x,s) > z\}\big) = 2\rho_r\big(s,(z,\infty)\big),\ z > 0\,. \tag{3.69}$$

Therefore, using Fubini's theorem and changing the variable of integration, we obtain

$$\int_0^\infty E\big([\![H(t;x,(\varepsilon_1,Y_1))]\!]\big)^2\,dx = E\left(\int_0^\infty [\![G(x,Y_1)f(t,Y_1)]\!]^2\,dx\right)$$

$$= 2E\left(\int_0^\infty [\![f(t,Y_1)z]\!]^2\,\rho_r(Y_1,dz)\right)$$

$$= 2\int_S\int_0^\infty [\![f(t,s)z]\!]^2\,\rho_r(s,dz)\,\gamma(ds) = \int_S\int_{\mathbb{R}} [\![f(t,s)z]\!]^2\,\rho(s,dz)\,m(ds) < \infty$$

for every $t \in T$ by (3.63). This verifies the assumptions of Corollary 3.4.2 and shows that $(Y(t), t \in T)$ is a well-defined symmetric infinitely divisible stochastic process. If B is a cylindrical set in $\mathbb{R}^T$, then the corollary tells us that the mass assigned by a weak Lévy measure of the process to the set B is

$$\frac{1}{2}(\lambda\times\gamma)\big(\{(x,s) :\ G(x,s)f(\cdot,s) \in B\}\big) + \frac{1}{2}(\lambda\times\gamma)\big(\{(x,s) :\ G(x,s)f(\cdot,s) \in -B\}\big)\,.$$

Appealing to (3.69), we see that this is also equal to

$$\int_S\int_0^\infty \mathbf{1}\big(xf(\cdot,s) \in B\big)\,\rho_r(s,dx)\,\gamma(ds) + \int_S\int_0^\infty \mathbf{1}\big(xf(\cdot,s) \in -B\big)\,\rho_r(s,dx)\,\gamma(ds)$$

$$= \int_S\int_{\mathbb{R}} \mathbf{1}\big(xf(\cdot,s) \in B\big)\,\rho(s,dx)\,m(ds)\,,$$

which is, by Theorem 3.3.10, the weight a weak Lévy measure of the process $(X(t),\ t \in T)$ assigns to the set B. Therefore, the two processes have the same finite-dimensional distributions. □

Example 3.4.4. Let M be an SαS random measure on S with a modified control measure $\tilde{m}$. Let $(f(t,\cdot),\ t \in T)$ be a family of measurable functions on T such that $f(t,\cdot) \in L^\alpha(\tilde{m})$ for each $t \in T$. Then the stochastic process $(X(t),\ t \in T)$ in (3.64) is a well-defined SαS process; see Example 3.3.14.

Let γ be a probability measure on S equivalent to $\tilde{m}$. Define

$$\tilde{r}(s) = \frac{d\tilde{m}}{d\gamma}(s),\ s \in S\,.$$

We apply Theorem 3.4.3. A straightforward calculation shows that

$$G(x,s) = (2C_\alpha)^{1/\alpha}\tilde{r}(s)^{1/\alpha}x^{-1/\alpha} \quad \text{for } x > 0 \text{ and } s \in S.$$

Therefore, a series representation (in law) of the SαS process $(X(t),\ t \in T)$ is given by

$$X(t) = (2C_\alpha)^{1/\alpha}\sum_{n=1}^{\infty}\varepsilon_n\Gamma_n^{-1/\alpha}\,\tilde{r}(Y_n)^{1/\alpha}f(t,Y_n),\ t \in T\,. \tag{3.70}$$

Certain other representations of SαS processes of the previous example are indicated in Exercise 3.8.14.

3.5 Examples of Infinitely Divisible Self-Similar Processes

A stochastic process $\big(X(t),\ t \in T\big)$ with $T = [0,\infty)$ or $T = \mathbb{R}$ is called *self-similar* if for some $H \in \mathbb{R}$,

$$\big(X(ct),\ t \in T\big) \stackrel{d}{=} \big(c^H X(t),\ t \in T\big) \tag{3.71}$$

in the sense of equality of the finite-dimensional distributions for every $c > 0$. The number H is the exponent of self-similarity. Some of the best-known examples of stochastic processes regarded as long-memory processes are the increment processes of certain self-similar processes with stationary increments. In turn, the best-known of the latter processes are infinitely divisible. In this section, we present several examples of infinitely divisible self-similar processes with stationary increments by constructing them as stochastic integrals defined in the previous section. The general theory of self-similar processes will be discussed in Chapter 8.

Example 3.5.1. Fractional Brownian motion A fractional Brownian motion, commonly abbreviated *FBM*, is the most famous self-similar process with stationary increments. Let $Q_t = \big(Q_t(x),\, x \in \mathbb{R}\big)$, $t \in \mathbb{R}$, be a family of kernels on $\mathbb{R}$ with $Q_t \in L^2\big(\lambda, \mathbb{R}\big)$ for $t \in \mathbb{R}$ (recall that λ is the one-dimensional Lebesgue measure) satisfying the following conditions: for all s, t and $c > 0$,

$$Q_t(x) - Q_s(x) = Q_{t-s}(x - s), \;\; Q_{ct}(cx) = c^{H-1/2} Q_t(x), \tag{3.72}$$

up to a set of Lebesgue measure zero, where H is a real number. Let $B = \big(B(x),\, x \in \mathbb{R}\big)$ be the standard Brownian motion, which we view as a Gaussian random measure on $\mathbb{R}$ with the Lebesgue Gaussian variance measure (Example 3.2.3). The stochastic process

$$X(t) = \int_{\mathbb{R}} Q_t(x)\, B(dx),\; t \in \mathbb{R}, \tag{3.73}$$

is a well-defined zero-mean Gaussian process. Notice that for every $c > 0$, by (3.72) and Exercise 3.8.17,

$$\begin{aligned} \big(X(ct),\, t \in \mathbb{R}\big) &= \Big(\int_{\mathbb{R}} Q_{ct}(x)\, B(dx),\; t \in \mathbb{R}\Big) \\ &= \Big(\int_{\mathbb{R}} c^{H-1/2} Q_t\big(c^{-1}x\big)\, B(dx),\; t \in \mathbb{R}\Big) \stackrel{d}{=} \Big(c^H \int_{\mathbb{R}} Q_t(x)\, B(dx),\; t \in \mathbb{R}\Big) \\ &= \big(c^H X(ct),\, t \in \mathbb{R}\big), \end{aligned}$$

and so the process (3.73) is self-similar with exponent H. Similarly, for every $s > 0$, by (3.72) and Exercise 3.8.16,

$$\begin{aligned} \Big(X(t+s) - X(s),\, t \in \mathbb{R}\Big) &= \Big(\int_{\mathbb{R}} \big(Q_{t+s}(x) - Q_s(x)\big)\, B(dx),\; t \in \mathbb{R}\Big) \\ = \Big(\int_{\mathbb{R}} Q_t(x-s)\, B(dx),\; t \in \mathbb{R}\Big) &\stackrel{d}{=} \Big(\int_{\mathbb{R}} Q_t(x)\, B(dx),\; t \in \mathbb{R}\Big) = \big(X(t),\, t \in \mathbb{R}\big). \end{aligned}$$

Therefore, the process (3.73) has stationary increments. We will see in Section 8.2 that up to a scale change, there is only one self-similar Gaussian process with stationary increments, and this process is called a fractional Brownian motion. Therefore, the process (3.73) is, in fact, a fractional Brownian motion.

We will also see in the sequel that the only possible range of the exponent of self-similarity is $0 \le H \le 1$ (unless the process is the trivial zero process), and if the process is continuous in probability (or only measurable), then the value $H = 0$ is impossible. Furthermore, the only such process with $H = 1$ is the trivial straight line process $X(t) = tX(1)$ a.s. for each $t \in \mathbb{R}$. Therefore, kernels not identically equal to zero satisfying (3.72) do not exist outside of the above range, and only in the range $0 < H < 1$ can nontrivial processes appear. In the above range, however, a number of kernels possess property (3.72). If $H \neq 1/2$, one choice is

$$g_t(c_1, c_2; H; x) = c_1\Big[\big((t-x)_+\big)^{H-1/2} - \big((-x)_+\big)^{H-1/2}\Big] \tag{3.74}$$

$$+c_2\Big[\big((t-x)_-\big)^{H-1/2} - \big((-x)_-\big)^{H-1/2}\Big],$$

where 0^a is interpreted as 0 for all $a \in \mathbb{R}$, and c_1, c_2 are real constants. It is elementary to check that this kernel has the required properties. Using kernel (3.74) leads to the so-called *moving-average representations* of the fractional Brownian motion. When $H = 1/2$, a fractional Brownian motion is just a Brownian motion, but using the kernel in (3.74) with $H = 1/2$ produces the zero process. Nontrivial moving-average representations of a Brownian motion can be obtained using the kernel

$$g_t(c_1, c_2; 1/2; x) := c_1\big(\mathbf{1}_{[0,t]}(x) - \mathbf{1}_{[t,0]}(x)\big) + c_2\big(\log|t-x| - \log|x|\big), \tag{3.75}$$

where c_1, c_2 are real constants. (It is transparent that kernel (3.75) with $c_2 = 0$ produces a Brownian motion, but the above discussion shows that the same is true in general.) Another well-known kernel is given as follows. Let $1/2 < \gamma < 1$. For $0 < H < 1, H \neq 3/2 - \gamma$, the kernel

$$Q_t(x) = \int_x^\infty (v-x)^{-\gamma}\big(|v|^{H+\gamma-3/2} - |v-t|^{H+\gamma-3/2}\big)dv,\ x \in \mathbb{R}, \tag{3.76}$$

for $t \geq 0$ is in $L^2\big(\lambda, \mathbb{R}\big)$ and satisfies (3.72). We will see in the sequel that there are natural generalizations of (3.76) to a family of functions on $\mathbb{R}^d$ that lead to self-similar processes with stationary increments that can be written as multiple integrals with respect to a Brownian motion. However, it turns out that the kernel (3.76) is actually a special case of the kernel (3.74) if $H \neq 1/2$, and of (3.75) if $H = 1/2$, as is, in fact, every kernel satisfying (3.76) and additional mild regularity assumptions; see Exercise 3.8.22.

One can also represent a fractional Brownian motion as a stochastic integral without using kernels of the type (3.72). For example, let $B_j = \big(B_j(x),\ x > 0\big)$, $j = 1, 2$, be independent standard Brownian motions. The stochastic process

$$X(t) = \int_0^\infty \frac{1-\cos tx}{x^{H+1/2}}\, B_1(dx) + \int_0^\infty \frac{\sin tx}{x^{H+1/2}}\, B_2(dx),\ t \in \mathbb{R}, \tag{3.77}$$

is, for $0 < H < 1$, a well-defined zero-mean Gaussian process. By Theorem 3.3.10, its incremental variance is

$$E\big(X(t) - X(s)\big)^2 = \int_0^\infty \frac{\big(\cos sx - \cos tx\big)^2}{x^{2H+1}}\, dx + \int_0^\infty \frac{\big(\sin tx - \sin sx\big)^2}{x^{2H+1}}\, dx$$

$$= \int_0^\infty \frac{2\Big(1 - \cos\big((t-s)x\big)\Big)}{x^{2H+1}}\, dx = c|t-s|^{2H},$$

where

$$c = 2\int_0^\infty \frac{1-\cos x}{x^{2H+1}}\,dx\,.$$

As we will see in Section 8.2, this means that $\big(X(t),\ t\in\mathbb{R}\big)$ is a fractional Brownian motion.

A more direct argument, using the properties of the integral, is also possible. The self-similarity with exponent H of the process defined in (3.77) follows using Exercise 3.8.17 in the same manner as above. Let us check the stationarity of the increments. For $s>0$, by the linearity of the integral,

$$\begin{aligned}
&\big(X(t+s)-X(s),\ t\in\mathbb{R}\big)\\
&=\left(\int_0^\infty \frac{\cos sx-\cos\big((t+s)x\big)}{s^{H+1/2}}\,B_1(dx)+\int_0^\infty \frac{\sin\big((t+s)x\big)-\sin tx}{x^{H+1/2}}\,B_2(dx),\ t\in\mathbb{R}\right)\\
&=\left(\int_0^\infty \frac{\cos sx-\big(\cos tx\cos sx-\sin tx\sin sx\big)}{s^{H+1/2}}\,B_1(dx)\right.\\
&\qquad\left.+\int_0^\infty \frac{\big(\sin tx\cos sx+\cos tx\sin sx\big)-\sin tx}{x^{H+1/2}}\,B_2(dx),\ t\in\mathbb{R}\right)\\
&=\left(\int_0^\infty \frac{1-\cos tx}{x^{H+1/2}}\cos sx\,B_1(dx)-\int_0^\infty \frac{1-\cos tx}{x^{H+1/2}}\sin sx\,B_2(dx)\right.\\
&\qquad\left.+\int_0^\infty \frac{\sin tx}{x^{H+1/2}}\sin sx\,B_1(dx)+\int_0^\infty \frac{\sin tx}{x^{H+1/2}}\cos sx\,B_2(dx),\ t\in\mathbb{R}\right).
\end{aligned}$$

For a Borel subset B of $(0,\infty)$ with a finite Lebesgue measure, let

$$\begin{aligned}
M_1(B) &= \int_B \cos sx\,B_1(dx)-\int_B \sin sx\,B_2(dx)\,,\\
M_2(B) &= \int_B \sin sx\,B_1(dx)+\int_B \cos sx\,B_2(dx)\,.
\end{aligned}$$

It follows from Corollary 3.3.17 and Proposition 3.2.7 that M_1 and M_2 are centered Gaussian random measures with the Lebesgue control measure each. Moreover, M_1 and M_2 are independent; see Exercise 3.8.21. Proposition 3.3.18 gives us

$$\begin{aligned}
&\big(X(t+s)-X(s),\ t\in\mathbb{R}\big)\\
&\stackrel{d}{=}\left(\int_0^\infty \frac{1-\cos tx}{x^{H+1/2}}\,M_1(dx)+\int_0^\infty \frac{\sin tx}{x^{H+1/2}}\,M_2(dx),\ t\in\mathbb{R}\right)\stackrel{d}{=}\big(X(t),\ t\in\mathbb{R}\big)\,,
\end{aligned}$$

and so the process in (3.77) has stationary increments.

The representation of a fractional Brownian motion given in (3.77) is its so-called *harmonizable representation*.

In the remainder of this section, we will consider self-similar symmetric α-stable processes with stationary increments, $0 < \alpha < 2$. Such processes are often referred to as fractional SαS motions. We will see in the sequel that the only interesting range of the exponent H of self-similarity of such a process is

$$H \in \begin{cases} (0, 1/\alpha] & \text{if } 0 < \alpha \leq 1, \\ (0, 1) & \text{if } 1 < \alpha < 2. \end{cases} \tag{3.78}$$

In a notable departure from the Gaussian case, where the only self-similar centered Gaussian process with stationary increments is a fractional Brownian motion, for "most" of the feasible pairs (α, H) of the exponents of stability and self-similarity, there exist many (substantially) different fractional SαS motions with exponent H of self-similarity.

Most of the known fractional SαS motions are represented as stochastic integrals with respect to SαS random measures. Often, one starts with an integral representation of a fractional Brownian motion and modifies it in an appropriate way.

Example 3.5.2. Linear fractional symmetric stable motions Let $Q_t = \big(Q_t(x), x \in \mathbb{R}\big)$, $t \in \mathbb{R}$, be a family of kernels on $\mathbb{R}$ with $Q_t \in L^\alpha\big(\lambda, \mathbb{R}\big)$ for $t \in \mathbb{R}$, satisfying the following conditions: for all s, t and $c > 0$,

$$Q_t(x) - Q_s(x) = Q_{t-s}(x - s), \quad Q_{ct}(cx) = c^{H-1/\alpha} Q_t(x), \tag{3.79}$$

up to a set of Lebesgue measure zero, where H is a real number (note the similarity to (3.72)). The number H has to be in a subset of the range described by (3.78). Let $L = \big(L(x), x \in \mathbb{R}\big)$ be the unit-scale SαS motion, which we now view as an SαS random measure on $\mathbb{R}$ with the Lebesgue modified control measure (Example 3.2.3). The stochastic process

$$X(t) = \int_{\mathbb{R}} Q_t(x)\, L(dx), \ t \in \mathbb{R}, \tag{3.80}$$

is a well-defined SαS process. The same arguments as in the case of the fractional Brownian motion shows that the process (3.80) is self-similar with exponent H and has stationary increments.

Let $0 < H < 1$. If we exclude the value $H = 1/\alpha$ (possible if $1 < \alpha < 2$), then a kernel satisfying the above properties is the following modification of (3.74):

$$\begin{aligned} g_t(c_1, c_2; H; x) = c_1 \Big[\big((t - x)_+\big)^{H-1/\alpha} - \big((-x)_+\big)^{H-1/\alpha} \Big] \\ + c_2 \Big[\big((t - x)_-\big)^{H-1/\alpha} - \big((-x)_-\big)^{H-1/\alpha} \Big], \end{aligned} \tag{3.81}$$

where once again, c_1, c_2 are real constants. If $1 < \alpha < 2$ and $H = 1/\alpha$, then the kernel (3.75) has the required properties in the stable case as well. Using kernels (3.81) and (3.75) leads to the so-called *linear fractional stable motions*. Notice the difference in terminology between the Gaussian and stable cases. In the Gaussian case, different kernels lead to different representations of the same fractional Brownian motion, and our terminology reflects that. In the α-stable case, $0 < \alpha < 2$, different kernels (or even different parameters in the same kernel) typically lead to different processes, and we have changed the terminology accordingly.

As in the Gaussian case, every kernel satisfying (3.79) and mild regularity assumptions reduces to the form (3.81) or (3.75), with parameters c_1 and c_2 depending on the kernel. To see this, one needs an obvious modification of Exercise 3.8.22. An example is the following extension of (3.76) to the case $1 < \alpha < 2$: for $0 < H < 1$, $1/\alpha < \gamma < 1$ such that $\gamma \neq 1 + 1/\alpha - H$, the kernel

$$Q_t(x) = \int_x^\infty (v-x)^{-\gamma} \left(|v|^{H+\gamma-1-1/\alpha} - |v-t|^{H+\gamma-1-1/\alpha} \right) dv, \ x \in \mathbb{R}, \tag{3.82}$$

for $t \geq 0$ is in $L^\alpha\big(\lambda, \mathbb{R}\big)$. It also clearly satisfies (3.79) (and hence has to be of the form (3.81) or (3.75), with parameters c_1 and c_2 depending on γ).

Recall that in the Gaussian case, all the different choices of the parameters c_1 and c_2 in (3.74) and (3.75) led to the same (up to a multiplicative constant) fractional Brownian motion. The situation is very different in the α-stable case, $0 < \alpha < 2$, as the following proposition shows.

Proposition 3.5.3. *Let $0 < H < 1$, and let $\big(X_i(t),\, t \in \mathbb{R}\big)$, $i = 1, 2$, be two linear fractional stable motions corresponding to (3.81) and (3.75) with the respective choices of the constants $\big(c_1^{(i)}, c_2^{(i)}\big)$, $i = 1, 2$. Then $\big(X_1(t),\, t \in \mathbb{R}\big) \stackrel{d}{=} \big(X_2(t),\, t \in \mathbb{R}\big)$ if and only if for some $\varepsilon \in \{-1, 1\}$, $c_1^{(1)} = \varepsilon c_1^{(2)}$ and $c_2^{(1)} = \varepsilon c_2^{(2)}$.*

Proof. Only the necessity has to be proved. Let $\big(X_1(t),\, t \in \mathbb{R}\big) \stackrel{d}{=} \big(X_2(t),\, t \in \mathbb{R}\big)$. The measures $\mu_{\mathbf{X}_1}$ and $\mu_{\mathbf{X}_2}$ given by (3.52) for the two processes (based on the integral representation (3.80) with the kernel (3.74) and (3.75)) are Lévy measures on $\mathbb{R}^{\mathbb{R}}$ (see Exercise 3.8.18). Hence, by the uniqueness statement of Theorem 3.1.7, we must have

$$\mu_{\mathbf{X}_1} = \mu_{\mathbf{X}_2}\,. \tag{3.83}$$

Suppose first that $H \neq 1/\alpha$. Notice that by (3.52), for every $a, b > 0$, for $i = 1, 2$,

$$\mu_{\mathbf{X}_i}\left(\psi \in \mathbb{R}^{\mathbb{R}} :\ \psi'(0)\psi'(1) < 0,\ \lim_{n\to\infty} \frac{\psi(2n) - \psi(n)}{n^{H-1/\alpha}} \text{ exists and is greater than } a, \right.$$

$$\lim_{n\to\infty} \frac{\psi(-2n)-\psi(-n)}{n^{H-1/\alpha}} \text{ exists and is greater than } b\Big)$$
$$= C_\alpha \Big(2^{H-1/\alpha}-1\Big)^\alpha \min\Big(\big|c_1^{(i)}\big|^\alpha a^{-\alpha},\ \big|c_2^{(i)}\big|^\alpha b^{-\alpha}\Big)$$

if $c_1^{(i)}c_2^{(i)} > 0$, and the measure is equal to zero if $c_1^{(i)}c_2^{(i)} \le 0$ (see (3.31)).

If $c_1^{(1)}c_2^{(1)} > 0$, then it follows from (3.83) that $c_1^{(2)}c_2^{(2)} > 0$ as well, and letting $a \to 0$ or $b \to 0$, we obtain the statement of the proposition. If $c_1^{(1)}c_2^{(1)} \le 0$, then it follows from (3.83) that $c_1^{(2)}c_2^{(2)} \le 0$ as well. In the case $c_1^{(1)}c_2^{(1)} < 0$, we obtain the statement of the proposition by looking at

$$\mu_{\mathbf{X}_i}\Big(\psi \in \mathbb{R}^{\mathbb{R}} :\ \psi'(0)\psi'(1) < 0,\ \lim_{n\to\infty} \frac{\psi(2n)-\psi(n)}{n^{H-1/\alpha}} \text{ exists and is greater than } a,$$
$$\lim_{n\to\infty} \frac{\psi(-2n)-\psi(-n)}{n^{H-1/\alpha}} \text{ exists and is less than } -b\Big).$$

We leave it to the reader to consider the final case $c_1^{(1)}c_2^{(1)} = c_1^{(2)}c_2^{(2)} = 0$ (Exercise 3.8.23).

If $1 < \alpha < 2$ and $H = 1/\alpha$, we proceed in a similar manner. For every $a, b > 0$, for $i = 1, 2$,

$$\mu_{\mathbf{X}_i}\Big(\psi \in \mathbb{R}^{\mathbb{R}} :\ \psi'(0)\psi'(1) < 0,\ \lim_{n\to\infty} \big(\psi(2n)-\psi(n)\big) \text{ exists and is greater than } a,$$
$$\lim_{n\to\infty} \big(\psi(n)-\psi(-n)\big) \text{ exists and is greater than } b\Big)$$
$$= C_\alpha \min\Big(\big|c_2^{(i)}\big|^\alpha (\log 2)^\alpha a^{-\alpha},\ \big|c_1^{(i)}\big|^\alpha b^{-\alpha}\Big)$$

if $c_1^{(i)}c_2^{(i)} > 0$, and the measure is equal to zero if $c_1^{(i)}c_2^{(i)} \le 0$. Once again, if $c_1^{(1)}c_2^{(1)} > 0$, then it follows from (3.83) that $c_1^{(2)}c_2^{(2)} > 0$ as well, and we obtain the statement of the proposition by letting $a \to 0$ or $b \to 0$. If $c_1^{(1)}c_2^{(1)} \le 0$ (and then also $c_1^{(2)}c_2^{(2)} \le 0$), we consider instead

$$\mu_{\mathbf{X}_i}\Big(\psi \in \mathbb{R}^{\mathbb{R}} :\ \psi'(0)\psi'(1) < 0,\ \lim_{n\to\infty} \big(\psi(2n)-\psi(n)\big) \text{ exists and is greater than } a,$$
$$\lim_{n\to\infty} \big(\psi(n)-\psi(-n)\big) \text{ exists and is less than } -b\Big),$$

and the case $c_1^{(1)}c_2^{(1)} = c_1^{(2)}c_2^{(2)} = 0$ is left to Exercise 3.8.23. □

Example 3.5.4. Harmonizable fractional stable motion Let $(\Omega', \mathcal{F}', P')$ be a probability space supporting two measurable stochastic processes, $\big(G_j(x) = G_j(\omega', x),\ x > 0\big)$, $j = 1, 2$, such that for each x, $G_1(x)$ and $G_2(x)$ are independent

standard normal random variables. Let M be an SαS infinitely divisible random measure on $\Omega' \times (0,\infty)$ with modified control measure $P' \times \lambda$. The random measure itself lives on some other probability space, $(\Omega, \mathcal{F}, P)$, and so does the stochastic process defined for $0 < H < 1$ as

$$X(t) = \int_{\Omega'} \int_0^\infty \frac{(1-\cos tx)G_1(\omega', x) + \sin tx\, G_2(\omega', x)}{x^{H+1/\alpha}} M(d\omega', dx), \ t \in \mathbb{R}; \tag{3.84}$$

we are using the double integral notation instead of the more awkward

$$\int_{\Omega' \times (0,\infty)} \cdots M(d(\omega', x)) .$$

It is elementary that the process (3.84) is a well-defined SαS process. It follows from (3.55) that for all $k = 1, 2, \ldots$, real numbers $t_1, \ldots, t_k$, and $\theta_1, \ldots, \theta_k$,

$$E \exp\left\{ i \sum_{j=1}^k \theta_j X(t_j) \right\} \tag{3.85}$$

$$= \exp\left\{ -E' \int_0^\infty \left| \sum_{j=1}^k \theta_j \frac{\big(1-\cos(t_j x)\big)G_1(x) + \sin(t_j x)\, G_2(x)}{x^{H+1/\alpha}} \right|^\alpha dx \right\}$$

$$= \exp\left\{ -\int_0^\infty E' \left| G_1(x) \sum_{j=1}^k \theta_j \big(1-\cos(t_j x)\big) + G_2(x) \sum_{j=1}^k \theta_j \sin(t_j x) \right|^\alpha x^{-(\alpha H+1)}\, dx \right\}$$

$$= \exp\left\{ -\int_0^\infty E|G_1(1)|^\alpha \left[\left(\sum_{j=1}^k \theta_j \big(1-\cos(t_j x)\big) \right)^2 + \left(\sum_{j=1}^k \theta_j \sin(t_j x) \right)^2 \right]^{\alpha/2} \right.$$

$$\left. x^{-(\alpha H+1)}\, dx \right\} .$$

Observe that the finite-dimensional distributions of the process $\big(X(t),\, t \in \mathbb{R}\big)$ defined in (3.84) do not depend on further properties of the processes $\big(G_j(x)$ $x > 0\big)$, $j = 1, 2$, apart from the fact that for each x, $G_1(x)$ and $G_2(x)$ are independent standard normal random variables.

Applying (3.85), we see that for every $c > 0$,

$$E \exp\left\{ i \sum_{j=1}^k \theta_j X(ct_j) \right\} = E \exp\left\{ i \sum_{j=1}^k \theta_j c^H X(t_j) \right\} ,$$

and so the process defined in (3.84) is self-similar with exponent H. Furthermore, for $s \in \mathbb{R}$,

$$X(t+s) - X(s) =$$

$$\int_{\Omega'} \int_0^\infty \frac{\big(\cos sx - \cos(t+s)x\big) G_1(\omega', x) + \big(\sin(t+s)x - \sin sx\big)\, G_2(\omega', x)}{x^{H+1/\alpha}} M(d\omega', dx)$$

$$= \int_{\Omega'} \int_0^\infty \frac{(1 - \cos tx)\, \tilde{G}_1(\omega', x) + \sin tx\, \tilde{G}_2(\omega', x)}{x^{H+1/\alpha}} M(d\omega', dx)\,,$$

where

$$\tilde{G}_1(x) = \cos sx\, G_1(x) - \sin sx\, G_2(x), \quad \tilde{G}_2(x) = \sin sx\, G_1(x) + \cos sx\, G_2(x),\ x > 0.$$

Clearly, it is still true that for each x, $\tilde{G}_1(x)$ and $\tilde{G}_2(x)$ are independent standard normal random variables. Therefore, the process $\big(X(t+s) - X(s),\ t \in \mathbb{R}\big)$ has, for each s, the same finite-dimensional distributions as the process $\big(X(t),\ t \in \mathbb{R}\big)$, and so the latter process has stationary increments.

The process $\big(X(t),\ t \in \mathbb{R}\big)$ defined in (3.84) is called the real harmonizable SαS motion, and it is an extension to the α-stable case of the harmonizable representation (3.77), even though the two integrals do not quite look alike.

The real harmonizable fractional SαS motion is a very different process from the linear fractional SαS motions of Example 3.5.2; see Section 3.6.

Example 3.5.5. FBM-local time fractional stable motions As in Example 3.5.4, we begin with a probability space $\big(\Omega', \mathcal{F}', P'\big)$. This time, we would like this probability space to support a fractional Brownian motion $\big(B_{H'}(t),\ t \geq 0\big)$ with exponent H' of self-similarity and consequently also its jointly continuous local time process $\big(l(x,t) = l(x,t)(\omega'),\ x \in \mathbb{R},\ t \geq 0\big)$. The fact that $\big(\Omega', \mathcal{F}', P'\big)$ supports such a local time process follows from Proposition 10.4.6.

Let M be an SαS random measure on the space $\Omega' \times \mathbb{R}$ with modified control measure $P' \times \lambda$. Once again, the random measure itself lives on some other probability space, $\big(\Omega, \mathcal{F}, P\big)$. We define

$$X(t) = \int_{\Omega'} \int_{\mathbb{R}} l(x,t)(\omega')\, M(d\omega', dx),\ t \geq 0. \tag{3.86}$$

The following proposition shows that this is, in fact, a well-defined SαS process that is also self-similar and has stationary increments.

Proposition 3.5.6. *The process* $(X(t),\ t \geq 0)$ *in* (3.86) *is a well-defined SαS process. It has stationary increments, and is self-similar with exponent*

$$H = 1 - H' + H'/\alpha = 1 + H'\left(\frac{1}{\alpha} - 1\right). \tag{3.87}$$

Proof. To show that the process in (3.86) is well defined, we need to show that

$$\int_{\Omega'} \int_{\mathbb{R}} l(x,t)(\omega')^\alpha\, P'(d\omega')\, dx = E' \int_{\mathbb{R}} l(x,t)^\alpha\, dx < \infty.$$

By Proposition 10.4.3, Fubini's theorem, and the Cauchy–Schwartz inequality, we have

$$\begin{aligned} E' \int_{\mathbb{R}} l(x,t)^{\alpha}\, dx &= \int_{\mathbb{R}} E'\Big(l(x,t)^{\alpha} \mathbf{1}\big(\sup_{0\le s\le t} |B_{H'}(s)| \ge |x| \big)\Big)\, dx \\ &\le \int_{\mathbb{R}} \Big(E' l(x,t)^{2\alpha} \Big)^{1/2} \Big(P\big(\sup_{0\le s\le t} |B_{H'}(s)| \ge |x| \big)\Big)^{1/2} dx \\ &\le \Big(\sup_{x\in\mathbb{R}} E' l(x,t)^{2\alpha} \Big)^{1/2} \int_{\mathbb{R}} \Big(P\big(\sup_{0\le s\le t} |B_{H'}(s)| \ge |x| \big)\Big)^{1/2} dx\,. \end{aligned}$$

By Proposition 10.4.7, the first term on the right-hand side above is finite. Furthermore, by the Borell-TIS inequality of Theorem 10.7.8, for all x with $|x|$ large enough,

$$\begin{aligned} P\big(\sup_{0\le s\le t} |B_{H'}(s)| \ge |x| \big) &\le P\big(\big| \sup_{0\le s\le t} |B_{H'}(s)| - E \sup_{0\le s\le t} |B_{H'}(s)| \big| \ge |x|/2 \big) \\ &\le 2\exp\big\{ -(x/2)^2 / 2\mathrm{Var}\big(B_{H'}(t)\big) \big\} = 2\exp\big\{ -x^2/8t^{2H'} \big\}\,, \end{aligned}$$

and so the integral

$$\int_{\mathbb{R}} \Big(P\big(\sup_{0\le s\le t} |B_{H'}(s)| \ge |x| \big)\Big)^{1/2} dx$$

is clearly finite. Therefore, the process in (3.86) is well defined.

Notice that by Proposition 10.4.6 and Theorem 10.7.8, fractional Brownian motion satisfies the assumptions of parts (i) and (iii) of Proposition 10.4.8. In particular, for $c > 0$, $k \ge 1$, $\theta_1, \dots, \theta_k \in \mathbb{R}$ and $t_1, \dots, t_k \ge 0$, we can use first (3.55) and then part (i) of Proposition 10.4.8 to obtain

$$\begin{aligned} E \exp\left(i \sum_{j=1}^{k} \theta_j X(ct_j) \right) &= \exp\left(- \int_{\mathbb{R}} \mathbf{E}' \left| \sum_{j=1}^{k} \theta_j l(x, ct_j) \right|^{\alpha} dx \right) \\ &= \exp\left(- \int_{\mathbb{R}} \mathbf{E}' \left| \sum_{j=1}^{k} \theta_j c^{1-H'} l\big(\frac{x}{c^{H'}}, t_j\big) \right|^{\alpha} dx \right) \\ &= \exp\left(-c^{\alpha(1-H')} \mathbf{E}' \int_{\mathbb{R}} \left| \sum_{j=1}^{k} \theta_j l\big(\frac{x}{c^{H'}}, t_j\big) \right|^{\alpha} dx \right) \\ &= \exp\left(-c^{\alpha(1-H')+H'} \mathbf{E}' \int_{\mathbb{R}} \left| \sum_{j=1}^{k} \theta_j l(y, t_j) \right|^{\alpha} dy \right) \end{aligned}$$

$$= E \exp\left(i \sum_{j=1}^{k} \theta_j c^{1-H'+H'/\alpha} X(t_j) \right)$$

$$= E \exp\left(i \sum_{j=1}^{k} \theta_j c^{H} X(t_j) \right) .$$

Therefore, the process $\big(X(t), t \geq 0\big)$ in (3.86) is self-similar with exponent of self-similarity given by (3.87). Next, using first (3.55) and then part (iii) of Proposition 10.4.8, we see that for every $h \geq 0$, $k \geq 1$, $\theta_1, \ldots, \theta_k \in \mathbb{R}$ and $t_1, \ldots, t_k \geq 0$,

$$E \exp\left(i \sum_{j=1}^{k} \theta_j (X(t_j + h) - X(h)) \right)$$

$$= \exp\left(- \int_{\mathbb{R}} E' \left| \sum_{j=1}^{k} \theta_j (l(x, t_j + h) - l(x, h)) \right|^{\alpha} dx \right)$$

$$= \exp\left(- \int_{\mathbb{R}} E' \left| \sum_{j=1}^{k} \theta_j l(x, t_j) \right|^{\alpha} dx \right)$$

$$= E \exp\left(i \sum_{j=1}^{k} \theta_j X(t_j) \right) .$$

Therefore, the process (3.86) has stationary increments. □

Interestingly, the exponent of self-similarity of the FBM-local time fractional stable motion (3.86) belongs to the range

$$\begin{cases} H \in \big(1, 1/\alpha\big) & \text{if } 0 < \alpha < 1 \\ H = 1 & \text{if } \alpha = 1, \\ H \in \big(1/\alpha, 1\big) & \text{if } 1 < \alpha < 2. \end{cases}$$

In spite of the fact that in the case $\alpha = 1$, all choices of $H' \in (0, 1)$ of the exponent of self-similarity of the underlying fractional Brownian motion lead to the same value $H = 1$ of the exponent of self-similarity of the FBM-local time fractional stable motion, the finite-dimensional distributions of the latter depend on the value of H'; see Proposition 4.4 in Cohen and Samorodnitsky (2006).

The FBM-local time fractional stable motion turns out to be a very different process from both the linear fractional SαS motions of Example 3.5.2 and the real harmonizable fractional SαS motion of Example 3.5.4, as will be seen in Section 3.6.

3.6 Stationary Infinitely Divisible Processes

In this section, we discuss certain properties of stationary infinitely divisible processes. We will use frequently the following notation. For $t \in \mathbb{R}$ let $\theta_t : \mathbb{R}^{\mathbb{R}} \to \mathbb{R}^{\mathbb{R}}$ be defined by

$$\theta_t(\mathbf{x})(w) = x(t+w),\ w \in \mathbb{R}$$

for $\mathbf{x} = \big(x(w),\ w \in \mathbb{R}\big)$. The shifts $(\theta_t,\ t \in \mathbb{R})$ define a group of measurable maps.

We begin with a basic structural result on stationary infinitely divisible processes.

Theorem 3.6.1. *Let $\big(X(t),\ t \in \mathbb{R}\big)$ be an infinitely divisible process with a characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$. Then the process $\mathbf{X}$ is stationary if and only if the following conditions are satisfied:*

$$\Sigma(s,t) = R(t-s),\ s,t \in \mathbb{R} \tag{3.88}$$

for some nonnegative definite function R on $\mathbb{R}$;

$$\text{the function } \mathbf{b} \text{ is constant on } \mathbb{R}; \tag{3.89}$$

$$\mu \circ \theta_r^{-1} = \mu \ \text{ for all } r \in \mathbb{R}. \tag{3.90}$$

Proof. For every $r \in \mathbb{R}$, the shifted process $\big(X(t+r),\ t \in \mathbb{R}\big)$ is clearly an infinitely divisible process with a characteristic triple $\big(\Sigma_r, \mu_r, \mathbf{b}_r\big)$ given by

$$\Sigma_r(s,t) = \Sigma(s+r, t+r),\ s,t \in \mathbb{R},\ \ \mu_r = \mu \circ \theta_r^{-1},\ \ \mathbf{b}_r = \theta_r(\mathbf{b})\,.$$

The process is stationary if and only if $\big(\Sigma_r, \mu_r, \mathbf{b}_r\big) = \big(\Sigma, \mu, \mathbf{b}\big)$ for all r, which is equivalent to the conditions (3.88), (3.89), and (3.90). □

Remark 3.6.2. It should be clear that the statement of Theorem 3.6.1 remains true for an infinitely divisible process in discrete time, $\big(X_n,\ n \in \mathbb{Z}\big)$. In this case, all the functions are defined on $\mathbb{Z}$, and the shifts $\theta_n : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ for $n \in \mathbb{Z}$ are now shifts of functions defined on the integers. We will set $\theta = \theta_1$, so that $\theta_n = \theta^n$ for each $n \in \mathbb{Z}$. In fact, in the rest of this section, we will deal mostly with stationary infinitely divisible processes in discrete time.

Let $\big(X_n,\ n \in \mathbb{Z}\big)$ be a stationary infinitely divisible process. Its Lévy measure μ is, by Theorem 3.6.1, invariant under the measure-preserving map θ. It turns out that it is useful to classify stationary infinitely divisible processes according to the recurrence properties of this map as discussed in Section 2.4.

Definition 3.6.3. A stationary infinitely divisible process with Lévy measure μ is said to be generated by a conservative (respectively dissipative, null, positive) flow if the shift θ is a conservative (respectively dissipative, null, positive) map on $\big(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}^{\mathbb{Z}}, \mu\big)$.

The decomposition given in (2.44) immediately gives us the following result.

Theorem 3.6.4. *Every stationary infinitely divisible process* $\mathbf{X} = \big(X_n,\, n \in \mathbb{Z}\big)$ *can be decomposed into a sum* $\mathbf{X} = \mathbf{X}^{(\mathrm{p})} + \mathbf{X}^{(\mathrm{cn})} + \mathbf{X}^{(\mathrm{d})}$ *of three independent stationary infinitely divisible processes such that* $\mathbf{X}^{(\mathrm{p})}$ *is generated by a positive flow,* $\mathbf{X}^{(\mathrm{cn})}$ *is generated by a conservative null flow, and* $\mathbf{X}^{(\mathrm{d})}$ *is generated by a dissipative flow. Moreover, the Lévy measures of* $\mathbf{X}^{(\mathrm{p})}$, $\mathbf{X}^{(\mathrm{cn})}$, *and* $\mathbf{X}^{(\mathrm{d})}$ *are uniquely determined.*

Remark 3.6.5. The uniqueness part of Theorem 3.6.4 cannot be strengthened to the uniqueness of the decomposition, because both the Gaussian part and the shift of the process $\mathbf{X}$ can be decomposed in multiple ways. When the Gaussian part is absent, however, the decomposition is automatically unique up to a shift by a constant. If there is a "canonical" shift (such as zero shift in the symmetric case, for instance), the decomposition becomes truly unique.

Since in most cases we work with infinitely divisible processes given as stochastic integrals with respect to infinitely divisible random measures, we discuss next the relationship between an integral representation of a stationary infinitely divisible process and its decomposition in Theorem 3.6.4. We will consider several situations.

The first situation we consider is both very general and the most transparent. Let $\big(E, \mathcal{E}, m\big)$ be a σ-finite measure space, and let $\phi : E \to E$ be a nonsingular measure-preserving map. Let ρ be a one-dimensional Lévy measure, and $b \in \mathbb{R}$ a constant. Let M be an infinitely divisible random measure on E with control measure m and constant local characteristics, equal to $(0, \rho, b)$. In particular, M has no Gaussian component. Let $f \in L_0(M)$. We consider an infinitely divisible process given by

$$X_n = \int_E f \circ \phi^n(s)\, M(ds),\ n \in \mathbb{Z}\,. \tag{3.91}$$

Notice that by Theorem 3.3.12, every stationary process $\mathbf{X} = \big(X_n,\, n \in \mathbb{Z}\big)$ has a representation of the form (3.91). See Exercise 3.8.24.

Theorem 3.6.6. *(i) The process* $\mathbf{X}$ *given by* (3.91) *is a well-defined stationary infinitely divisible process.*

(ii) Let $\mathcal{P}(\phi)$, $\mathcal{CN}(\phi)$, *and* $\mathcal{D}(\phi)$ *be the decomposition* (2.44) *of the map* ϕ. *Then*

$$X_n^{(\mathrm{p})} = \int_{\mathcal{P}(\phi)} f \circ \phi^n(s)\, M(ds),\ n = 0, 1, \dots\,,$$

$$X_n^{(\mathrm{cn})} = \int_{\mathcal{CN}(\phi)} f \circ \phi^n(s)\, M(ds),\ n = 0, 1, \dots,$$

and

$$X_n^{(\mathrm{d})} = \int_{\mathcal{D}(\phi)} f \circ \phi^n(s)\, M(ds),\ n = 0, 1, \dots,$$

are well-defined independent stationary infinitely divisible processes, $\mathbf{X}^{(\mathrm{p})}$ *is generated by a positive flow,* $\mathbf{X}^{(\mathrm{cn})}$ *is generated by a conservative null flow, and* $\mathbf{X}^{(\mathrm{d})}$ *is generated by a dissipative flow. Furthermore,* $\mathbf{X} = \mathbf{X}^{(\mathrm{p})} + \mathbf{X}^{(\mathrm{cn})} + \mathbf{X}^{(\mathrm{d})}$.

(iii) *If* ϕ *is positive, then* $\mathbf{X}$ *is generated by a positive flow. If* ϕ *is conservative null, then* $\mathbf{X}$ *is generated by a conservative null flow. If* ϕ *is dissipative, then* $\mathbf{X}$ *is generated by a dissipative flow. The converse of these statements is true if the function* f *has full support, i.e.,*

$$m\Big(s \in E : f \circ \phi^n(s) = 0 \text{ for all } n = 0, 1, 2, \dots\Big) = 0\,. \tag{3.92}$$

Proof. The process $\mathbf{X}$ is well defined because the fact that $f \in L_0(M)$ means that $f \circ \phi^n \in L_0(M)$ for every n, since ϕ preserves the measure m. Furthermore, by Theorem 3.3.10, the process $\mathbf{X}$ is infinitely divisible, without a Gaussian component, with a constant shift function, and Lévy measure given by

$$\mu_X = (m \times \rho) \circ H^{-1}\,, \tag{3.93}$$

with $H : E \times \mathbb{R} \to \mathbb{R}^{\mathbb{Z}}$ given by $H(s,x)(n) = xf \circ \phi^n(s)$, $n \in \mathbb{Z}$, for $s \in E$ and $x \in \mathbb{R}$. By Theorem 3.6.1, we conclude that $\mathbf{X}$ is a stationary process.

Next, since $\mathcal{P}(\phi)$, $\mathcal{CN}(\phi)$, and $\mathcal{D}(\phi)$ are invariant under the map ϕ, we see that

$$\begin{aligned}\int_{\mathcal{P}(\phi)} f \circ \phi^n(s)\, M(ds) &= \int_E \mathbf{1}_{\mathcal{P}(\phi)}(s) f \circ \phi^n(s)\, M(ds) \\ &= \int_E \big[\mathbf{1}_{\mathcal{P}(\phi)} f\big] \circ \phi^n(s)\, M(ds)\,.\end{aligned}$$

Further, (3.45) tells us that $\mathbf{1}_{\mathcal{P}(\phi)} f \in L_0(M)$. Therefore, the already established part (i) of the theorem tells us that $\mathbf{X}^{(\mathrm{p})}$ is a well-defined stationary infinitely divisible process. Similarly, $\mathbf{X}^{(\mathrm{cn})}$ and $\mathbf{X}^{(\mathrm{d})}$ are also well-defined stationary infinitely divisible processes. Furthermore, since $\mathcal{P}(\phi)$, $\mathcal{CN}(\phi)$, and $\mathcal{D}(\phi)$ are invariant under the map ϕ, we obtain independence of the processes $\mathbf{X}^{(\mathrm{p})}$, $\mathbf{X}^{(\mathrm{cn})}$, and $\mathbf{X}^{(\mathrm{d})}$ by Corollary 3.3.11. The fact that $\mathbf{X} = \mathbf{X}^{(\mathrm{p})} + \mathbf{X}^{(\mathrm{cn})} + \mathbf{X}^{(\mathrm{d})}$ follows from Theorem 3.3.2.

Let us prove that $\mathbf{X}^{(\mathrm{d})}$ is generated by a dissipative flow. If the Lévy measure μ_{d} of this process vanishes, then there is nothing to prove, so assume that μ_{d} does not vanish. Let h be a positive function in $L^1(\mu_{\mathrm{d}})$. By Theorem 2.4.5, we need to prove that

$$\mu_{\mathrm{d}}\left(\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \sum_{n=1}^{\infty} h\big(\theta^n(\mathbf{x})\big) = \infty\right) = 0\,.$$

By Theorem 3.3.10, this is the same as to show that

$$(m \times \rho)\left((s,x) \in \mathcal{D}(\phi) \times \mathbb{R} : \sum_{n=1}^{\infty} h\big(x(f \circ \phi^{n+k}(s),\, k \in \mathbb{Z})\big) = \infty\right) = 0\,.$$

By Fubini's theorem, there is a Borel subset B of $\mathbb{R}$ of a full ρ-measure such that for every $x \in B$, the function

$$h_x(s) = h\big(x(f \circ \phi^k(s),\ k \in \mathbb{Z})\big),\ s \in \mathcal{D}(\phi)\,,$$

is in $L^1\big(m|\,\mathcal{D}(\phi)\big)$ (the measure m restricted to $\mathcal{D}(\phi)$), and it is enough to prove that for every $x \in B$,

$$m\left(s \in \mathcal{D}(\phi) : \sum_{n=1}^{\infty} h_x \circ \phi^n(s) = \infty\right) = 0\,.$$

This, however, follows from the fact that the map ϕ, restricted to $\mathcal{D}(\phi)$, is dissipative and from Theorem 2.4.5. Therefore, $\mathbf{X}^{(\mathrm{d})}$ is generated by a dissipative flow, and in the same way, we can prove that $\mathbf{X}^{(\mathrm{cn})}$ is generated by a conservative flow. Let us check that the latter process is also generated by a null flow. If the process is not generated by a null flow, then by Theorem 3.6.4, this process has a positive part with a nonvanishing Lévy measure. We now use part (ii) of Theorem 2.4.11 (instead of Theorem 2.4.5) to show, in the above notation, that for every sequence $(w_n,\ n = 1, 2, \ldots)$ in $\mathcal{W}$, there is a set of positive measure ρ such that for every x in this set, we have

$$m\left(s \in \mathcal{CN}(\phi) : \sum_{n=1}^{\infty} w_n h_x \circ \phi^n(s) = \infty\right) > 0\,.$$

This, however, contradicts part (ii) of Theorem 2.4.11 applied to $\mathcal{N}(\phi)$.

Finally, to see that $\mathbf{X}^{(\mathrm{p})}$ is generated by a positive flow, assume that $m(\mathcal{P}(\phi)) > 0$, and choose a ϕ-invariant probability measure $\tilde{m}$ supported by $\mathcal{P}(\phi)$ and equivalent to m on this set. Then by Theorem 3.3.10, there exists a shift-invariant probability measure on $\mathbb{R}^{\mathbb{Z}}$ equivalent to the Lévy measure μ_p of the process $\mathbf{X}^{(\mathrm{p})}$. This means that the left shift is positive with respect to μ_p, and hence the process $\mathbf{X}^{(\mathrm{p})}$ is generated by a positive flow.

If ϕ is a dissipative map, then $\mathcal{P}(\phi)$ and $\mathcal{CN}(\phi)$ have m-measure zero, so that the processes $\mathbf{X}^{(\mathrm{p})}$ and $\mathbf{X}^{(\mathrm{cn})}$ vanish, and hence by part (ii) of the theorem, the process $\mathbf{X} = \mathbf{X}^{(\mathrm{d})}$ is generated by a dissipative flow. Similarly, if ϕ is a positive (respectively conservative null) map, then the process $\mathbf{X}$ is generated by a positive (respectively conservative null) flow.

Suppose now that (3.92) holds. Let $\mathbf{X}$ be generated by a dissipative flow. If ϕ is not dissipative, then either $m(\mathcal{P}(\phi)) > 0$ or $m(\mathcal{CN}(\phi)) > 0$, and hence by (3.92), either the process $\mathbf{X}^{(\mathrm{p})}$ or the process $\mathbf{X}^{(\mathrm{cn})}$ has a nonvanishing Lévy measure. By Theorem 3.6.4, this means that the process $\mathbf{X}$ is not generated by a dissipative flow. This contradiction shows that ϕ is a dissipative map. Similarly, if $\mathbf{X}$ is generated by a positive (respectively conservative null) flow and (3.92) holds, then ϕ is a positive (respectively conservative null) map. □

We consider next stationary SαS processes, $0 < \alpha < 2$, represented as stochastic integrals with respect to an SαS random measure, i.e., processes given in the form

$$X_n = \int_E f_n(s)\, M(ds),\ n \in \mathbb{Z}\,, \tag{3.94}$$

where M is an SαS random measure on a measurable space $\big(E, \mathcal{E}\big)$, with modified control measure $\tilde{m}$, and $f_n \in L^\alpha(\tilde{m})$ for $n \in \mathbb{Z}$; see Example 3.3.14. Note that in the first part of the theorem, we do not assume that the functions (f_n) are of the special form they take in (3.91).

Theorem 3.6.7. *Let* $\mathbf{X}$ *be a stationary SαS process,* $0 < \alpha < 2$, *given in the form* (3.94).

(i) *If*

$$\sum_{n=1}^{\infty} |f_n(s)|^\alpha < \infty \ \ \tilde{m}\text{-a.e.}, \tag{3.95}$$

then $\mathbf{X}$ *is generated by a dissipative flow. If for some sequence* $(w_n,\, n = 1, 2, \ldots)$ *in* $\mathcal{W}$,

$$\sum_{n=1}^{\infty} w_n |f_n(s)|^\alpha < \infty \ \ \tilde{m}\text{-a.e.}, \tag{3.96}$$

then $\mathbf{X}$ *is generated by a null flow. If for some* $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} |f_n(s)|^{\alpha+\varepsilon} = \infty \ \ \tilde{m}\text{-a.e.}, \tag{3.97}$$

then $\mathbf{X}$ *is generated by a conservative flow. If for every sequence* $(w_n,\, n = 1, 2, \ldots)$ *in* $\mathcal{W}$, *for some* $0 < \varepsilon < \alpha$,

$$\sum_{n=1}^{\infty} w_n \min\big(|f_n(s)|^{\alpha+\varepsilon}, |f_n(s)|^{\alpha-\varepsilon}\big) = \infty \ \ \tilde{m}\text{-a.e.}, \tag{3.98}$$

then $\mathbf{X}$ *is generated by a positive flow.*

(ii) *Suppose that a stationary SαS process,* $0 < \alpha < 2$, *is given in the form* (3.91), *with* M *an SαS random measure on a measurable space* $\big(E, \mathcal{E}\big)$, *with modified control measure* $\tilde{m}$ *(so that* $f_n(s) = f \circ \phi^n(s)$*). If*

$$\sum_{n=1}^{\infty} |f_n(s)|^\alpha = \infty \ \ \tilde{m}\text{-a.e.}, \tag{3.99}$$

then $\mathbf{X}$ *is generated by a conservative flow. If for every sequence* $(w_n,\ n = 1, 2, \ldots)$ *in* $\mathcal{W}$,

$$\sum_{n=1}^{\infty} w_n |f_n(s)|^\alpha = \infty \ \tilde{m}\text{-a.e.}, \tag{3.100}$$

then $\mathbf{X}$ *is generated by a positive flow.*

Proof. Suppose that (3.95) holds. Then by (3.93),

$$\sum_{n=1}^{\infty} h\big(\theta^n(\mathbf{x})\big) < \infty \ \mu_X\text{-a.e.},$$

where $h : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is given by $h(\mathbf{x}) = |x(0)|^\alpha$. This implies that $\mathcal{C}(\theta) = \emptyset$ μ_X-a.e. (see Exercise 2.6.12), and so $\mathbf{X}$ is generated by a dissipative flow. Similarly, if (3.96) holds for some sequence $(w_n,\ n = 1, 2, \ldots)$ in $\mathcal{W}$, then the function h above also satisfies

$$\sum_{n=1}^{\infty} w_n h\big(\theta^n(\mathbf{x})\big) < \infty \ \mu_X\text{-a.e.},$$

which implies, by Theorem 2.4.11, that $\mathcal{P}(\theta) = \emptyset$ μ_X-a.e.

If, on the other hand, (3.97) holds, then we define a function $h : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ by

$$h(\mathbf{x}) = \min\big(|x(0)|^{\alpha-\varepsilon}, |x(0)|^{\alpha+\varepsilon}\big).$$

It is elementary to check that $h \in L^1(\mu_X)$, and by the assumption,

$$\sum_{n=1}^{\infty} h\big(\theta^n(\mathbf{x})\big) = \infty \ \mu_X\text{-a.e.} \tag{3.101}$$

By Theorem 2.4.5, this implies that $\mathcal{D}(\theta) = \emptyset$ μ_X-a.e., and so $\mathbf{X}$ is generated by a conservative flow. Using the same function h and Theorem 2.4.11, we see that if (3.98) holds for every sequence $(w_n,\ n = 1, 2, \ldots)$ in $\mathcal{W}$, then $\mathcal{N}(\theta) = \emptyset$ μ_X-a.e., and so $\mathbf{X}$ is generated by a positive flow.

If we have $f_n(s) = f \circ \phi^n$, $n \in \mathbb{Z}$, then the sufficiency of the condition (3.99) follows from Theorem 3.6.6 and Theorem 2.4.5, while the sufficiency of the condition (3.100) follows from Theorem 3.6.6 and Theorem 2.4.11. □

The following proposition is sometimes helpful in conjunction with Theorem 3.6.7, because it allows us to relate properties of stationary SαS processes to different values of α.

Proposition 3.6.8. *Let* α_1, α_2 *be numbers in* $(0, 2)$. *Suppose that both* $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$ *have integral representation* (3.94), *with the same functions* (f_n) *and the same modified control measure* $\tilde{m}$. *In one case, the random measure* M *is an* Sα_1S *measure, while in the second case, the random measure* M *is an* Sα_2S *measure.*

Assume that both $\mathbf{X}^{(1)}$ *and* $\mathbf{X}^{(2)}$ *are well-defined stationary processes. If* $\mathbf{X}^{(1)}$ *is generated by a dissipative (respectively conservative, positive, null) flow, then also* $\mathbf{X}^{(2)}$ *is generated by a dissipative (respectively conservative, positive, null) flow.*

Proof. It follows from Theorem 3.3.10 that the Lévy measures $\mu_{X^{(1)}}$ and $\mu_{X^{(2)}}$ of the two processes are equivalent. Since both the Hopf decomposition and the positive–null decomposition are clearly invariant under a switch from one measure to an equivalent one, the statement of the proposition follows. □

Example 3.6.9. The increment process $X_n = X(n) - X(n-1)$, $n \geq 1$, of the linear fractional stable motion $\mathbf{X}$ of Example 3.5.2 is generated by a dissipative flow. Indeed, if $H \neq 1/\alpha$, we have

$$X_n = \int_{\mathbb{R}} f_n(s)\, L(ds),\ n = 1, 2, \dots,$$

with

$$\begin{aligned} f_n(s) =& c_1\Big[\big((n-s)_+\big)^{H-1/\alpha} - \big((n-1-s)_+\big)^{H-1/\alpha}\Big] \\ &+ c_2\Big[\big((n-s)_-\big)^{H-1/\alpha} - \big((n-1-s)_-\big)^{H-1/\alpha}\Big], \end{aligned}$$

so that

$$f_n(s) \sim c_1(H - 1/\alpha)\, n^{H-1/\alpha-1} \ \text{ as } n \to \infty,$$

and (3.95) holds because $H < 1$.

A similar computation can easily be done in the case $1 < \alpha < 1$ and $H = 1/\alpha$ with the kernel (3.75).

Example 3.6.10. The increment process $X_n = X(n) - X(n-1)$, $n \geq 1$, of the harmonizable fractional stable motion $\mathbf{X}$ of Example 3.5.4 is generated by a positive flow. Indeed, we have

$$X_n = \int_{\Omega'} \int_0^\infty f_n(\omega', x)\, M(d\omega', dx),\ n = 1, 2, \dots,$$

with

$$f_n(\omega', x) = \frac{\big(\cos((n-1)x) - \cos nx\big) G_1(\omega', x) + \big(\sin nx - \sin((n-1)x)\big)\, G_2(\omega', x)}{x^{H+1/\alpha}}.$$

Elementary trigonometry shows that

$$f_n(\omega', x) = 2\sin((n-1/2)x)\Big[\cos(x/2) G_2(\omega', x) - \sin(x/2) G_1(\omega', x)\Big].$$

For every x that is irrational relative to π, the trajectory of the sequence $\big(\sin((n - 1/2)x),\ n = 1, 2, \ldots\big)$ visits the interval $(1/2, 1)$ (say) at a linear frequency. This implies that (3.98) holds for every sequence $(w_n,\ n = 1, 2, \ldots)$ in $\mathcal{W}$ (with $w_1 = 1$), since the probabilities $p_n = w_n - w_{n+1}$, $n = 1, 2, \ldots$, and $p_\infty = \lim_{n\to\infty} w_n$ have infinite mean, as in the proof of Theorem 2.4.11.

Example 3.6.11. The increment process $X_n = X(n) - X(n-1)$, $n \geq 1$, of the FBM-local time fractional stable motion $\mathbf{X}$ of Example 3.5.5 is generated by a conservative null flow. To see this, notice that we can write

$$X_n = \int_{\Omega'} \int_{\mathbb{R}} (l(x, n+1)(\omega') - l(x, n)(\omega'))M(d\omega', dx),\ n \geq 0\,. \tag{3.102}$$

Note that

$$\sum_{n=0}^{m} \big[l(x, n+1)(\omega') - l(x, n)(\omega')\big] = l(x, m+1)(\omega') \to \infty \quad \text{as } m \to \infty$$

outside of a subset of $\Omega' \times \mathbb{R}$ of measure 0. By part (i) of Theorem 3.6.7, this implies that in the case $0 < \alpha < 1$, the FBM-H-local time fractional SαS noise is generated by a conservative flow. An appeal to Proposition 3.6.8 shows that the same is true for all $0 < \alpha < 2$.

In order to prove that the FBM-H-local time fractional SαS noise is generated by a null flow, we use, once again, Theorem 3.6.7 and Proposition 3.6.8. It is enough to exhibit a sequence $(w_n) \in \mathcal{W}$ such that

$$\sum_{n=0}^{\infty} w_n \big[l(x, n+1)(\omega') - l(x, n)(\omega')\big] < \infty \tag{3.103}$$

for $\mathbf{P}' \times$ Leb-almost every (ω', x). We claim that the sequence $w_n = (1+n)^{-\theta}$ with $\theta \in (1 - H, 1]$ satisfies (3.103). To see this, it is clearly enough to find a strictly positive measurable function g such that

$$\mathbf{E}' \int_{\mathbb{R}} g(x) \sum_{n=0}^{\infty} w_n \big[l(x, n+1)(\omega') - l(x, n)(\omega')\big]\, dx < \infty. \tag{3.104}$$

Note that by (10.11),

$$\mathbf{E}' \int_{\mathbb{R}} g(x) \sum_{n=0}^{\infty} w_n \big[l(x, n+1)(\omega') - l(x, n)(\omega')\big]\, dx = \sum_{n=0}^{\infty} w_n \int_n^{n+1} \mathbf{E}' g(B_H(t))dt\,.$$

Choose $g(x) = \exp(-x^2/2)$, $x \in \mathbb{R}$. Then for all $t \geq 0$, we have

$$\mathbf{E}' g(B_H(t)) = \frac{1}{(1 + t^{2H}\sigma^2)^{1/2}}\,,$$

where $\sigma^2 = \operatorname{Var} B_H(1)$. Then the left-hand side of (3.104) is

$$\sum_{n=0}^{\infty} w_n \int_n^{n+1} \frac{dt}{(1+t^{2H}\sigma^2)^{1/2}} \le \sum_{n=0}^{\infty} w_n \frac{1}{(1+n^{2H}\sigma^2)^{1/2}} < \infty \tag{3.105}$$

by the choice of θ. Hence (3.103) follows, and so the process is generated by a null flow.

In particular, we have learned from the last three examples that the increment processes of the linear fractional stable motion, of the harmonizable fractional stable motion, and of the FBM-local time fractional stable motion are different, and hence the three motions themselves are different processes.

3.7 Comments on Chapter 3

Comments on Section 3.1

An encyclopedic treatment of finite-dimensional infinitely divisible random vectors and Lévy processes can be found in Sato (1999).

A thorough discussion of α-stable processes, $0 < \alpha < 2$, is in Samorodnitsky and Taqqu (1994).

Comments on Section 3.2

The theory of infinitely divisible random measures and stochastic integrals with respect to these measures was presented in the fullest generality in Rajput and Rosiński (1989).

Comments on Section 3.4

Early work on series representations of infinitely divisible random variables often dealt with the so-called *shot noise models*, representing sums of certain impulses (usually nonnegative) generated by Poisson arrivals; see, e.g., Vervaat (1979). An approach to the series representations focused on ordering the underlying Poisson jumps by their size was introduced in Ferguson and Klass (1972). An explicit form of such a representation in the stable case was given in LePage et al. (1981). A very general discussion of series representations is in Rosiński (1990).

Comments on Section 3.5

The moving-average representations of the fractional Brownian motion originated with Mandelbrot and Van Ness (1968); see also Section 7.2.1 in Samorodnitsky and Taqqu (1994). The origins of the harmonizable representation of the fractional Brownian motion are in Kolmogorov (1940) and Yaglom (1955).

The fractional Brownian motion on $[0, \infty)$ can also be represented as a stochastic integral on compact intervals, of the form

$$X(t) = \int_0^t K_H(t,x)\, B(dx), \ t \ge 0\,,$$

where $K_H : \{(t,x) : 0 \le x \le t\} \to \mathbb{R}$ is a kernel such that $K_H(t,\cdot) \in L^2(\lambda)$ for each $t \ge 0$, and B is the standard Brownian motion; see Decreusefond and Üstünel (1999).

The FBM-local time fractional stable motion was first introduced in Cohen and Samorodnitsky (2006). One can also obtain new classes of self-similar stationary increment SαS processes using the local times of stable Lévy processes, as was shown in Dombry and Guillotin-Plantard (2009).

More examples of self-similar stable processes with stationary increments are in Example 8.4.1 and Example 9.6.3.

Comments on Section 3.6

Integral representations of stationary infinitely divisible processes are discussed in Kabluchko and Stoev (2016). The line of work connecting the recurrence properties of nonsingular flows to the structure of stationary α-stable processes originated with the paper Rosiński (1995). The role of the positive–null decomposition in the same context was pointed out in Samorodnitsky (2005). A very general treatment for general stationary infinitely divisible processes without a Gaussian component was given in Roy (2007).

3.8 Exercises to Chapter 3

Exercise 3.8.1. *Show that the gamma random variable and the negative binomial random variable are all infinitely divisible by computing their characteristic triples.*

Exercise 3.8.2. *Let X be a one-dimensional infinitely divisible random variable with a characteristic triple (σ^2, μ, b). Assume that X has a finite variance. Show that*

$$\mathrm{Var}(X) = \sigma^2 + \int_{\mathbb{R}} x^2 \, \mu(dx) \, .$$

Exercise 3.8.3. *Let $\mathbf{X} = (X^{(1)}, \ldots, X^{(d)})$ be an infinitely divisible random vector with characteristic triple $\big(\Sigma, \mu, \mathbf{b}\big)$, and let $A \cup B = \{1, \ldots, d\}$ be a partition of $\{1, \ldots, d\}$ into two nonempty parts. Prove, using the uniqueness of the characteristic triple, that the random vectors $(X^{(j)}, j \in A)$ and $(X^{(j)}, j \in B)$ are independent if and only if the following two conditions hold:*

$$\Sigma_{j_1 j_2} = 0 \ \textit{for each } j_1 \in A, \, j_2 \in B;$$

$$\mu\left(\mathbf{x} = (x^{(1)}, \ldots, x^{(d)}) : \sum_{j \in A} (x^{(j)})^2 > 0 \ \textit{and} \ \sum_{j \in B} (x^{(j)})^2 > 0\right) = 0 \, .$$

Exercise 3.8.4 (Rosiński and Żak (1997)). *Let (X_1, X_2) be an infinitely divisible random vector. Suppose that neither the one-dimensional Lévy measure of X_1 nor the one-dimensional Lévy measure of X_2 puts any mass on the set $\{2\pi k,\ k \in \mathbb{Z}\}$. Prove that if both*

$$E\big(e^{i(X_1+X_2)}\big) = E\big(e^{iX_1}\big)E\big(e^{iX_2}\big) \ \text{ and } \ E\big(e^{i(X_1-X_2)}\big) = E\big(e^{iX_1}\big)E\big(e^{-iX_2}\big),$$

then X_1 and X_2 are independent.

Exercise 3.8.5. *Prove Proposition 3.1.3.*

Exercise 3.8.6. *Consider the following condition on a measure on $\mathbb{R}^T$: for every countable subset T_1 of T, there is a countable set T_2 of T such that $T_1 \subset T_2$, and*

$$\mu\Big(\mathbf{x} \in \mathbb{R}^T : x(t) = 0 \ \text{for all } t \in T_1\Big) \tag{3.106}$$

$$= \mu\Big(\mathbf{x} \in \mathbb{R}^T : x(t) = 0 \ \text{for all } t \in T_1, \text{ but not for all } t \in T_2\Big).$$

Prove that a measure satisfies this condition if and only if it satisfies Condition 3.1.5 when T is countable, and Condition 3.1.6 if T is uncountable.

Exercise 3.8.7. *Prove that a Lévy measure on a countable set T is necessarily σ-finite.*

Exercise 3.8.8. *Let T be an arbitrary set, and $\big(X(t),\ t \in T\big)$ a stochastic process whose finite-dimensional distributions consist of i.i.d. components such that $X(t) \stackrel{d}{=} X$ for each $t \in T$, with X an infinitely divisible random variable. Prove that $\big(X(t),\ t \in T\big)$ is an infinitely divisible stochastic process, and express its Lévy measure in terms of the one-dimensional Lévy measure of X. Show that if the latter does not vanish and T is uncountable, then the Lévy measure of the infinitely divisible process $\big(X(t),\ t \in T\big)$ is not σ-finite.*

Exercise 3.8.9. *The notion of a compound Poisson random variable of Example 3.1.1 has an extension to infinitely divisible processes. Let $\big(Y_i(t),\ t \in T\big)$, $i = 1, 2, \ldots$, be a sequence of i.i.d. stochastic processes on T, independent of a mean-λ Poisson random variable N. Prove that*

$$X(t) = \sum_{i=1}^{N} Y_i(t),\ t \in T,$$

is an infinitely divisible stochastic process with a characteristic triple $(0, \lambda F_{\mathbf{Y}}, \mathbf{b})$, where $F_{\mathbf{Y}}$ is the law of the process $\mathbf{Y}$ on $\mathbb{R}^T$, and $b(t) = \lambda E[\![Y_1(t)]\!]$, $t \in T$. The process $\mathbf{X}$ is then a compound Poisson process.

Exercise 3.8.10. *Prove Proposition 3.1.10.*

Exercise 3.8.11. *Prove (3.13).*

Exercise 3.8.12. *(i) Prove that an α-stable process is strictly α-stable if and only if*

$$\mathbf{b} = \int_{\mathbb{R}^T} [\![\mathbf{x}]\!]\, \mu(d\mathbf{x}) \quad \text{if } 0 < \alpha < 1;$$

Lévy measure μ is symmetric if $\alpha = 1$;

$$\mathbf{b} = \int_{\mathbb{R}^T} \big([\![\mathbf{x}]\!] - \mathbf{x}\big)\, \mu(d\mathbf{x}) \quad \text{if } 1 < \alpha < 2.$$

(ii) Prove that an α-stable process is symmetric α-stable if and only if the Lévy measure μ is symmetric and $\mathbf{b} = 0$.

Exercise 3.8.13. *Prove that the process $(X(t),\ t \in \mathbb{R})$ constructed in Example 3.2.3 is continuous in probability, that it has stationary and independent increments, and that $X(1)$ is an infinitely divisible random variable with the characteristic triple (σ^2, ρ, b).*

Exercise 3.8.14. *Prove that the representation (3.70) of an SαS process can be generalized in the following way: the Rademacher sequence (ε_n) can be replaced by any other i.i.d. sequence (W_n) of symmetric random variables such that $E|W_1|^\alpha = 1$.*

Note that this allows the choice of (W_n) being zero-mean normal and shows that an SαS process of the form (3.64) is, distributionally, a mixture of zero-mean Gaussian processes.

Exercise 3.8.15. *Prove the comparison property (3.45).*

Exercise 3.8.16. *Let L be a Lévy motion and $f \in L_0(L)$. Prove that for every $s \in \mathbb{R}$, one has $f(\cdot + s) \in L_0(L)$ and*

$$\int_{\mathbb{R}} f(x+s)\, L(dx) \stackrel{d}{=} \int_{\mathbb{R}} f(x)\, L(dx)\,.$$

Exercise 3.8.17. *(i) Let B be a Brownian motion on $\mathbb{R}$ and $f \in L^2\big(\lambda, \mathbb{R}\big)$. Prove that for every $c > 0$,*

$$\int_{\mathbb{R}} f(x)\, B(dx) \stackrel{d}{=} \int_{\mathbb{R}} c^{1/2} f(cx)\, B(dx)\,.$$

(ii) Let $0 < \alpha < 2$, and let L be an SαS Lévy motion on $\mathbb{R}$. Let $f \in L^\alpha\big(\lambda, \mathbb{R}\big)$. Prove that for every $c > 0$,

$$\int_{\mathbb{R}} f(x)\, L(dx) \stackrel{d}{=} \int_{\mathbb{R}} c^{1/\alpha} f(cx)\, L(dx)\,.$$

Exercise 3.8.18. *Let M be an infinitely divisible random measure on a measurable space $(S, \mathcal{S})$, and let $f(t,\cdot) \in L_0(M)$ for each $t \in T$. Let m_0 be the σ-finite measure on $(S, \mathcal{S})$ given in (3.14). If T is countable, assume that*

$$m_0\Big(s \in S : f(t,s) = 0 \text{ for all } t \in T\Big) = 0\,.$$

If T is uncountable, assume that for every countable subset T_1 of T such that

$$m_0\Big(s \in S : f(t,s) = 0 \text{ for all } t \in T_1\Big) > 0,$$

there is $t_0 \in T_1^c$ such that

$$m_0\Big(s \in S : f(t,s) = 0 \text{ for all } t \in T_1, f(t_0,s) \neq 0\Big) > 0\,.$$

Prove that the measure $\mu_{\mathbf{X}}$ in (3.52) is a Lévy measure on $\mathbb{R}^T$ (and not only a weak Lévy measure).

Exercise 3.8.19. *The following is a convenient construction of a Poisson random measure with a finite mean measure. Let m be a finite measure on a measurable space $(S,\mathcal{S})$. Let $X_1, X_2, \ldots$ be i.i.d. S-valued random variables with a common law $m/m(S)$, independent of a Poisson random variable K with mean $m(S)$. For $B \in \mathcal{S}$, let*

$$M(B) = \sum_{k=1}^{K} \mathbf{1}(X_k \in B)\,,$$

the number of points among $X_1, \ldots, X_K$ that are in the set B. Prove that $\big(M(B), B \in \mathcal{S}\big)$, is a Poisson random measure with mean measure m.

Exercise 3.8.20. *Prove the claim of Example 3.3.14.*

Exercise 3.8.21. *Let $M^{(1)}$ and $M^{(2)}$ be independent centered Gaussian random measures on $(S,\mathcal{S})$ with control measure m each. Let $f, g : S \to \mathbb{R}$ be measurable. Prove that $M_f^{(1)} + M_g^{(2)}$ and $M_g^{(1)} - M_f^{(2)}$ are, once again, independent centered Gaussian random measures on $(S,\mathcal{S})$ with control measure $\tilde{m} \ll m$ each, where $d\tilde{m}/dm = f^2 + g^2$.*

Exercise 3.8.22. *Let $0 < H < 1$, and let $Q_t = \big(Q_t(x),\, x \in \mathbb{R}\big)$, $t \geq 0$, be a family of kernels on $\mathbb{R}$ with $Q_t \in L^2\big(\lambda, \mathbb{R}\big)$ for $t \geq 0$ satisfying (3.72). Assume that for each $t > 0$, Q_t is continuous at each $x \notin \{0, t\}$. Prove that there are real numbers c_1, c_2 such that $Q_t(x) = g_t(c_1, c_2; H; x)$ for all $t \geq 0$ and $x \notin \{0, t\}$, where $g_t(c_1, c_2; H; \cdot)$ is given by (3.74) if $H \neq 1/2$ and by (3.75) if $H = 1/2$.*

Exercise 3.8.23. *Complete the proof of Proposition 3.5.3 by considering the case $c_1^{(1)} c_2^{(1)} = c_1^{(2)} c_2^{(2)} = 0$.*

Exercise 3.8.24. *Let $\mathbf{X} = \big(X_n,\, n \in \mathbb{Z}\big)$ be a stationary infinitely divisible process without a Gaussian component. Prove that the integral representation of $\mathbf{X}$ given in Theorem 3.3.12 has the form (3.91) (in particular, the random measure M has constant local characteristics, i.e., it is a homogeneous infinitely divisible random measure).*

Chapter 4
Heavy Tails

4.1 What Are Heavy Tails? Subexponentiality

When we talk about the "tails" of a one-dimensional random variable X, we usually think about probabilities of the type $P(X > x)$ and $P(X < -x)$ for a large positive x, with the appropriate meaning of "right tail" and "left tail." If $(X(t),\, t \in \mathbb{R})$ or $(X_n,\, n \in \mathbb{Z})$ is a stationary stochastic process, the kind of marginal tails the process has may significantly impact the way memory expresses itself in the process. Particularly important is the distinction between stochastic processes with "light tails" and those with "heavy tails."

Defining the notions of light tails and heavy tails precisely is neither possible nor actually necessary, since the distinction varies from application to application. Let us begin with a nonnegative random variable X, so that we have only one tail to worry about. Let F be the c.d.f. of X. Intuitively, if $\overline{F}(x) = P(X > x)$ decays fast as $x \to \infty$, then X has a light tail, whereas a slow decay of $\overline{F}(x)$ means that X has a heavy tail. If

$$Ee^{\theta X} < \infty \text{ for some } \theta > 0,$$

then the tail of X decays at least exponentially fast at infinity, and in such situations, we will always say that the tail is light. On the other hand, in some cases the much weaker finite variance assumption $EX^2 < \infty$ is adequate enough to say that the tail of X is light.

Somewhat analogously, if the infinite second moment assumption $EX^2 = \infty$ holds, then we say that the tail of X is heavy, but in some cases, much weaker assumptions on X suffice to induce us to say that X has a heavy tail. As a rule, most formal definitions of heavy tails impose a certain regularity on the tail function $\overline{F}$ besides its slow decay. We begin with one of the broadest available definitions.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_4

Definition 4.1.1. A nonnegative random variable X, or its distribution F, is called *subexponential* if $\overline{F}(x) > 0$ for all $x \geq 0$ and

$$\lim_{x\to\infty} \frac{\overline{F \star F}(x)}{\overline{F}(x)} = 2\,. \tag{4.1}$$

Note that this simply says that X is subexponential if

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} = 2\,, \tag{4.2}$$

where X_1 and X_2 are i.i.d. copies of X.

Unless stated otherwise, in the rest of this section, X_1 and X_2 are assumed to be i.i.d. copies of a given random variable X. It is clear that $X_1 + X_2 \geq \max(X_1, X_2)$ for nonnegative random variables. Furthermore, we also have

$$\begin{aligned} P(\max(X_1, X_2) > x) &= 2P(X > x) - \big(P(X > x)\big)^2 \\ &\sim 2P(X > x) \end{aligned}$$

as $x \to \infty$. In particular, for every nonnegative random variable X with $P(X > x) > 0$ for all $x \geq 0$, we have

$$\liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(\max(X_1, X_2) > x)} = \liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{2P(X > x)} \geq 1\,. \tag{4.3}$$

We also see that an alternative way of defining subexponentiality is by

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(\max(X_1, X_2) > x)} = 1\,. \tag{4.4}$$

That is, for a subexponential random variable X, the tail of the sum of two independent copies of X is equivalent to the tail of the maximum of these random variables. Note a very important message in this sentence: for a subexponential random variable, such a sum becomes very large because one of the two terms in this sum becomes as large as required for the entire sum (and not, say, because both terms in the sum become about half of the total value each)!

A quintessential example of subexponential random variables is that of regularly varying random variables.

Definition 4.1.2. A nonnegative random variable X, or its distribution F, is called *regularly varying* if $\overline{F}(x) > 0$ for all $x \geq 0$ and there is $\alpha \geq 0$ such that for every $b > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(bx)}{\overline{F}(x)} = b^{-\alpha}\,. \tag{4.5}$$

We discuss regularly varying random variables in detail in Section 4.2.

Proposition 4.1.3. *A regularly varying random variable is subexponential.*

Proof. Let $0 < \varepsilon < 1$. We can write

$$\begin{aligned} P(X_1 + X_2 > x) &\le P(X_1 > (1-\varepsilon)x \text{ or } X_2 > (1-\varepsilon)x) \\ &+ P\big(X_1 + X_2 > x,\, X_1 \le (1-\varepsilon)x,\, X_2 \le (1-\varepsilon)x\big) \\ &\le 2P(X > (1-\varepsilon)x) + P(X_1 > \varepsilon x,\, X_2 > \varepsilon x) \\ &= 2P(X > (1-\varepsilon)x) + \big(P(X > \varepsilon x)\big)^2 . \end{aligned}$$

By the definition of a regularly varying random variable,

$$\frac{P(X > (1-\varepsilon)x)}{P(X > x)} \to (1-\varepsilon)^{-\alpha}, \quad \frac{P(X > \varepsilon x)}{P(X > x)} \to \varepsilon^{-\alpha}$$

as $x \to \infty$. Therefore,

$$\limsup_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \le 2(1-\varepsilon)^{-\alpha} .$$

Letting $\varepsilon \downarrow 0$, we obtain

$$\limsup_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \le 2,$$

which, according to (4.3), implies subexponentiality. □

Next, we discuss the basic properties of subexponential random variables. The property described in the following proposition is often called the *long tail property*.

Proposition 4.1.4. *If X is a subexponential random variable, then*

$$\lim_{x\to\infty} \frac{P(X > x + y)}{P(X > x)} = 1 \tag{4.6}$$

uniformly in y over compact sets.

Proof. By monotonicity, it is enough to check (4.6) for a fixed negative y. Suppose that, to the contrary,

$$\limsup_{x\to\infty} \frac{P(X > x + y)}{P(X > x)} = 1 + \rho > 1 .$$

Then we can write

$$\begin{aligned} P(X_1 + X_2 > x) &\ge P\Big(\max(X_1, X_2) > x \text{ or } x + y < X_1 \le x,\, X_2 \ge -y\Big) \\ &\ge 2P(X > x) + P(x + y < X \le x)P(X \ge -y) - o\Big(P(X > x)\Big) , \end{aligned}$$

implying that

$$\limsup_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \geq 2 + \rho P(X \geq -y) > 2\,.$$

This contradicts the assumption of subexponentiality. □

The long tail property of subexponential random variables has the following corollary, which says that the tail of a subexponential random variable decays more slowly than any exponential function (and hence explains why a random variable satisfying (4.1) is called "subexponential").

Corollary 4.1.5. *If X is a subexponential random variable, then for every $\varepsilon > 0$,*

$$\lim_{x\to\infty} e^{\varepsilon x} P(X > x) = \infty\,. \tag{4.7}$$

Proof. By Proposition 4.1.4, we know that for every $0 < \rho < 1$, there is $x_0 > 0$ such that for every $x \geq x_0$, we have

$$P(X > x + 1) \geq \rho P(X > x)\,.$$

Iterating, we see that for every $n \geq 1$,

$$P(X > x_0 + n) \geq \rho^n P(X > x_0)\,.$$

By monotonicity, this means that there is a finite positive constant C such that for all $x \geq x_0$,

$$P(X > x) \geq C e^{(\log \rho)x}\,.$$

This shows that (4.7) holds for every $\varepsilon > -\log \rho$. Since ρ can be taken arbitrarily close to 1, (4.7) holds for every $\varepsilon > 0$. □

We have said previously that the sum of two independent copies of a subexponential random variable becomes large only because one of the two terms becomes large. The next lemma formalizes this statement.

Lemma 4.1.6. *If X is a subexponential random variable, then*

$$\lim_{M\to\infty} \limsup_{x\to\infty} \frac{P(X_1 + X_2 > x,\, X_1 > M,\, X_2 > M)}{P(X > x)} = 0\,.$$

Proof. For every $M > 0$, we have by the long tail property,

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x,\, X_1 \leq M)}{P(X > x)} = P(X \leq M)\,,$$

so that

$$\lim_{M\to\infty}\lim_{x\to\infty}\frac{P(X_1+X_2>x,\ \min(X_1,X_2)\le M)}{P(X>x)}=2\,,$$

and the claim of the lemma follows from the definition of subexponentiality. □

It turns out that tail equivalence to a subexponential random variable implies subexponentiality.

Proposition 4.1.7. *Let X and Y be nonnegative random variables. If X is subexponential and $P(Y>x)/P(X>x)\to c\in(0,\infty)$ as $x\to\infty$, then Y is subexponential as well.*

Proof. Denote by F_X and F_Y the corresponding distributions of X and Y. For every $\varepsilon>0$, there is $x_\varepsilon>0$ such that for all $x\ge x_\varepsilon$, $P(Y>x)/P(X>x)\le c(1+\varepsilon)$.

Let $M\ge x_\varepsilon$, and denote by $y_M>x_\varepsilon$ a positive number such that for all $x\ge y_M$, $P(X>x-M)/P(X>x)\le 1+\varepsilon$. For $x>2M$, we have

$$P(Y_1+Y_2>x)=2P(Y_1+Y_2>x,\ Y_1\le M)+P(Y_1+Y_2>x,\ Y_1>M,\ Y_2>M)\,.$$

Now for all $x\ge y_M+M$,

$$P(Y_1+Y_2>x,\ Y_1\le M)=\int_{[0,M]}P(Y>x-y)\,F_Y(dy)$$

$$\le c(1+\varepsilon)\int_{[0,M]}P(X>x-y)\,F_Y(dy)\le c(1+\varepsilon)^2P(X>x)\,P(Y\le M)\,.$$

Similarly,

$$P(Y_1+Y_2>x,\ Y_1>M,\ Y_2>M)=\int_{(M,\infty)}P(Y>\max(x-y,M))\,F_Y(dy)$$

$$\le c(1+\varepsilon)\int_{(M,\infty)}P(X>\max(x-y,M))\,F_Y(dy)$$

$$=c(1+\varepsilon)\int_{(M,\infty)}P(Y>\max(x-y,M))\,F_X(dy)$$

$$\le c^2(1+\varepsilon)^2\int_{(M,\infty)}P(X>\max(x-y,M))\,F_X(dy)$$

$$=c^2(1+\varepsilon)^2P(X_1+X_2>x,\ X_1>M,\ X_2>M)\,.$$

Therefore, for all $M\ge x_\varepsilon$,

$$\limsup_{x\to\infty}\frac{P(Y_1+Y_2>x)}{P(X>x)}\le 2c(1+\varepsilon)^2P(Y\le M)$$

$$+c^2(1+\varepsilon)^2\limsup_{x\to\infty}\frac{P(X_1+X_2>x,\ X_1>M,\ X_2>M)}{P(X>x)}.$$

Applying Lemma 4.1.6, we obtain

$$\limsup_{x\to\infty} \frac{P(Y_1 + Y_2 > x)}{P(X > x)} \le 2c(1+\varepsilon)\,.$$

Since this is true for all $\varepsilon > 0$, we conclude that

$$\limsup_{x\to\infty} \frac{P(Y_1 + Y_2 > x)}{P(Y > x)} \le 2\,.$$

By (4.3), this means that Y is indeed subexponential. □

The tail equivalence (4.2) with two random variables in the definition of a subexponential random variable implies an analogous tail equivalence with an arbitrary number of terms in the sum.

Proposition 4.1.8. *Let X be a subexponential random variable. Then for every $k = 1, 2, \ldots$,*

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_k > x)}{P(X > x)} = k\,. \tag{4.8}$$

Proof. We begin by proving (4.8) for k of the form $k = 2^m$, $m = 1, 2, \ldots$, and we proceed by induction on m. For $m = 1$, the statement is the definition of subexponentiality (4.2). Assume that the statement holds for some $m = 1, 2, \ldots$. Then by Proposition 4.1.7, $Y^{(m)} := X_1 + \ldots + X_{2^m}$ is subexponential, so that by the induction hypothesis,

$$\frac{P(X_1 + \ldots + X_{2^{m+1}} > x)}{P(X > x)}$$

$$= \frac{P(Y_1^{(m)} + Y_2^{(m)} > x)}{P(Y^{(m)} > x)} \frac{P(Y^{(m)} > x)}{P(X > x)} \to 2 \cdot 2^m = 2^{m+1}\,,$$

completing the induction step. Next, we prove (4.8) for a general k, and since the statement has already been established for k of the form $k = 2^m$, $m = 1, 2, \ldots$, it is enough to verify the backward induction step: if (4.8) holds for some $k \ge 2$, then it also holds for $k - 1$. Since by the inclusion–exclusion formula, we clearly have

$$\liminf_{x\to\infty} \frac{P(X_1 + \ldots + X_{k-1} > x)}{P(X > x)} \ge k - 1\,,$$

we need to prove only the appropriate asymptotic upper bound. However,

$$P(X_1 + \ldots + X_k > x)$$

$$\ge P(X_1 + \ldots + X_{k-1} > x) + P(X_k > x) - o\Big(P(X_1 + \ldots + X_{k-1} > x)\Big)\,,$$

so that by the assumption of the backward induction,

$$k = \lim_{x\to\infty} \frac{P(X_1 + \ldots + X_k > x)}{P(X > x)}$$
$$\geq \limsup_{x\to\infty} \frac{P(X_1 + \ldots + X_{k-1} > x)}{P(X > x)} + 1\,,$$

from which the required upper bound

$$\limsup_{x\to\infty} \frac{P(X_1 + \ldots + X_{k-1} > x)}{P(X > x)} \leq k - 1$$

follows. This shows the backward induction step and hence completes the proof. □

The backward induction step used in the proof of the last proposition immediately establishes the following corollary.

Corollary 4.1.9. *Let X be a nonnegative random variable. If (4.8) holds for some $k \geq 2$, then X is subexponential (and hence (4.8) holds for all $k \geq 2$).*

The result of Proposition 4.1.8 is a tail equivalence statement, in the sense that it compares the tail function of a subexponential random variable to the tail function of a sum of several independent copies of the random variable in the asymptotic sense. The next proposition provides a useful comparison of these tail functions, valid for all values of the argument.

Proposition 4.1.10. *Let X be a subexponential random variable. For every $\varepsilon > 0$, there is $K = K_\varepsilon \in (0, \infty)$ such that*

$$\frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} \leq K(1 + \varepsilon)^n \tag{4.9}$$

for all $x \geq 0$ and $n = 1, 2, \ldots$.

Proof. Clearly,

$$P(X_1 + X_2 > x) = P(X_1 + X_2 > x,\ \min(X_1, X_2) \leq x) + o\big(P(X > x)\big)\,,$$

implying that

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x,\ \min(X_1, X_2) \leq x)}{P(X > x)} = 2\,.$$

Further, it follows from (4.4) that

$$P(X_1 + X_2 > x,\ X_1 \leq x,\ X_2 \leq x) = P(X_1 + X_2 > x) - P(\max(X_1, X_2) > x)$$
$$= o\big(P(X > x)\big)\,.$$

These two facts together show that

$$\lim_{x\to\infty} \frac{P(X_1 \le x,\, X_1 + X_2 > x)}{P(X > x)} = 1\,. \tag{4.10}$$

Given $\varepsilon > 0$, choose $x_0 \ge 0$ such that

$$\frac{P(X_1 \le x,\, X_1 + X_2 > x)}{P(X > x)} \le 1 + \frac{\varepsilon}{2}$$

for all $x \ge x_0$, and let

$$K = \max\left(\frac{1}{P(X > x_0)},\, \frac{2}{\varepsilon(1+\varepsilon)}\right).$$

We will prove (4.9) by induction on n. The statement is trivially valid for $n = 1$. Assume that it is valid for some $n \ge 1$. Using the notation F for the distribution of X, we have for every $x \ge x_0$, by the induction hypothesis,

$$\begin{aligned} P(X_1 + \ldots + X_n + X_{n+1} > x) &= P(X > x) + \int_{[0,x]} P(X_1 + \ldots + X_n > x - y)\, F(dy) \\ &\le P(X > x) + K(1+\varepsilon)^n \int_{[0,x]} P(X > x - y)\, F(dy) \\ &= P(X > x) + K(1+\varepsilon)^n P(X_1 \le x,\, X_1 + X_2 > x) \\ &\le P(X > x) + K(1+\varepsilon)^n (1 + \varepsilon/2) P(X > x) \\ &\le K(1+\varepsilon)^{n+1} P(X > x), \end{aligned}$$

since $K \ge 2/(\varepsilon(1+\varepsilon))$. On the other hand, for $0 \le x < x_0$, since $K \ge 1/P(X > x_0)$,

$$P(X_1 + \ldots + X_n + X_{n+1} > x) \le 1 \le KP(X > x_0) \le K(1+\varepsilon)^{n+1} P(X > x)\,.$$

This completes the inductive argument and hence the proof of the proposition. □

A consequence of Proposition 4.1.10 is the following extension of the tail equivalence in Proposition 4.1.8 to the sum of a random number of terms.

Corollary 4.1.11. *Let N be a nonnegative integer-valued random variable independent of a sequence $X_1, X_2, \ldots$ of i.i.d. copies of a subexponential random variable X. If $Ee^{\theta N} < \infty$ for some $\theta > 0$, then*

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X > x)} = EN\,.$$

Proof. By Proposition 4.1.10 with $0 < \varepsilon \leq e^{\theta} - 1$, the dominated convergence theorem, and Proposition 4.1.8,

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X > x)} = \lim_{x\to\infty} \sum_{n=0}^{\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} P(N = n)$$

$$= \sum_{n=0}^{\infty} \lim_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} P(N = n)$$

$$= \sum_{n=0}^{\infty} n\, P(N = n) = EN\,.$$

□

Example 4.1.12. The most important application of Corollary 4.1.11 is to the compound Poisson random variables of Example 3.1.1. Let

$$Y = X_1 + \ldots + X_N, \quad N \text{ a mean } \lambda \text{ Poisson random variable}$$

and let $X_1, X_2, \ldots$ be i.i.d. random variables independent of N. Then subexponentiality of X implies

$$\lim_{x\to\infty} \frac{P(Y > x)}{P(X > x)} = \lambda\,.$$

Remark 4.1.13. The converse statement of Corollary 4.1.9 easily generalizes to the case of the sum of a random number of terms. For the converse statement, one does not even need the assumption of finite exponential moments of the random number of terms. Let N be a nonnegative integer-valued random variable independent of a sequence $X_1, X_2, \ldots$ of i.i.d. random variables. Assume that $P(N \geq 2) > 0$ and $EN < \infty$. If

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X > x)} = EN\,,$$

then X is subexponential. To see this, let $k \geq 2$ be such that $P(N = k) > 0$. By Corollary 4.1.9, it is enough to show that

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_k > x)}{P(X > x)} = k\,.$$

Indeed, if this fails, then we must have

$$\limsup_{x\to\infty} \frac{P(X_1 + \ldots + X_k > x)}{P(X > x)} > k\,,$$

since the obvious counterpart of (4.3) says that for every $n = 1, 2, \ldots$,

$$\liminf_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} \geq n\,.$$

In that case, we would have by Fatou's lemma

$$\begin{aligned}\limsup_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X > x)} &\geq P(N = k) \limsup_{x\to\infty} \frac{P(X_1 + \ldots + X_k > x)}{P(X > x)} \\ &\quad + \sum_{n\neq k} P(N = n) \liminf_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} \\ &> kP(N = k) + \sum_{n\neq k} nP(N = n) = EN\,,\end{aligned}$$

contradicting the assumption.

The long tail property (4.6) of subexponential random variables is a basic one, and one might conjecture that every nonnegative random variable with a long tail is subexponential. The conjecture is false, as the following example demonstrates.

Example 4.1.14 (Embrechts and Goldie (1980)). For $n = 1, 2, \ldots$, we set

$$\begin{aligned}a_n &= \max\{k = 1, 2, \ldots : (k + 1)! \leq n + 1\}, \\ k_n &= \max\{k = 1, 2, \ldots : (k + 1)! \leq (n + 1)a_n\}\,.\end{aligned}$$

The following properties of the two sequences are clear:

$$1 \leq a_n \uparrow \infty,\ k_n \geq a_n\,. \tag{4.11}$$

Let now X be a nonnegative random variable with the density

$$f_X(x) = \frac{1}{a_n(n + 1)!} \text{ for } (n + 1)! \leq x \leq (n + 1)! + na_n,\ n = 1, 2, \ldots,$$

and equal to zero otherwise. Clearly, f_X is a legitimate density, and

$$P(X > (n + 1)!) = \frac{1}{n!} \text{ for } n = 1, 2, \ldots.$$

We first check that X has a long tail. To this end, we need to check that for every $y > 0$,

$$\frac{\int_x^{x+y} f_X(t)\,dt}{P(X > x)} \to 0 \text{ as } x \to \infty. \tag{4.12}$$

Let $(n+1)! \le x < (n+2)!,\ n = 1, 2, \ldots$. Then

$$P(X > x) \ge P(X > (n+2)!) = \frac{1}{(n+1)!}\,,$$

while

$$\int_x^{x+y} f_X(t)\,dt \le y\frac{1}{a_n(n+1)!}\,,$$

so that the ratio in (4.12) is at most $y/a_n \to 0$ as $n \to \infty$, and (4.12) follows. Hence, X has a long tail.

Next, we check that X is not subexponential. We will check that

$$\lim_{n\to\infty} \frac{P(X_1 + X_2 > (n+1)! + na_n)}{P(X > (n+1)! + na_n)} = \infty\,. \tag{4.13}$$

To this end, observe that

$$\begin{aligned}
&P(X_1 + X_2 > (n+1)! + na_n)\\
&\ge P\Big((n+1)! < X_1 \le (n+1)! + na_n,\ X_1 + X_2 > (n+1)! + na_n\Big)\\
&= \frac{1}{a_n(n+1)!}\int_{(n+1)!}^{(n+1)!+na_n} P(X > (n+1)! + na_n - u)\,du\\
&= \frac{1}{a_n(n+1)!}\int_0^{na_n} P(X > x)\,dx \ge \frac{1}{a_n(n+1)!}\sum_{k=1}^{k_n}\int_{k!}^{(k+1)!} P(X > x)\,dx\\
&\ge \frac{1}{a_n(n+1)!}\sum_{k=1}^{k_n}\big((k+1)! - k!\big)P(X > (k+1)!)\\
&= \frac{1}{a_n(n+1)!}\sum_{k=1}^{k_n} k = \frac{k_n(k_n+1)}{2a_n(n+1)!}\,.
\end{aligned}$$

Since

$$P(X > (n+1)! + na_n) = P(X > (n+2)!) = \frac{1}{(n+1)!}\,,$$

we see that

$$\begin{aligned}
\frac{P(X_1 + X_2 > (n+1)! + na_n)}{P(X > (n+1)! + na_n)} &\ge \frac{k_n(k_n+1)}{2a_n}\\
&\ge \frac{a_n + 1}{2} \to \infty
\end{aligned}$$

by (4.11), which proves (4.13). Therefore, X is not subexponential, even though it has a long tail.

Remark 4.1.15. Somewhat surprisingly, the class of subexponential distributions turns out not to be closed under convolutions. That is, there exist independent nonnegative subexponential random variables, X and Y, such that the sum $X + Y$ does not have a subexponential distribution. An example showing this can be found in Leslie (1989).

It is useful to extend the notion of subexponentiality to not necessarily nonnegative random variables.

Definition 4.1.16. A real-valued random variable X, or its distribution F, is called subexponential if the positive part X_+ of X is subexponential according to Definition 4.1.1.

In fact, for a real-valued subexponential random variable X, the tail equivalence relation (4.2) still holds. To see this, note that

$$\limsup_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \le \limsup_{x\to\infty} \frac{P((X_1)_+ + (X_2)_+ > x)}{P(X_+ > x)} = 2,$$

and for every $M > 0$,

$$\begin{aligned} P(X_1 + X_2 > x) &\ge P(X_1 > x + M, X_2 > -M \text{ or } X_1 > -M, X_2 > x + M) \\ &= 2P(X_1 > x + M, X_2 > -M) - (P(X_1 > x + M))^2, \end{aligned}$$

so that

$$\begin{aligned} \liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} &\ge \liminf_{x\to\infty} 2\frac{P(X_1 > x + M, X_2 > -M)}{P(X > x)} \\ &= 2P(X > -M)\,. \end{aligned}$$

Letting $M \to \infty$, we obtain (4.2).

Most of the properties of nonnegative subexponential random variables extend easily to real-valued subexponential random variables. This includes the long tail property of Proposition 4.1.4, as well as Corollary 4.1.5 and Proposition 4.1.7. The statements of Proposition 4.1.8 and Corollary 4.1.11 are also true in this more general case. See Exercise 4.6.1.

The following statement gives an explicit criterion for membership in the class of subexponential distributions. It is due to Pitman (1980).

Theorem 4.1.17. *Let X be a nonnegative random variable with $P(X > x) > 0$ for all x, and set $g(x) = -\log P(X > x)$, $x \ge 0$. Suppose that there is $x_0 > 0$ such that on the interval (x_0, ∞), g is differentiable, and $g'(x)$ decreases to zero as x increases*

to infinity. If

$$\int_{x_0}^{\infty} \exp\{xg'(x) - g(x)\}g'(x)\,dx < \infty\,, \tag{4.14}$$

then X is subexponential.

Proof. It is easy to construct a random variable Y with $P(Y > x) = P(X > x)$ for $x > 2x_0$ such that Y satisfies the assumption of the theorem with $x_0 = 0$. By Proposition 4.1.7, it is enough to prove subexponentiality of Y. For notational simplicity, we write X instead of Y. Since

$$P(X_1 + X_2 > x) = P(X_1 \le x/2,\, X_1 + X_2 > x) + P(X_2 \le x/2,\, X_1 + X_2 > x) \\ + P(X_1 > x/2,\, X_2 > x/2)\,,$$

the claim of the theorem will follow once we check that

$$\lim_{x\to\infty} \frac{P(X_1 \le x/2,\, X_1 + X_2 > x)}{P(X > x)} = 1 \tag{4.15}$$

and

$$\lim_{x\to\infty} \frac{\big(P(X > x/2)\big)^2}{P(X > x)} = 0\,. \tag{4.16}$$

We can write the fraction in (4.15) as

$$\frac{P(X_1 \le x/2,\, X_1 + X_2 > x)}{P(X > x)} = \int_{[0,x/2]} e^{g(x)-g(x-y)} g'(y)\, e^{-g(y)}\, dy\,.$$

Notice that for every $y \ge 0$,

$$\mathbf{1}\big(y \le x/2\big) e^{g(x)-g(x-y)} = \mathbf{1}\big(y \le x/2\big) \exp\left\{\int_{x-y}^{x} g'(t)\,dt\right\} \to 1$$

as $x \to \infty$, since g' converges to zero. Furthermore, by the monotonicity of the derivative,

$$0 \le \mathbf{1}\big(y \le x/2\big) \exp\left\{\int_{x-y}^{x} g'(t)\,dt\right\} \\ \le \mathbf{1}\big(y \le x/2\big) \exp\{yg'(x-y)\} \le e^{yg'(y)}\,.$$

By (4.14) and the dominated convergence theorem, we obtain

$$\lim_{x\to\infty} \frac{P(X_1 \le x/2,\, X_1 + X_2 > x)}{P(X > x)} = \int_{[0,\infty)} g'(y)\, e^{-g(y)}\, dy = 1\,,$$

establishing (4.15). For (4.16), we can write, for every $a > 0$ and $x > 2a$, by the monotonicity of g',

$$\frac{\big(P(X > x/2)\big)^2}{P(X > x)} = e^{g(x)-2g(x/2)} = \exp\left\{\int_{x/2}^{x} g'(t)\,dt - \int_0^{x/2} g'(t)\,dt\right\}$$

$$= \exp\left\{\int_a^{x/2} \big(g'(t + x/2) - g'(t)\big)\,dt + \int_{x/2}^{x/2+a} g'(t)\,dt - \int_0^a g'(t)\,dt\right\}$$

$$\le \exp\left\{\int_{x/2}^{x/2+a} g'(t)\,dt - \int_0^a g'(t)\,dt\right\} \le \exp\big\{ag'(x/2) - g(a)\big\}\,.$$

Since the derivative g' converges to zero, we have

$$\limsup_{x\to\infty} \frac{\big(P(X > x/2)\big)^2}{P(X > x)} \le e^{-g(a)} = P(X > a)\,.$$

Letting $a \to \infty$ establishes (4.16). □

Example 4.1.18. If X is a (nondegenerate) normal random variable, then $Y = e^X$ is a lognormal random variable. It is easy to check, using Proposition 4.1.7 and Theorem 4.1.17, that a lognormal random variable is subexponential. It is equally easy to check that every random variable whose right tail satisfies

$$P(X > x) \sim \exp\big\{-cx^a(\log x)^\theta\big\} \text{ as } x \to \infty$$

for $c > 0$ is subexponential as well, under any of the following conditions:

- $0 < a < 1$ and $\theta \in \mathbb{R}$;
- $a = 0$ and $\theta > 0$;
- $a = 1$ and $\theta < 0$.

See Exercise 4.6.2.

4.2 Regularly Varying Random Variables

Regularly varying nonnegative random variables were introduced in Definition 4.1.2 as an example of subexponential random variables. In this section, we discuss this important class in more detail. The parameter $\alpha \ge 0$ in (4.5) is sometimes called the index of regular variation, or the exponent of regular variation. Sometimes, the expression "tail exponent" is used. When the tail exponent of a random variable is equal to zero, the random variable is said to have a *slowly varying tail*. The tail of such a random variable is particularly heavy.

Clearly, to say that a random variable is regularly varying with exponent $\alpha \geq 0$ is the same as to say that its distribution function F satisfies the requirement that $\overline{F}(x) = 1 - F(x)$ be regularly varying at infinity with exponent $-\alpha \leq 0$. We discuss general regularly varying functions in Section 10.5. By its definition, the tail distribution function $\overline{F}$ is not only regularly varying, but also monotone.

It follows immediately that a nonnegative random variable X is regularly varying with tail exponent α if and only if

$$P(X > x) = x^{-\alpha} L(x), \; x \geq 0, \tag{4.17}$$

for a positive slowly varying function L. When $\alpha = 0$, then L has to be nonincreasing. One of the reasons why random variables with regularly varying tails (or distributions of random variables with regularly varying tails) are popular is that the tail exponent α provides a linear scale of how heavy a tail is; the lower the value of α, the heavier the tail. One consequence of this is that the tail exponent of a nonnegative random variable largely determines which moments of the random variable are finite.

Proposition 4.2.1. *Let X be a nonnegative regularly varying random variable with tail index α. Then for $p > 0$,*

$$EX^p \text{ is } \begin{cases} < \infty \text{ if } p < \alpha, \\ = \infty \text{ if } p > \alpha. \end{cases}$$

Proof. Since we can write for $p > 0$

$$EX^p = p \int_0^\infty x^{p-1} P(X > x)\, dx\,, \tag{4.18}$$

the statement of the proposition follows from the comparison (10.33) of a regularly varying function and power functions. □

Remark 4.2.2. Note that a nonnegative regularly varying random variable with tail index α may or may not have a finite moment EX^α. In fact, finiteness of that moment or lack thereof is determined by further properties of the slowly varying function L in (4.17).

When a moment of a regularly varying random variable is infinite, it is an often very useful fact that the corresponding truncated moment is regularly varying as well.

Proposition 4.2.3. *Let X be a nonnegative regularly varying random variable with tail index α. Then for $p \geq \alpha$, the truncated pth moment*

$$m_p(x) = E(X \wedge x)^p, \; x \geq 0\,,$$

is regularly varying with exponent $p - \alpha$, and

$$\lim_{x\to\infty} \frac{E(X \wedge x)^p}{x^p P(X > x)} = \frac{p}{p-\alpha}\,,$$

with $p/0 = +\infty$.

Proof. Applying (4.18) to the random variable $X \wedge x$ gives us

$$E(X \wedge x)^p = p \int_0^x y^{p-1} P(X > y)\, dy$$

and the claim follows from Theorem 10.5.6. □

A related statement for upper truncated moments is in Exercise 4.6.4.

If X is a not necessarily nonnegative random variable, we say that it has a regularly varying right tail with tail exponent $\alpha \geq 0$ if its positive part X_+ is regularly varying with tail exponent α. Clearly, (4.17) still characterizes a regularly varying right tail, regardless of whether the random variable is nonnegative.

Regular variation of the tails of random variables is a very robust property, one that is preserved under various operations. We begin with a simple observation described in the following lemma.

Lemma 4.2.4. *Let X have a regularly varying right tail. Suppose that Y is independent of X and such that*

$$\lim_{x\to\infty} \frac{P(Y > x)}{P(X > x)} = A \geq 0\,.$$

Then

$$\lim_{x\to\infty} \frac{P(X + Y > x)}{P(X > x)} = 1 + A\,.$$

In particular, $X + Y$ also has a regularly varying right tail.

Proof. For every $0 < \varepsilon < 1$ and $x > 0$,

$$P(X + Y > x) \leq P(X > (1-\varepsilon)x) + P(Y > (1-\varepsilon)x) + P(X > \varepsilon x,\, Y > \varepsilon x)\,.$$

Therefore, by the assumption of regular variation and independence,

$$\begin{aligned}\limsup_{x\to\infty} \frac{P(X + Y > x)}{P(X > x)} &\leq \limsup_{x\to\infty} \frac{P(X > (1-\varepsilon)x)}{P(X > x)} + \limsup_{x\to\infty} \frac{P(Y > (1-\varepsilon)x)}{P(X > x)} \\ &\quad + \limsup_{x\to\infty} P(Y > \varepsilon x) \frac{P(X > \varepsilon x)}{P(X > x)} \\ &= (1-\varepsilon)^{-\alpha} + A(1-\varepsilon)^{-\alpha} + 0 \cdot \varepsilon^{-\alpha} = (1+A)(1-\varepsilon)^{-\alpha}\,.\end{aligned}$$

Letting $\varepsilon \to 0$, we conclude that

$$\limsup_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)} \le 1 + A\,.$$

For the matching lower bound, write for $\theta > 0$,

$$P(X+Y>x) \ge P\left(\left\{X > x+\theta, Y > -\theta\right\} \cup \left\{Y > x+\theta, X > -\theta\right\}\right)$$
$$\ge P(X > x+\theta, Y > -\theta) + P(Y > x+\theta, X > -\theta) - P(X > x, Y > x)\,.$$

Since a random variable with a regularly varying tail is subexponential, we conclude by the long tail property of subexponential random variables in Proposition 4.1.4 that

$$\begin{aligned}\liminf_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)} &\ge \liminf_{x\to\infty} P(Y>-\theta)\frac{P(X>x+\theta)}{P(X>x)} \\ &\quad + \liminf_{x\to\infty} P(X>-\theta)\frac{P(Y>x+\theta)}{P(X>x)} \\ &= P(Y>-\theta) + AP(X>-\theta)\,.\end{aligned}$$

Letting $\theta \to \infty$ completes the proof. □

The next statement shows that if X has a regularly varying right tail, then adding a random variable with a lighter tail does not change the tail of X even in the absence of independence.

Proposition 4.2.5. *Suppose that X has a regularly varying right tail and Y is a random variable such that*

$$\lim_{x\to\infty} \frac{P(|Y|>x)}{P(X>x)} = 0\,. \tag{4.19}$$

Then

$$\lim_{x\infty} \frac{P(X+Y>x)}{P(X>x)} = 1\,. \tag{4.20}$$

In particular, $X+Y$ also has a regularly varying right tail.

Proof. For every $0 < \varepsilon < 1$,

$$P(X > (1+\varepsilon)x) - P(Y < -\varepsilon x) \le P(X+Y>x) \le P(X > (1-\varepsilon)x) + P(Y > \varepsilon x)\,.$$

Therefore, by (4.19) and the regular variation of X,

$$\limsup_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)} \le (1-\varepsilon)^{-\alpha} + \limsup_{x\to\infty} \frac{P(|Y|>\varepsilon x)}{P(X>\varepsilon x)}\varepsilon^{-\alpha} = (1-\varepsilon)^{-\alpha}\,,$$

where α is the tail exponent of X. Now the upper bound in (4.20) follows by letting $\varepsilon \to 0$. The lower bound in (4.20) can be obtained in the same way. □

It turns out, further, that multiplying a random variable X with a regularly varying right tail by an independent positive random variable with a sufficiently light tail leaves the tail of X asymptotically unchanged apart from multiplying it by a positive constant.

Proposition 4.2.6. *Suppose that X has a regularly varying right tail with exponent $\alpha \geq 0$, and Y is a positive random variable such that*

$$EY^{\alpha+\varepsilon} < \infty \tag{4.21}$$

for some $\varepsilon > 0$. Then

$$\lim_{x\to\infty} \frac{P(YX > x)}{P(X > x)} = EY^{\alpha} . \tag{4.22}$$

In particular, YX also has a regularly varying right tail.

Proof. Let F_Y be the law of the random variable Y. Write

$$P(YX > x) = \int_0^\infty P(X > x/y)\, F_Y(dy) .$$

Let $y \geq 1$. By the Potter bounds of Corollary 10.5.8, there exist $C > 0$ and $a_\varepsilon > 0$ such that if $x/y \geq a_\varepsilon$, then

$$\frac{P(X > x/y)}{P(X > x)} \leq Cy^{\alpha+\varepsilon} . \tag{4.23}$$

On the other hand, if $x/y < a_\varepsilon$, then

$$\frac{P(X > x/y)}{P(X > x)} \leq \frac{1}{P(X > a_\varepsilon y)} .$$

This last expression, according to (10.33), is smaller than $y^{\alpha+\varepsilon}$ for all $y > y_0$, for some $y_0 \geq 1$, and hence is also bounded from above by $C_1 y^{\alpha+\varepsilon}$ for all $y \geq 1$ if we choose

$$C_1 = \frac{1}{P(X > a_\varepsilon y_0)} .$$

Therefore, (4.23) holds for all $y \geq 1$ and $x > 0$ if we increase the constant C. By the dominated convergence theorem, we conclude that

$$\lim_{x\infty} \frac{P(YX > x)}{P(X > x)} = \int_0^\infty \lim_{x\infty} \frac{P(X > x/y)}{P(X > x)} F_Y(dy)$$
$$= \int_0^\infty y^\alpha F_Y(dy) = EY^\alpha .$$

□

It is sometimes appropriate to discuss regular variation of both the left and the right tails of a random variable at the same time. The key notion is that of *balanced regularly varying tails*.

Definition 4.2.7. A real-valued random variable X, or its distribution F, is said to have balanced regularly varying tails with exponent $\alpha \geq 0$ if the nonnegative random variable $|X|$ is regularly varying with exponent α and for some $0 \leq p, q \leq 1$, $p + q = 1$,

$$\lim_{x\to\infty} \frac{P(X > x)}{P(|X| > x)} = p, \quad \lim_{x\to\infty} \frac{P(X < -x)}{P(|X| > x)} = q . \tag{4.24}$$

Note that the numbers p and q in Definition 4.2.7 describe the relative weights of the right and left tails of the random variable X. The right tail of X is actually guaranteed to be regularly varying only if $p > 0$, and the left tail does not need to be regularly varying unless $q > 0$. If, however, $0 < p < 1$, then both tails of X are regularly varying with exponent α, and moreover, asymptotically equivalent to each other.

Example 4.2.8. Recall that for $0 < \alpha < 2$, an α-stable random variable is an infinitely divisible random variable X with a characteristic triplet (σ^2, μ, b), where $\sigma = 0$, and the Lévy measure μ is of the form

$$\mu(dx) = \left[a_- |x|^{-(1+\alpha)} \mathbf{1}(x < 0) + a_+ x^{-(1+\alpha)} \mathbf{1}(x > 0)\right] dx .$$

Here $a_-, a_+ \geq 0$, $a_- + a_+ > 0$. For an α-stable random variable,

$$P(|X| > x) \sim \frac{a_- + a_+}{\alpha} x^{-\alpha} \text{ as } x \to \infty,$$

and (4.24) holds with

$$p = \frac{a_+}{a_- + a_+}, \quad q = \frac{a_-}{a_- + a_+} ;$$

see Samorodnitsky and Taqqu (1994). Therefore, α-stable random variables have balanced regularly varying tails. In particular, for an α-stable random variable, the absolute moment of order α is infinite, but all absolute moments of orders strictly smaller than α are finite.

A two-sided version of Proposition 4.2.6 is contained in the following easy corollary.

Corollary 4.2.9. *Let X have balanced regularly varying tails with exponent $\alpha \geq 0$, and let Y be independent of X such that for some $\varepsilon > 0$, $E|Y|^{\alpha+\varepsilon} < \infty$. Then*

$$\lim_{x\to\infty} \frac{P(|YX| > x)}{P(|X| > x)} = E|Y|^{\alpha}$$

and

$$\lim_{x\to\infty} \frac{P(YX > x)}{P(|YX| > x)} = p\frac{EY_+^{\alpha}}{E|Y|^{\alpha}} + q\frac{EY_-^{\alpha}}{E|Y|^{\alpha}},$$
$$\lim_{x\to\infty} \frac{P(YX < -x)}{P(|YX| > x)} = p\frac{EY_-^{\alpha}}{E|Y|^{\alpha}} + q\frac{EY_+^{\alpha}}{E|Y|^{\alpha}}.$$

In particular, YX also has balanced regularly varying tails with exponent α.

See Exercise 4.6.5.

An immediate consequence of Lemma 4.2.4 and Corollary 4.2.9 says that if $X_1, \ldots, X_n$ are i.i.d. random variables with balanced regularly varying tails with exponent $\alpha \geq 0$, independent of a sequence $Y_1, \ldots, Y_n$ of independent random variables such that for some $\varepsilon > 0$, $E|Y_i|^{\alpha+\varepsilon} < \infty$, $i = 1, \ldots, n$, then the sum $\sum_{i=1}^n Y_i X_i$ also has balanced regularly varying tails with exponent α, which are asymptotically equivalent to the tails of a generic representative X of the sequence $X_1, \ldots, X_n$.

This statement, however, can be greatly generalized. First of all, it remains true even if $Y_1, \ldots, Y_n$ are not independent. Second, we can even allow a certain type of dependence between the sequences $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$. Finally, and perhaps most importantly, the statement remains true even for infinite sequences (i.e., for $n = \infty$), under additional assumptions on the random variables $Y_1, Y_2, \ldots$.

The following theorem, stated here without proof, is a very general result of this type. The proof can be found in Hult and Samorodnitsky (2008).

Theorem 4.2.10. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with balanced regularly varying tails with exponent $\alpha > 0$. If $\alpha > 1$, we assume that $EX = 0$. Let $Y_1, Y_2, \ldots$ be a sequence of random variables such that $\sup_j |Y_j| > 0$ a.s.*

Assume that there is a filtration $(\mathcal{F}_j, j \geq 1)$ such that

$$\begin{aligned} &Y_j \text{ is } \mathcal{F}_j\text{-measurable, } X_j \text{ is } \mathcal{F}_{j+1}\text{-measurable,} \\ &\mathcal{F}_j \text{ is independent of } \sigma(X_j, X_{j+1}, \ldots), j \geq 1. \end{aligned} \tag{4.25}$$

Suppose that there is $0 < \varepsilon < \alpha$ *such that*

$$\sum_{n=1}^{\infty} E|Y_n|^{\alpha-\varepsilon} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} E|Y_n|^{\alpha+\varepsilon} < \infty, \quad \text{if} \quad \alpha \in (0,1) \cup (1,2), \tag{4.26}$$

$$E\left(\sum_{n=1}^{\infty} |Y_n|^{\alpha-\varepsilon}\right)^{\frac{\alpha+\varepsilon}{\alpha-\varepsilon}} < \infty, \quad \text{if} \quad \alpha \in \{1,2\}, \tag{4.27}$$

$$E\left(\sum_{n=1}^{\infty} |Y_n|^2\right)^{\frac{\alpha+\varepsilon}{2}} < \infty, \quad \text{if} \quad \alpha \in (2,\infty). \tag{4.28}$$

Then the series

$$S = \sum_{n=1}^{\infty} Y_n X_n$$

converges with probability 1. Moreover,

$$\lim_{x\to\infty} \frac{P(|S| > x)}{P(|X| > x)} = \sum_{n=1}^{\infty} E|Y_n|^{\alpha} \tag{4.29}$$

and

$$\lim_{x\to\infty} \frac{P(S > x)}{P(|X| > x)} = \sum_{n=1}^{\infty} \left[pE(Y_n)_+^{\alpha} + qE(Y_n)_-^{\alpha}\right]. \tag{4.30}$$

In particular, S has balanced regularly varying tails with exponent α.

Remark 4.2.11. Under the assumptions of the theorem, the series S converges absolutely if $\alpha \le 1$, and unconditionally if $\alpha > 1$; see Remark 1.4.2.

The condition (4.25) means that the sequence $Y_1, Y_2, \ldots$ is, in a sense, predictable with respect to the sequence $X_1, X_2, \ldots$. This framework includes, as a special case, the situation in which the sequence $Y_1, Y_2, \ldots$ is independent of the sequence $X_1, X_2, \ldots$, since in this case, one can choose $\mathcal{F}_j = \sigma\big(Y_n, n = 1, 2, \ldots, X_n, n < j\big)$ for $j = 1, 2, \ldots$.

It follows from Proposition 4.2.5 that even without the assumption of zero mean in the case $\alpha > 1$ in Theorem 4.2.10, the conclusions of that theorem will still be valid under the following additional assumption on the sequence $Y_1, Y_2, \ldots$:

$$\text{the series } S_Y = \sum_{n=1}^{\infty} Y_n \text{ converges and } \lim_{x\to\infty} \frac{P(|S_Y| > x)}{P(|X| > x)} = 0\,. \tag{4.31}$$

When the sequence $Y_1, Y_2, \ldots$ in Theorem 4.2.10 is actually deterministic, the result of the theorem gives sufficient conditions for the linear processes of Section 1.4 to consist of random variables with balanced regularly varying tails. We state these conditions as a corollary; it is an immediate consequence of Theorem 4.2.10. This corollary complements the result of Theorem 1.4.1.

Corollary 4.2.12. *Let $\ldots, X_{-1}, X_0, X_1, X_2, \ldots$ be i.i.d. random variables with balanced regularly varying tails with exponent $\alpha > 0$, and let $(\varphi_n, n \in \mathbb{Z})$ be real numbers such that*

$$\begin{array}{l} \sum_{n=-\infty}^{\infty} |\varphi_n|^{\alpha-\varepsilon} < \infty \text{ for some } 0 < \varepsilon < \alpha, \text{ if } \alpha \le 2, \\ \sum_{n=-\infty}^{\infty} |\varphi_n|^2 < \infty \qquad \text{if } \alpha > 2. \end{array} \tag{4.32}$$

If $\alpha > 1$, assume that either $EX = 0$ or (1.25) holds (i.e., the series $\sum \varphi_n$ converges). Then the series

$$S = \sum_{n=-\infty}^{\infty} \varphi_n X_n$$

converges with probability 1, and

$$\lim_{x\to\infty} \frac{P(|S| > x)}{P(|X| > x)} = \sum_{n=-\infty}^{\infty} |\varphi_n|^{\alpha}, \tag{4.33}$$

$$\lim_{x\to\infty} \frac{P(S > x)}{P(|X| > x)} = \sum_{n=-\infty}^{\infty} \left[p(\varphi_n)_+^{\alpha} + q(\varphi_n)_-^{\alpha}\right]. \tag{4.34}$$

Therefore, S has balanced regularly varying tails with exponent α.

Remark 4.2.13. It is clear that Theorem 4.2.10 can be applied in the situation of Corollary 4.2.12, even though the series in the latter is doubly infinite. Note that the convergence statement in Corollary 4.2.12 follows also from Theorem 1.4.1.

4.3 Multivariate Regularly Varying Tails

It is often desirable to have a notion of heavy tails for a random vector. In the context of a stochastic process, this corresponds to the notion of heavy tails for the finite-dimensional distributions of the process. A multivariate version of the subexponential tails of Section 4.1 exists, but it has not proved to be particularly useful. On the other hand, there is a very useful notion of multivariate regularly varying tails.

Definition 4.3.1. A random vector $\mathbf{X} \in \mathbb{R}^d$, or its law, is said to be regularly varying if $P(\|\mathbf{X}\| > x) > 0$ for all $x \geq 0$, and there exist $\alpha \geq 0$ and a probability measure σ on the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ such that for every $b > 0$,

$$\lim_{x\to\infty} \frac{P(\|\mathbf{X}\| > bx,\, \mathbf{X}/\|\mathbf{X}\| \in \cdot)}{P(\|\mathbf{X}\| > x)} = b^{-\alpha}\sigma(\cdot) \tag{4.35}$$

vaguely in $\mathbb{S}^{d-1}$.

Remark 4.3.2. Since the unit sphere $\mathbb{S}^{d-1}$ is compact, there is very little difference between vague and weak convergence in $\mathbb{S}^{d-1}$. The reason we call convergence in (4.35) vague is that the measures involved are not probability measures. See Section 10.2 for a discussion of weak and vague convergence.

The parameter α is called, as in the one-dimensional case, the index of regular variation, the exponent of regular variation, or the tail exponent. The probability measure σ is called the *spectral measure* of $\mathbf{X}$.

Of course, (4.3.1) simply says that for every σ-continuity Borel set $A \in \mathbb{S}^{d-1}$,

$$\lim_{x\to\infty} \frac{P(\|\mathbf{X}\| > bx,\, \mathbf{X}/\|\mathbf{X}\| \in A)}{P(\|\mathbf{X}\| > x)} = b^{-\alpha}\sigma(A)\,. \tag{4.36}$$

In particular, using (4.36) with $A = \mathbb{S}^{d-1}$ shows that the norm of a regularly varying random vector is itself a regularly varying nonnegative random variable. Even more precisely, the statement (4.36) can be restated as follows.

Proposition 4.3.3. *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector. On the event $\mathbf{X} \neq \mathbf{0}$, define $\boldsymbol{\theta}_{\mathbf{X}} = \mathbf{X}/\|\mathbf{X}\|$. Then the vector $\mathbf{X}$ is multivariate regularly varying with tail exponent α and spectral measure σ if and only if its norm $\|\mathbf{X}\|$ is regularly varying with tail exponent α and the conditional law of $\boldsymbol{\theta}_{\mathbf{X}}$ given that $\|\mathbf{X}\| > x$ converges weakly to σ as $x \to \infty$.*

Example 4.3.4. What is multivariate regular variation in $\mathbb{R}^1$? In this case, $\mathbb{S}^0 = \{-1, 1\}$, and for a probability measure σ on $\mathbb{S}^0$, set $p = \sigma(\{1\})$, $q = \sigma(\{-1\})$. Then (4.36) becomes

$$\lim_{x\to\infty} \frac{P(X > x)}{P(|X| > x)} = p, \quad \lim_{x\to\infty} \frac{P(X < -x)}{P(|X| > x)} = q\,,$$

so that multivariate regular variation in $\mathbb{R}^1$ is exactly the balanced regular variation of Definition 4.2.7.

In the case $\alpha > 0$, there is an alternative characterization of multivariate regular variation in which the unit sphere does not play a special role. An advantage of this characterization is that it is possible to use it as a definition of regular variation in metric spaces more general than $\mathbb{R}^d$, in which an obvious unit sphere may not be available.

Consider the space $\overline{\mathbb{R}}_0^d = [-\infty,\infty]^d \setminus \{\mathbf{0}\}$, which we endow with the subspace topology inherited from $[-\infty,\infty]^d$. It is not difficult to check that $\overline{\mathbb{R}}_0^d$ is metrizable as a complete separable metric space; see Exercise 4.6.6. Note that relatively compact sets in $\overline{\mathbb{R}}_0^d$ are precisely the sets that are bounded away from the origin.

Theorem 4.3.5. *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector. The following statements are equivalent.*

(a) $\mathbf{X}$ *is multivariate regularly varying with a tail index $\alpha > 0$.*
(b) *There is a sequence $a_n \uparrow \infty$ such that*

$$nP(a_n^{-1}\mathbf{X} \in \cdot) \xrightarrow{v} \mu(\cdot) \tag{4.37}$$

vaguely in $\overline{\mathbb{R}}_0^d$, where a Radon measure μ on $\overline{\mathbb{R}}_0^d$ is assumed to be nonzero and to satisfy

$$\mu\Big(\Big\{(x_1,\dots,x_d) : x_j \in \{\pm\infty\} \text{ for some } j = 1,\dots,d\Big\}\Big) = 0\,. \tag{4.38}$$

Furthermore, if (4.37) holds, then there is $\alpha > 0$ such that the measure μ has the scaling property

$$\mu\big(aA\big) = a^{-\alpha}\mu(A) \tag{4.39}$$

for every $a > 0$ and every Borel set $A \subset \overline{\mathbb{R}}_0^d$, and the parameter α is the exponent of regular variation of the vector $\mathbf{X}$.

Proof. Suppose first that $\mathbf{X}$ is multivariate regularly varying with tail index $\alpha > 0$. Define a measure μ on $\overline{\mathbb{R}}_0^d$ by assigning it zero mass on infinite points (i.e., guaranteeing (4.38) by definition), and setting

$$\mu(B) = \int_{\mathbb{S}^{d-1}} \int_0^\infty \mathbf{1}(x\mathbf{s} \in B)\, \alpha x^{-(\alpha+1)}\, dx\, \sigma(d\mathbf{s})$$

for Borel sets $B \subset (-\infty,\infty)^d \setminus \{\mathbf{0}\}$. Clearly, μ assigns finite values to sets bounded away from the origin; hence it is a Radon measure on $\overline{\mathbb{R}}_0^d$. Next, we set

$$a_n = \inf\big\{x > 0 : P(\|\mathbf{X}\| > x) \le 1/n\big\},\ n = 1,2,\dots, \tag{4.40}$$

and define

$$m_n(\cdot) = nP(a_n^{-1}\mathbf{X} \in \cdot) \text{ on } \overline{\mathbb{R}}_0^d,\ \ n = 1,2,\dots.$$

Let $B \subset \overline{\mathbb{R}}_0^d$ be a compact set. Then there is $\delta > 0$ such that $\|\mathbf{x}\| \ge \delta$ for all $\mathbf{x} \in B$, so that

$$m_n(B) \le nP(\|\mathbf{X}\| \ge a_n\delta) \sim \frac{P(\|\mathbf{X}\| \ge a_n\delta)}{P(\|\mathbf{X}\| \ge a_n)} \to \delta^{-\alpha}$$

as $n \to \infty$ (see Exercise 4.6.3). Therefore, each measure m_n is Radon, and by Theorem 10.2.8, the family (m_n) is relatively compact in the vague topology on $\overline{\mathbb{R}}_0^d$. In order to prove (4.37), it is enough to prove that all subsequential limits of the sequence (m_n) coincide with μ.

Suppose that $m_{n_k} \xrightarrow{v} \gamma$ for some subsequence $n_k \to \infty$ and a Radon measure γ. Let $C_\gamma = \{\lambda > 0 : \gamma(\{\mathbf{x} : \|\mathbf{x}\| = \lambda\}) = 0\}$. Note that the complement of C_γ in $(0, \infty)$ is at most countable. For every $\lambda \in C_\gamma$,

$$\gamma\big(\{\mathbf{x} : \|\mathbf{x}\| \geq \lambda\}\big) = \lim_{k\to\infty} m_{n_k}\big(\{\mathbf{x} : \|\mathbf{x}\| \geq \lambda\}\big) = \lambda^{-\alpha},$$

as above. Letting $\lambda \to \infty$, we see that γ does not charge infinite points.

Next, let $\lambda \in C_\gamma$, and let A be a Borel subset of $\mathbb{S}^{d-1}$ that is a continuity set both for the spectral measure σ and for the measure

$$\gamma_*(\cdot) = \gamma\Big(\big\{\mathbf{x} : \lambda \leq \|\mathbf{x}\| < \infty,\, \mathbf{x}/\|\mathbf{x}\| \in \cdot\big\}\Big)$$

on $\mathbb{S}^{d-1}$. Then the set

$$B = \Big\{\mathbf{x} \in \overline{\mathbb{R}}_0^d : \lambda < \|\mathbf{x}\| < \infty \text{ and } \mathbf{x}/\|\mathbf{x}\| \in A\Big\}$$

is a continuity set for γ. In particular, $m_{n_k}(B) \to \gamma(B)$ as $k \to \infty$. On the other hand,

$$\begin{aligned} m_n(B) &= n\,P\big(\|\mathbf{X}\| > a_n\lambda,\, \mathbf{X}/\|\mathbf{X}\| \in A\big) \\ &\sim \frac{P\big(\|\mathbf{X}\| > a_n\lambda,\, \mathbf{X}/\|\mathbf{X}\| \in A\big)}{P(\|\mathbf{X}\| > a_n)} \to \sigma(A)\lambda^{-\alpha} = \mu(B) \end{aligned}$$

as $n \to \infty$ by the multivariate regular variation of $\mathbf{X}$. Therefore, $\mu(B) = \gamma(B)$. Since sets B as above form a π-system generating the Borel σ-field on $\mathbb{R}^d \setminus \{\mathbf{0}\}$, the measures μ and γ coincide on that Borel σ-field. Since neither μ nor γ charges the set of infinite points, we conclude that $\gamma = \mu$, and so all subsequential limits of the sequence (m_n) indeed coincide with μ.

In the opposite direction, suppose that (4.37) holds and μ satisfies (4.38). For $x > 0$, define $h(x) = P(\|\mathbf{X}\| > x)$, $g(x) = \mu\big(\{\mathbf{x} : \|\mathbf{x}\| > x\}\big)$. The functions h and g satisfy the assumptions of Lemma 10.5.15, so we conclude that the sequence (a_n) is regularly varying with some exponent $\beta > 0$. Let $a > 0$ and take any $b_j \geq 0$, $j = 1, \ldots, d$, not all equal to zero, such that

$$\mu\{\mathbf{x} : x_j = b_j \text{ or } x_j = ab_j \text{ for some } j = 1, \ldots, d\} = 0\,.$$

Let $A = \prod_{j=1}^{d} [b_j, \infty)$. Then both A and aA are relatively compact μ-continuity sets in $\overline{\mathbb{R}}_0^d$, so by (4.37) and the regular variation of the sequence (a_n),

$$\mu(aA) = \lim_{n\to\infty} nP(a_n^{-1}\mathbf{X} \in aA)$$
$$= \lim_{n\to\infty} nP\big(a_{\lceil na^{-1/\beta}\rceil}^{-1}\mathbf{X} \in A\big) = a^{1/\beta}\mu(A)\,.$$

Setting $\alpha = 1/\beta$, and noticing that the sets A as above form a π-system generating the Borel σ-field on $[0,\infty)^d \setminus \{\mathbf{0}\}$, we conclude that the measures $\mu(a\cdot)$ and $a^\alpha\mu(\cdot)$ coincide on $[0,\infty)^d \setminus \{\mathbf{0}\}$. Reversing the directions of some of the half-lines in the definition of the set A above and using the same argument shows that these two measures coincide on all the other quadrants in $\mathbb{R}^d \setminus \{\mathbf{0}\}$ as well. This proves the scaling property (4.39). One consequence of (4.39) is that the measure μ must assign zero weight to all spheres $b\mathbb{S}^{d-1}$, $b > 0$, for if it assigns a positive weight to any one of these spheres, by the scaling property it assigns a positive weight to all of them, and a σ-finite measure cannot assign a positive weight to each of uncountably many disjoint sets.

Because of (4.39), the measure

$$\sigma(A) = \frac{\mu\big\{\mathbf{x}:\ \|\mathbf{x}\| > b,\ \mathbf{x}/\|\mathbf{x}\| \in A\big\}}{\mu\{\mathbf{x}:\ \|\mathbf{x}\| > b\}},\quad A \text{ a Borel subset of } \mathbb{S}^{d-1},$$

is a probability measure on $\mathbb{S}^{d-1}$, independent of the choice of $b > 0$. If A is a σ-continuity Borel subset of $\mathbb{S}^{d-1}$, then for every $b > 0$, the set

$$B = \big\{\mathbf{x}:\ \|\mathbf{x}\| > b,\ \mathbf{x}/\|\mathbf{x}\| \in A\big\}$$

is a μ-continuity relatively compact Borel subset of $\overline{\mathbb{R}}_0^d$. If we set, for $x > 0$, $n_x = \inf\{n \geq 1:\ a_n \geq x\}$, then $n_x \to \infty$ as $x \to \infty$. For every $b > 0$, we have, by (4.37), with sets A and B as above,

$$n_x P\big(\|\mathbf{X}\| > bx,\ \mathbf{X}/\|\mathbf{X}\| \in A\big) \sim n_x P\big(\|\mathbf{X}\| > ba_{n_x},\ \mathbf{X}/\|\mathbf{X}\| \in A\big)$$
$$= n_x P(\mathbf{X} \in a_{n_x}B) \to \mu(B) = \sigma(A)\mu\big(b\mathbb{S}^{d-1}\big)$$

as $x \to \infty$. Since

$$n_x P(\|\mathbf{X}\| > x) \sim n_x P\big(\|\mathbf{X}\| > a_{n_x}\big) \to \mu\big(\mathbb{S}^{d-1}\big)\,,$$

we conclude by (4.39) that

$$\lim_{x\to\infty} \frac{P(\|\mathbf{X}\| > bx,\ \mathbf{X}/\|\mathbf{X}\| \in A)}{P(\|\mathbf{X}\| > x)} = \sigma(A)\frac{\mu\big(b\mathbb{S}^{d-1}\big)}{\mu\big(\mathbb{S}^{d-1}\big)} = b^{-\alpha}\sigma(A)\,,$$

and (4.36) follows. Therefore, $\mathbf{X}$ is multivariate regularly varying with a tail index $\alpha > 0$. □

The limiting measure μ in (4.37) is called the *tail measure* of the random vector $\mathbf{X}$. It is clear that the tail measure is defined uniquely up to a positive multiplicative constant.

Corollary 4.2.9 has the following, very natural, extension to the case of multivariate regular variation.

Proposition 4.3.6. *Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector. Assume that $\mathbf{X}$ is multivariate regularly varying with a tail index $\alpha > 0$ and tail measure μ_X, corresponding to some sequence (a_n). Let $\mathbf{A}$ be an $m \times d$ random matrix, independent of $\mathbf{X}$, such that for some $\varepsilon > 0$, $E\|\mathbf{A}\|^{\alpha+\varepsilon} < \infty$. Assume that the measure $E\big(\mu_X \circ \mathbf{A}^{-1}\big)$ on $\mathbb{R}^m$ is not the zero measure. Then the random vector $\mathbf{Z} = \mathbf{A}\mathbf{X} \in \mathbb{R}^m$ is multivariate regularly varying with a tail index $\alpha > 0$. If one uses the same sequence (a_n), then the tail measure of $\mathbf{Z}$ is given by $\mu_Z = E\big(\mu_X \circ \mathbf{A}^{-1}\big)$.*

Proof. Let B be a Borel subset of $\mathbb{R}^m$, bounded away from the origin. Assume that B is a continuity set for the measure $E\big(\mu_X \circ \mathbf{A}^{-1}\big)$. Then the set $\mathbf{A}^{-1}(B)$ is a Borel subset of $\mathbb{R}^d$ bounded away from the origin. It is, further, with probability 1, a continuity set for the measure μ_X. Notice that

$$nP(a_n\mathbf{Z} \in B) = \int_{\mathbb{R}^{m\times d}} nP\big(\mathbf{X} \in \mathbf{a}^{-1}(B)\big)\, F_A(d\mathbf{a})\,.$$

Here F_A is the law of the random matrix on the space of $\mathbb{R}^{m\times d}$ of $m \times d$ matrices. For F_A-almost every $\mathbf{a}$,

$$nP\big(\mathbf{X} \in \mathbf{a}^{-1}(B)\big) \to \mu_X\big(\mathbf{a}^{-1}(B)\big)\,.$$

Let $\delta := \inf\{\|\mathbf{x}\| : \mathbf{x} \in B\} > 0$. For $\mathbf{a} \in \mathbb{R}^{m\times d}$, write

$$\begin{aligned} nP\big(\mathbf{X} \in \mathbf{a}^{-1}(B)\big) &\le nP\big(a_n^{-1}\|\mathbf{X}\| > \delta\|\mathbf{a}\|^{-1}\big) \\ &= nP\big(a_n^{-1}\|\mathbf{X}\| > \delta\big)\frac{P\big(a_n^{-1}\|\mathbf{X}\| > \delta\|\mathbf{a}\|^{-1}\big)}{P\big(a_n^{-1}\|\mathbf{X}\| > \delta\big)}\,. \end{aligned}$$

The argument, using the Potter bounds of Corollary 10.5.8 in the proof of Proposition 4.2.6, shows that there is a constant $C > 0$ such that

$$\frac{P\big(a_n^{-1}\|\mathbf{X}\| > \delta\|\mathbf{a}\|^{-1}\big)}{P\big(a_n^{-1}\|\mathbf{X}\| > \delta\big)} \le C\|\mathbf{a}\|^{\alpha+\varepsilon}\,.$$

Using the moment assumption on the norm of the random matrix $\mathbf{A}$, we apply the dominated convergence theorem to obtain

$$\begin{aligned} \lim_{n\to\infty} nP(a_n\mathbf{Z} \in B) &= \int_{\mathbb{R}^{m\times d}} \lim_{n\to\infty} nP\big(\mathbf{X} \in \mathbf{a}^{-1}(B)\big)\, F_A(d\mathbf{a}) \\ &= \int_{\mathbb{R}^{m\times d}} \mu_X\big(\mathbf{a}^{-1}(B)\big)\, F_A(d\mathbf{a}) = E\big(\mu_X \circ \mathbf{A}^{-1}(B)\big)\,, \end{aligned}$$

as required. □

Let $\mathbf{X} \in \mathbb{R}^d$ be a multivariate regularly varying random vector with a tail index $\alpha > 0$. Let μ_X be the tail measure of $\mathbf{X}$, corresponding to some sequence (a_n). Let $(b_1, \ldots, b_d) \in \mathbb{R}^d$. Then

$$\frac{P\big(\sum_{j=1}^d b_j X_j > x\big)}{P(\|\mathbf{X}\| > x)} \to \frac{\mu_X\big(\{\mathbf{x} : \sum_{j=1}^d b_j x_j > 1\}\big)}{\mu_X\big(\{\mathbf{x} : \|\mathbf{x}\| > 1\}\big)}$$

as $x \to \infty$. This means that every linear combination of the components of $\mathbf{X}$ is a one-dimensional random variable whose right tail is regularly varying with the same tail index α. Even more precisely, the right tails of the distributions of all the linear combinations are equivalent, and all are equivalent to the tail of the distribution of the norm of the vector $\mathbf{X}$. Strictly speaking, this is true only for those linear combinations for which

$$\mu_X\big(\{\mathbf{x} : \sum_{j=1}^d b_j x_j > 0\}\big) > 0\,. \tag{4.41}$$

If for some coefficients $(b_1, \ldots, b_d)$, the condition (4.41) fails, then the right tail of the distribution of the linear combination $\sum_{j=1}^d b_j X_j$ is still proportional to the right tail of the distribution of the norm $\|\mathbf{X}\|$, but the coefficient of proportionality is equal to zero. The fact that the tail measure μ_X is not the zero measure implies that the condition (4.41) holds at least for some vectors of coefficients.

This discussion is often summarized by saying that all linear combinations of the components of a multivariate regularly varying random vector are themselves regularly varying in one dimension.

The question to what extent a properly formulated assumption of regular variation of all linear combinations of the components of a random vector implies multivariate regular variation of the latter is much harder, and we address it next. Let $\mathbf{X} \in \mathbb{R}^d$ be a random vector, and let D be a regularly varying positive random variable with a tail index $\alpha > 0$. The assumption

$$\begin{aligned} &w(\mathbf{b}) = \lim_{x\to\infty} \frac{P\big(\sum_{j=1}^d b_j X_j > x\big)}{P(D > x)} \text{ exists and is finite for any } \mathbf{b} = (b_1, \ldots, b_d) \in \mathbb{R}^d \\ &w(\mathbf{b}^{(0)}) > 0 \text{ for at least one } \mathbf{b}^{(0)} \in \mathbb{R}^d \end{aligned} \tag{4.42}$$

is the form in which we think of regular variation of the linear combinations of the components of a random vector. We have seen that if a random vector $\mathbf{X}$ is multivariate regularly varying, then (4.42) holds with $D = \|\mathbf{X}\|$. On the other hand, suppose that (4.42) holds. Define for $t > 0$ a finite measure on $\overline{\mathbb{R}}_0^d$ by

$$m_t(\cdot) = \frac{P(\mathbf{X} \in t\cdot)}{P\big(\sum_{j=1}^d b_j^{(0)} X_j > t\big)}\,, \tag{4.43}$$

where $\mathbf{b}^{(0)} = (b_1^{(0)}, \ldots, b_d^{(0)})$ is as in (4.42). Note that for each $\delta > 0$,

$$m_t\big(\{\mathbf{x} : \|\mathbf{x}\| > \delta\}\big) \le \sum_{j=1}^{d} \frac{P(|X_j| > t\delta/d)}{P\big(\sum_{j=1}^{d} b_j^{(0)} X_j > t\big)}$$
$$\to \sum_{j=1}^{d} \frac{w\big(d\mathbf{e}^{(j)}/\delta\big) + w\big(-d\mathbf{e}^{(j)}/\delta\big)}{w(\mathbf{b}^{(0)})} < \infty$$

as $t \to \infty$. Here $\mathbf{e}^{(j)}$ is the vector in $\mathbb{R}^d$ with all zero coordinates except the jth coordinate, which is equal to one. By Theorem 10.2.8, the family (m_t) is relatively compact in the vague topology on $\overline{\mathbb{R}}_0^d$. Let γ be such that $m_{t_k} \xrightarrow{v} \gamma$ as $k \to \infty$ for some sequence $t_k \to \infty$. Then by the continuous mapping theorem for vague convergence (Theorem 10.2.9) and (4.42), for every $\mathbf{b} \in \mathbb{R}^d$ and $x > 0$,

$$\gamma\big(\{\mathbf{y} : \sum_{j=1}^{d} b_j y_j > x\}\big) = \lim_{k\to\infty} \frac{P\big(\sum_{j=1}^{d} b_j X_j > t_k x\big)}{P\big(\sum_{j=1}^{d} b_j^{(0)} X_j > t_k\big)} = x^{-\alpha} \frac{w(\mathbf{b})}{w(\mathbf{b}^{(0)})}\,; \tag{4.44}$$

note that the regular variation of D in (4.42) justifies the first equality above, because the set on which the measure γ is evaluated must be a γ-continuity set.

Note that the expression on the left-hand side of (4.44) is determined by the ratios $w(\mathbf{b})/w(\mathbf{b}^{(0)})$, $\mathbf{b} \in \mathbb{R}^d$. Therefore, all possible limiting measures γ as above coincide on the half-spaces of the type $\{\mathbf{y} : \sum_{j=1}^{d} b_j y_j > x\}$. If the values a Radon measure on such half-spaces were to determine the entire measure γ, we would immediately conclude that a vague limit of μ_t as $t \to \infty$ exists. Defining

$$a_n = \inf\big\{t > 0 : P\big(\sum_{j=1}^{d} b_j^{(0)} X_j > t\big) \le 1/n\big\},\ n = 1, 2, \ldots,$$

we see that (4.37) holds. Since by (4.44), the limiting measure γ does not charge the infinite points, we would conclude by Theorem 4.3.5 that $\mathbf{X}$ is multivariate regularly varying. Moreover, the spectral measure σ of $\mathbf{X}$ would be uniquely determined by the ratios $w(\mathbf{b})/w(\mathbf{b}^{(0)})$, $\mathbf{b} \in \mathbb{R}^d$.

Everything hangs, therefore, on the question whether certain Radon measures are uniquely determined by the values they assign to half-spaces of the type

$$\big\{\mathbf{y} : \sum_{j=1}^{d} b_j y_j > x\big\},\ \mathbf{b} \in \mathbb{R}^d,\ x > 0\,. \tag{4.45}$$

A simple characteristic functions argument shows that *probability measures* are indeed uniquely determined by their values on such half-spaces. Since we are

dealing with infinite measures, the situation is more complicated, and we have to use the natural scaling property of the possible limiting measures.

Theorem 4.3.7 below shows that in most cases, (4.42) implies multivariate regular variation of $\mathbf{X}$, because the possible limiting measures are, in fact, uniquely determined by their values on the half-spaces.

Theorem 4.3.7. *Suppose that a random vector* $\mathbf{X}$ *satisfies* (4.42). *If* $\alpha > 0$ *is a noninteger, then* $\mathbf{X}$ *is multivariate regularly varying, and the spectral measure* σ *of* $\mathbf{X}$ *is uniquely determined by the ratios* $w(\mathbf{b})/w(\mathbf{b}^{(0)})$, $\mathbf{b} \in \mathbb{R}^d$, *where* $\mathbf{b}^{(0)}$ *is such that* $w(\mathbf{b}^{(0)}) > 0$. *The same is true if* α *is an odd integer and* $\mathbf{X}$ *is either symmetric or takes values in* $[0, \infty)^d$.

Proof. We need to prove that under the assumptions of the theorem, all possible subsequential limits of the family (m_t) in (4.43) have the following property: if two such measures coincide on subspaces of type (4.45), then the two measures are in fact equal. The characterizing feature of the possible limiting measures γ turns out to be its scaling (in $x > 0$) property (4.44). While this property does not immediately imply that the measure γ enjoys a full scaling property as in (4.39), it carries enough information to establish finiteness of certain integrals with respect to γ. Specifically, we claim that

$$\begin{array}{l} \int_{\|\mathbf{y}\| \le 1} \|\mathbf{y}\|^p \, \gamma(d\mathbf{y}) < \infty \text{ for every } p > \alpha \\ \int_{\|\mathbf{y}\| \ge 1} \|\mathbf{y}\|^p \, \gamma(d\mathbf{y}) < \infty \text{ for every } p < \alpha \end{array} . \tag{4.46}$$

Indeed, let $p > \alpha$. By (4.44), for some $c_p \in (0, \infty)$,

$$\begin{aligned} \int_{\|\mathbf{y}\| \le 1} \|\mathbf{y}\|^p \, \gamma(d\mathbf{y}) &= \int_0^\infty \gamma\big(\{\mathbf{y} : \mathbf{1}(\|\mathbf{y}\| \le 1)\|\mathbf{y}\|^p > t\}\big)\, dt \\ \le \int_0^1 \gamma\big(\{\mathbf{y} : \|\mathbf{y}\|^p > t\}\big)\, dt &\le c_p \sum_{j=1}^d \int_0^1 \gamma\big(\{\mathbf{y} : |y_j| > t^{1/p}/d\}\big)\, dt \\ &= c_p \sum_{j=1}^d \int_0^1 \frac{w(\mathbf{e}_j) + w(-\mathbf{e}_j)}{w(\mathbf{b}^{(0)})} \big(t^{1/p}/d\big)^{-\alpha}\, dt < \infty . \end{aligned}$$

Similarly, for $0 < p < \alpha$,

$$\begin{aligned} \int_{\|\mathbf{y}\| \ge 1} \|\mathbf{y}\|^p \, \gamma(d\mathbf{y}) &= \int_0^\infty \gamma\big(\{\mathbf{y} : \mathbf{1}(\|\mathbf{y}\| \ge 1)\|\mathbf{y}\|^p > t\}\big)\, dt \\ &\le \gamma\big(\{\mathbf{y} : \mathbf{1}(\|\mathbf{y}\| \ge 1)\}\big) + \int_1^\infty \gamma\big(\{\mathbf{y} : \|\mathbf{y}\|^p > t\}\big)\, dt \\ \le \gamma\big(\{\mathbf{y} : \mathbf{1}(\|\mathbf{y}\| \ge 1)\}\big) &+ c_p \sum_{j=1}^d \int_1^\infty \frac{w(\mathbf{e}_j) + w(-\mathbf{e}_j)}{w(\mathbf{b}^{(0)})} \big(t^{1/p}/d\big)^{-\alpha}\, dt < \infty . \end{aligned}$$

The integrability properties (4.46) will allow us to use the characteristic function approach as if we were dealing with probability measures. Suppose first that α is a noninteger. Let γ_j, $j = 1, 2$, be two measures satisfying (4.46) that coincide on subspaces of type (4.45). We will prove that $\gamma_1 = \gamma_2$. For an integer $n \geq 1$ such that $\alpha \in (2n-2, 2n)$, we define for $j = 1, 2$,

$$\nu_j(B) = (-1)^n \int_B \left(e^{i(\mathbf{b}^{(0)},\mathbf{y})} - e^{-i(\mathbf{b}^{(0)},\mathbf{y})}\right)^{2n} \gamma_j(d\mathbf{y}), \quad B \text{ Borel}. \tag{4.47}$$

Since the integrand in this integral is bounded, and additionally, its absolute value is bounded by a constant factor of $\|\mathbf{y}\|^{2n}$ near the origin, it follows from (4.46) that the integral is a finite nonnegative real number. Therefore, ν_1 and ν_2 are two finite measures on $\mathbb{R}^d$. We will prove that their characteristic functions coincide. This will imply that the total masses of ν_1 and ν_2 are equal and, after a normalization, that $\nu_1 = \nu_2$. Since $d\nu_1/d\gamma_1 = d\nu_2/d\gamma_2$, the latter fact will imply that $\gamma_1 = \gamma_2$ as long as the common Radon–Nykodim derivative does not vanish outside a set of measure zero with respect to both γ_1 and γ_2. However, the derivative vanishes only on the countable collection of hyperplanes $\{\mathbf{y} : (\mathbf{b}^{(0)}, \mathbf{y}) = \pi + \pi m\}$, $m = 0, \pm 1, \ldots$, and by (4.44), each of these hyperplanes has measure zero with respect to both γ_1 and γ_2.

It remains, therefore, to prove that the characteristic functions of ν_1 and ν_2 coincide. For $\mathbf{a} \in \mathbb{R}^d$, we have by the binomial formula

$$(-1)^n \int_{\mathbb{R}^d} e^{i(\mathbf{a},\mathbf{y})}\, \nu_j(d\mathbf{y}) = \int_{\mathbb{R}^d} \sum_{k=0}^{2n} (-1)^k \binom{2n}{k} \exp\{i(\mathbf{a} + 2(n-k)\mathbf{b}^{(0)}, \mathbf{y})\}\, \gamma_j(d\mathbf{y}). \tag{4.48}$$

Recall that since γ_1 and γ_2 coincide on subspaces of type (4.45), we automatically have

$$\int_{\mathbb{R}^d} \varphi\big((\mathbf{b}, \mathbf{y})\big)\, \gamma_1(d\mathbf{y}) = \int_{\mathbb{R}^d} \varphi\big((\mathbf{b}, \mathbf{y})\big)\, \gamma_2(d\mathbf{y}) \tag{4.49}$$

for $\mathbf{b} \in \mathbb{R}^d$ and every measurable function φ for which the two integrals are well defined. Therefore, if we could interchange the sum and the integral in (4.48), we could apply (4.49) to each of the resulting integrals and conclude that the expression in (4.48) is the same for $j = 1$ and $j = 2$. However, even though the sum on the right-hand side of (4.48) is integrable, each individual term in the sum is easily seen not to be integrable. To overcome this problem, we modify the individual terms without changing the sum, as follows. Let $j < m$ be two nonnegative integers. By the binomial formula,

$$\sum_{k=0}^{m} (-1)^k \binom{m}{k} k(k-1)\ldots(k-j+1)$$

$$= (-1)^j\, m(m-1)\ldots(m-j+1) \sum_{k=0}^{m-j} (-1)^k \binom{m-j}{k} = 0,$$

implying that

$$\sum_{k=0}^{m}(-1)^k\binom{m}{k}k^j = 0$$

as well. In particular, for every pair of complex numbers h_1, h_2, and $0 \le j < m$,

$$\sum_{k=0}^{m}(-1)^k\binom{m}{k}e^{kh_1+h_2} = \sum_{k=0}^{m}(-1)^k\binom{m}{k}\left(e^{kh_1+h_2} - \sum_{l=0}^{j}\frac{(kh_1+h_2)^l}{l!}\right). \tag{4.50}$$

Suppose that $\alpha \in (2n-1, 2n)$. We use (4.50) with $m = 2n$, $j = 2n-1$, $h_1 = -2i(\mathbf{b}^{(0)}, \mathbf{y})$, and $h_2 = i(\mathbf{a} + 2n\mathbf{b}^{(0)}, \mathbf{y})$ to rewrite the right-hand side of (4.48) as

$$\int_{\mathbb{R}^d}\sum_{k=0}^{2n}(-1)^k\binom{2n}{k}R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)}, \mathbf{y})\big)\,\gamma_j(d\mathbf{y}) \tag{4.51}$$

$$= \sum_{k=0}^{2n}(-1)^k\binom{2n}{k}\int_{\mathbb{R}^d}R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)}, \mathbf{y})\big)\,\gamma_j(d\mathbf{y}).$$

Here

$$R_j(z) = e^z - \sum_{l=0}^{j}\frac{z^l}{l!},$$

and interchanging the sum and the integral is now permitted, since for each $k = 0, 1, \ldots, 2n$, for some finite constant c,

$$\left|R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)}, \mathbf{y})\big)\right| \le \begin{cases} c\|\mathbf{y}\|^{2n} & \text{if } \|\mathbf{y}\| \le 1, \\ c\|\mathbf{y}\|^{2n-1} & \text{if } \|\mathbf{y}\| \ge 1. \end{cases}$$

By (4.46), all the integrals on the right-hand side of (4.51) are finite. Finally, appealing to (4.49), we conclude that the right-hand side of (4.51) is the same for $j = 1$ and $j = 2$, and hence the same is true for (4.48). Thus, the characteristic functions of ν_1 and ν_2 coincide.

If $\alpha \in (2n-2, 2n-1)$, we can use the same argument as above, but with $j = 2n-2$. Therefore, we have proved the theorem in the case of a noninteger α.

We consider now the case that $\alpha = 2n-1$ is an odd integer. Suppose first that $\mathbf{X}$ is a symmetric random vector. In that case, all the measures m_t in (4.43) are symmetric, and hence so are all possible vague limits of these measures. Therefore, when proving that $\gamma_1 = \gamma_2$ above, we need to consider only symmetric measures

γ_1 and γ_2. By this symmetry, we can rewrite the integral on the left-hand side of (4.51) as

$$\begin{aligned}
&\frac{1}{2}\int_{\mathbb{R}^d}\sum_{k=0}^{2n}(-1)^k\binom{2n}{k}\Big[R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)\\
&\qquad +R_{2n-1}\big(-i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)\Big]\,\gamma_j(d\mathbf{y})\\
&=\frac{1}{2}\sum_{k=0}^{2n}(-1)^k\binom{2n}{k}\int_{\mathbb{R}^d}\Big[R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)\\
&\qquad +R_{2n-1}\big(-i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)\Big]\,\gamma_j(d\mathbf{y})\,.
\end{aligned}$$

Note that interchanging the sum and the integral is again permitted, because for each $k=0,1,\ldots,2n$, for some finite constant c,

$$\begin{aligned}
&\Big|R_{2n-1}\big(i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)+R_{2n-1}\big(-i(\mathbf{a}+2(n-k)\mathbf{b}^{(0)},\mathbf{y})\big)\Big|\\
&\qquad\le\begin{cases}c\|\mathbf{y}\|^{2n} & \text{if } \|\mathbf{y}\|\le 1\\ c\|\mathbf{y}\|^{2n-2} & \text{if } \|\mathbf{y}\|\ge 1\end{cases},
\end{aligned}$$

which by (4.46) provides the needed integrability when $\alpha=2n-1$. We complete the argument by appealing, once again, to (4.49).

Finally, we consider the case of $\alpha=2n-1$ and $\mathbf{X}$ taking values in $[0,\infty)^d$. Let $\mathbf{Y}$ be a random vector whose law is the equal-weight mixture of the laws of $\mathbf{X}$ and $-\mathbf{X}$. Clearly, $\mathbf{Y}$ is a symmetric random vector. By the assumption (4.42), for every $\mathbf{b}\in\mathbb{R}^d$,

$$\begin{aligned}
&\lim_{x\to\infty}\frac{P\big(\sum_{j=1}^d b_jY_j>x\big)}{P(D>x)}\\
&=\frac{1}{2}\left(\lim_{x\to\infty}\frac{P\big(\sum_{j=1}^d b_jX_j>x\big)}{P(D>x)}+\lim_{x\to\infty}\frac{P\big(\sum_{j=1}^d(-b_j)X_j>x\big)}{P(D>x)}\right)=\frac{1}{2}\big(w(\mathbf{b})+w(-\mathbf{b})\big)\,.
\end{aligned}$$

Therefore, the vector $\mathbf{Y}$ itself satisfies (4.42). Define a family of measures as in (4.43) but corresponding to $\mathbf{Y}$:

$$m_t^{(Y)}(\cdot)=\frac{P(\mathbf{Y}\in t\cdot)}{P\big(\sum_{j=1}^d b_j^{(0)}Y_j>t\big)},\ t>0\,.$$

These measures are related to the measures defined by (4.43) (for $\mathbf{X}$) by

$$m_t^{(Y)}(\cdot)=\frac{P\big(\sum_{j=1}^d b_j^{(0)}X_j>t\big)}{P\big(\sum_{j=1}^d b_j^{(0)}Y_j>t\big)}\frac{1}{2}\big(m_t(\cdot)+m_t(-\cdot)\big)\,.$$

If $m_{t_k} \xrightarrow{v} \gamma$ as $k \to \infty$ for some sequence $t_k \to \infty$, then

$$m_{t_k}^{(Y)}(\cdot) \xrightarrow{v} \frac{w(\mathbf{b}^{(0)})}{w(\mathbf{b}^{(0)}) + w(-\mathbf{b}^{(0)})}(\gamma(\cdot) + \gamma(-\cdot)) .$$

However, by the already considered symmetric case, the vague limit of $m_{t_k}^{(Y)}$ does not depend on the sequence $t_k \to \infty$. Therefore, the sum $\gamma(\cdot) + \gamma(-\cdot)$ is uniquely determined. Since γ is concentrated on $[0, \infty)^d \setminus \{\mathbf{0}\}$, γ itself is uniquely determined, and the proof of the theorem is complete. □

Remark 4.3.8. The statement of Theorem 4.3.7 is not true without restrictions on the value of $\alpha > 0$ and/or the support of the vector $\mathbf{X}$. In Hult and Lindskog (2006), examples are provided of random vectors $\mathbf{X}$ satisfying (4.42) with an integer α but not possessing multivariate regular variation. The authors also present an example of a random vector taking values in $[0, \infty)^d$ that does not have the property of multivariate regular variation for which all linear combinations of coordinates with nonnegative weights are regularly varying, with any given integer exponent of regular variation. This cannot happen if the exponent of regular variation of the linear combinations is a noninteger; see Exercise 4.6.7.

It is not known whether the statement of Theorem 4.3.7 remains true if α is an even integer and $\mathbf{X}$ takes values in $[0, \infty)^d$. However, a corresponding statement when $\mathbf{X}$ is symmetric is false. To see this, let $\alpha = 2k$, and let $Y_1, \ldots, Y_d$ be a random vector taking values in $[0, \infty)^d$ without the property of multivariate regular variation, for which all linear combinations of coordinates with nonnegative weights are regularly varying with exponent $\alpha = k$, in the sense of (4.42). Let $G_1, \ldots, G_d$ be i.i.d. standard normal random variables independent of $Y_1, \ldots, Y_d$, and define $X_i = G_i Y_2^{1/2}$, $i = 1, \ldots, d$. Clearly, $\mathbf{X} = (X_1, \ldots, X_d)$ is a symmetric random vector. For each $\mathbf{b} \in \mathbb{R}^d$, as $x \to \infty$,

$$P\big((\mathbf{b}, \mathbf{X}) > x\big) = P\left(\sum_{j=1}^{d} b_j G_j Y_j^{1/2} > x\right)$$

$$= P\left(G_1 \sum_{j=1}^{d} b_j^2 Y_j > x\right) \sim E\big((G_1)_+\big)^k P\left(\sum_{j=1}^{d} b_j^2 Y_j > x\right)$$

by Proposition 4.2.6. Therefore, the random vector $\mathbf{X}$ satisfies (4.42). However, it is not multivariate regularly varying. Indeed, if it were multivariate regular varying, so would be its coordinatewise square $(G_1^2 Y_1, \ldots, G_d^2 Y_d)$. The "cancellation property" of the powers of a normal random variable (see Theorem 4.1 in Damek et al. (2014) and Example 4.4 in Jacobsen et al. (2008)) tells us that in this case, the vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ would itself be multivariate regular varying, which, by the assumption, it is not.

4.4 Heavy Tails and Convergence of Random Measures

Various limit theorems concerning sequences of regularly varying random variables and vectors often become clearer when viewed through the lenses of point processes and, more generally, random measures associated with these sequences. Let $(\mathbf{X}_1, \mathbf{X}_2 \ldots)$ be a sequence of i.i.d. regularly varying random vectors in $\mathbb{R}^d$, with a tail index $\alpha > 0$. With the sequence (a_n) defined by (4.40), we define the point process associated with the first n observations in the sequence by

$$N_n = \sum_{j=1}^{n} \delta_{(j/n,\, \mathbf{X}_j/a_n)} \,; \tag{4.52}$$

recall that δ_b is the (Dirac) point mass at b. The next theorem shows that this sequence of point processes converges to a Poisson process, weakly in the space $M_+^{\mathrm{R}}\big([0,1]\times\overline{\mathbb{R}}_0^d\big)$ of Radon measures on $[0,1]\times\overline{\mathbb{R}}_0^d$, endowed with the vague topology; see Section 10.2. Here $\overline{\mathbb{R}}_0^d = [-\infty,\infty]^d \setminus \{\mathbf{0}\}$, and by Exercise 4.6.6, it is metrizable as a complete separable metric space. Therefore, so is the space $[0,1]\times\overline{\mathbb{R}}_0^d$, endowed with the product topology.

Theorem 4.4.1. *Let $(\mathbf{X}_1, \mathbf{X}_2 \ldots)$ be a sequence of i.i.d. regularly varying random vectors in $\mathbb{R}^d$, with a tail index $\alpha > 0$ and tail measure μ. Then*

$$N_n \Rightarrow N_* \ \text{ as } n \to \infty \tag{4.53}$$

weakly in $M_+^{\mathrm{R}}\big([0,1] \times \overline{\mathbb{R}}_0^d\big)$, where N_* is a Poisson random measure on $[0,1] \times \mathbb{R}^d$ with mean measure $m = \lambda \times \mu$, naturally considered as a Poisson random measure on $[0,1] \times \overline{\mathbb{R}}_0^d$.

Proof. By Theorem 10.2.13, it is enough to prove that the Laplace functional of N_n converges to the Laplace functional of N_* for every nonnegative continuous function f on $[0,1] \times \overline{\mathbb{R}}_0^d$ with compact support. As the expression for the Laplace functional of a Poisson random measure in Exercise 10.9.4 shows, this convergence, in turn, will follow from the following statement:

$$\lim_{n\to\infty} E \exp\left\{-\int_{[0,1]\times\mathbb{R}^d} f(t,\mathbf{x})\, N_n(dt,\, d\mathbf{x})\right\} \tag{4.54}$$

$$= \exp\left\{-\int_{[0,1]}\int_{\mathbb{R}^d} \left(1 - e^{-f(t,\mathbf{x})}\right)\, \mu(d\mathbf{x})\, dt\right\}$$

for every bounded and uniformly continuous function f on $[0,1] \times \overline{\mathbb{R}}_0^d$ such that for some $\varepsilon = \varepsilon(f) > 0$,

$$f(t,\mathbf{x}) = 0 \ \text{ for all } (t,\mathbf{x}) \text{ such that } \|\mathbf{x}\| \le \varepsilon. \tag{4.55}$$

We will begin by checking that for every $t \in [0,1]$,

$$\lim_{n\to\infty} \left(E e^{-f\left(t, \mathbf{X}_1/a_n\right)} \right)^n = \exp\left\{ -\int_{\mathbb{R}^d} \left(1 - e^{-f(t,\mathbf{x})}\right) \mu(d\mathbf{x}) \right\} . \tag{4.56}$$

To see this, reuse the notation $m_n(\cdot) = nP(a_n^{-1}\mathbf{X} \in \cdot)$ as a Radon measure on $\overline{\mathbb{R}}_0^d$. Theorem 4.3.5 says that $m_n \xrightarrow{v} \mu$. Since the function $1 - e^{-f(t,\cdot)}$ is continuous with compact support, we conclude that

$$\begin{aligned} \int_{\mathbb{R}^d} \left(1 - e^{-f(t,\mathbf{x})}\right) \mu(d\mathbf{x}) &= \lim_{n\to\infty} \int_{\mathbb{R}^d} \left(1 - e^{-f(t,\mathbf{x})}\right) m_n(d\mathbf{x}) \\ &= \lim_{n\to\infty} n\, E\left(1 - e^{-f\left(t, \mathbf{X}_1/a_n\right)}\right) . \end{aligned}$$

This implies (4.56). We proceed with discretizing the function f. For $m = 1, 2, \ldots$, let $f_m(0, \mathbf{x}) = f(0, \mathbf{x})$ and

$$f_m(t, \mathbf{x}) = f\big(i/m, \mathbf{x}\big) \text{ if } (i-1)/m < t \le i/m,\ i = 1, \ldots, m.$$

It is clear that for every $m = 1, 2, \ldots$,

$$\sup_{(t,\mathbf{x})} \big| f_m(t, \mathbf{x}) \big| \le \sup_{(t,\mathbf{x})} \big| f(t, \mathbf{x}) \big|,\ f_m(t, \mathbf{x}) = 0 \text{ for all } (t, \mathbf{x}) \text{ such that } \|\mathbf{x}\| \le \varepsilon.$$

Further, by the uniform continuity of f,

$$\lim_{m\to\infty} \sup_{(t,\mathbf{x})} \big| f(t, \mathbf{x}) - f_m(t, \mathbf{x}) \big| = 0 . \tag{4.57}$$

Fix $m = 1, 2, \ldots$. Using the notation $c_{i,m}(n),\ i = 1, \ldots, m$, for positive numbers satisfying $c_{i,m}(n) \to 1$ as $n \to \infty$, $i = 1, \ldots, m$, we obtain by (4.56),

$$\begin{aligned} E \exp\left\{ -\int_{[0,1]\times\mathbb{R}^d} f_m(t, \mathbf{x})\, N_n(dt,\ d\mathbf{x}) \right\} &= \prod_{i=1}^m \left(E e^{-f\left(i/m,\, \mathbf{X}_1/a_n\right)} \right)^{c_{i,m}(n) n/m} \\ &\to \exp\left\{ -\frac{1}{m} \sum_{i=1}^m \int_{\mathbb{R}^d} \left(1 - e^{-f(i/m,\mathbf{x})}\right) \mu(d\mathbf{x}) \right\} \end{aligned} \tag{4.58}$$

as $n \to \infty$. The latter expression converges, as $m \to \infty$, to the expression on the right-hand side of (4.54), since the Riemann sums converge to the corresponding integral. Therefore, the claim of the theorem follows from the following calculation:

$$\begin{aligned}
&\left| E\exp\left\{-\int_{[0,1]\times\mathbb{R}^d} f(t,\mathbf{x})\,N_n(dt,\,d\mathbf{x})\right\} - E\exp\left\{-\int_{[0,1]\times\mathbb{R}^d} f_m(t,\mathbf{x})\,N_n(dt,\,d\mathbf{x})\right\}\right| \\
&\quad \le E\left|\int_{[0,1]\times\mathbb{R}^d} f(t,\mathbf{x})\,N_n(dt,\,d\mathbf{x}) - \int_{[0,1]\times\mathbb{R}^d} f_m(t,\mathbf{x})\,N_n(dt,\,d\mathbf{x})\right| \\
&\quad \le \sup_{(t,\mathbf{x})}\left|f(t,\mathbf{x}) - f_m(t,\mathbf{x})\right| E\left|\left\{j=1,\dots,n:\ \|\mathbf{X}_j/a_n\| > \varepsilon\right\}\right| \\
&\quad = \sup_{(t,\mathbf{x})}\left|f(t,\mathbf{x}) - f_m(t,\mathbf{x})\right| nP\big(\|\mathbf{X}_1\| > \varepsilon a_n\big)\,,
\end{aligned}$$

since the latter expression is a product of two terms, one of which converges to a finite limit as $n \to \infty$ by the definition of the sequence (a_n), and the other converges to zero as $m \to \infty$ by (4.57). □

Theorem 4.4.1 has a very simple and useful intuitive meaning. Dividing the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ by a large number a_n makes each individual observation small (in norm) with a very high probability. The number n of observations is exactly large enough that the exceptionally large observations (with norm of order at least a_n) arrange themselves into a Poisson random measure. The fact that the mean measure of the limiting Poisson random measure has a uniform factor corresponding to the time coordinate t reflects the obvious point that in an i.i.d. sample, the exceptionally large observations are equally likely to be found anywhere in the sample. On the other hand, the factor of the mean measure corresponding to the space coordinate $\mathbf{x}$ is the tail measure of the observations. This measure puts an infinite mass in every neighborhood of the origin, reflecting the obvious point that most observations are not exceptionally large, and hence after the normalization, will appear near the origin.

The next theorem presents a result related to Theorem 4.4.1. In the new setup, the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$ are still divided by a large number, but the number is now much smaller than a_n. This makes for a growing number of the corresponding exceptionally large observations. In order to obtain a sensible limit, one must divide the entire point process by an appropriately large number. This has an averaging effect, resulting in a nonrandom limit.

Let (k_n) be a sequence of positive numbers satisfying

$$k_n \to \infty, \quad k_n/n \to 0 \ \text{ as } n \to \infty. \tag{4.59}$$

Define

$$M_n = \frac{1}{k_n}\sum_{j=1}^{n} \delta_{\mathbf{X}_j/a_{nk_n^{-1}}}\,,\ n = 1, 2, \dots\,. \tag{4.60}$$

Note that the notation a_v for a real number $v \ge 1$ really means $a_{\lfloor v \rfloor}$. Further, the fact that M_n in (4.60) has atoms with a weight generally different from 1 means that

it is no longer a point process (but it is still, clearly, a random Radon measure). It is sometimes called the *tail empirical measure*.

Theorem 4.4.2. *Let* $(\mathbf{X}_1, \mathbf{X}_2 \ldots)$ *be a sequence of i.i.d. regularly varying random vectors in* $\mathbb{R}^d$*, with a tail index* $\alpha > 0$ *and tail measure* μ*, and let* (k_n) *be a sequence of positive numbers satisfying* (4.59)*. Then*

$$M_n \Rightarrow \mu \quad as\ n \to \infty \tag{4.61}$$

weakly in $M_+^{\mathrm{R}}\left(\overline{\mathbb{R}}_0^d\right)$.

Proof. We proceed as in the proof of Theorem 4.4.1. By Theorem 10.2.13, it is enough to prove convergence of the Laplace transforms, i.e., that

$$\lim_{n\to\infty} E \exp\left\{-\int_{\mathbb{R}^d} f(\mathbf{x})\, M_n(d\mathbf{x})\right\} = \exp\left\{-\int_{\mathbb{R}^d} f(\mathbf{x})\, \mu(d\mathbf{x})\right\} \tag{4.62}$$

for every bounded and uniformly continuous function f on $\overline{\mathbb{R}}_0^d$ that vanishes in a neighborhood of the origin. Since

$$E \exp\left\{-\int_{\mathbb{R}^d} f(\mathbf{x})\, M_n(d\mathbf{x})\right\} = \left(E \exp\left\{-\frac{1}{k_n} f\left(\mathbf{X}_1 / a_{nk_n^{-1}}\right)\right\}\right)^n,$$

(4.62) will follow from the statement

$$\lim_{n\to\infty} n E\left(1 - \exp\left\{-\frac{1}{k_n} f\left(\mathbf{X}_1 / a_{nk_n^{-1}}\right)\right\}\right) = \int_{\mathbb{R}^d} f(\mathbf{x})\, \mu(d\mathbf{x})\,.$$

This statement is equivalent (since the function f is bounded and $k_n \to \infty$) to the statement

$$\lim_{n\to\infty} n k_n^{-1} E f\left(\mathbf{X}_1 / a_{nk_n^{-1}}\right) \to \int_{\mathbb{R}^d} f(\mathbf{x})\, \mu(d\mathbf{x})\,,$$

which follows immediately from the fact that $m_{\lfloor nk_n^{-1} \rfloor} \xrightarrow{v} \mu$, as in the proof of (4.56). □

Notice that we have eliminated the time coordinate t from the definition of the tail empirical measure M_n in (4.60). This is because in the limit, all the atoms of M_n disappear, and it is no longer of interest to keep track of when the exceptionally large observations appear. Exercise 4.6.9 describes what happens when one insists on preserving the time coordinate.

4.5 Comments on Chapter 4

Comments on Section 4.1

The general result regarding sums with random numbers of terms of Corollary 4.1.11 and the subsequent discussion is due to Embrechts et al. (1979); the special case of geometric sums was discussed even earlier in Pakes (1975) and Teugels (1975). The paper Embrechts et al. (1979) describes a number of additional important properties of the family of subexponential distributions, including its closure under convolution roots: if $X_1 + \ldots + X_n$ is subexponential for some n, then so is X.

Comments on Section 4.2

The importance of regular variation in probability became clear after Feller (1968, 1971) and deHaan (1970).

Balanced regular variation of the tails of real-valued random variable with exponent $0 < \alpha < 2$ turns out to be a necessary and sufficient condition for this random variable, or its law, to be in the domain of attraction of an α-stable distribution; for the domain of attraction of the normal distribution, a condition can be stated in terms of the slow variation of the truncated second moment of the random variable. See Theorem 2, p. 577, Feller (1971).

Regular variation of the tails of infinite weighted sums of i.i.d. random variables with regularly varying tails has been considered by many authors. A frequently cited result was given in the PhD thesis Cline (1983). The result in Corollary 4.2.12 was proved in Mikosch and Samorodnitsky (2000).

Comments on Section 4.3

Most of the content of Theorem 4.3.7 showing that in some cases, regular variation of the linear combinations of the components of a random vector implies multivariate regular variation of the vector itself, is due to Basrak et al. (2002). In Hult and Lindskog (2006) one can find examples of random vectors that are not multivariate regularly varying, for which all linear combinations of the coordinates are regularly varying with any given integer exponent of regular variation (their examples are based on unpublished notes of H. Kesten), as well as an example of a random vector taking values in $[0, \infty)^d$ that does not have the property of multivariate regular variation for which all linear combinations of coordinates with nonnegative weights are regularly varying, with any given integer exponent of regular variation. A related example for $\alpha = 2$ is in Basrak et al. (2002).

Inverse problems for regular variation have been studied in Jacobsen et al. (2008) and Damek et al. (2014). An example of such an inverse problem is a converse to Proposition 4.2.6: if X and Y are independent, XY is regularly varying, and Y has a sufficiently light tail, does it follow that X is regularly varying? Similar questions can be asked about the multivariate version of Proposition 4.2.6 in Proposition 4.3.6, or about certain versions of Theorem 4.2.10 and its multivariate analogues. The "cancellation property" of the powers of a normal random variable in Remark 4.3.8 is an example of such a result.

Comments on Section 4.4

S. Resnick introduced the idea of using weak convergence of point processes to a Poisson limit in order to derive various limit theorems for regular varying random vectors in Resnick (1986), and the monograph Resnick (1987) contains a detailed exposition; some of it is also in Resnick (2007). Versions of Theorem 4.4.1 have been proved for stationary sequences of regularly varying random vectors, under weak dependence assumptions, including a version of strong mixing assumptions; see Davis and Hsing (1995), Davis and Mikosch (1998). Such theorems have also been established for specific families of stationary sequences of regularly varying random vectors, such as certain ARMA and GARCH processes; see Davis and Resnick (1985), Mikosch and Stărică (2000b). A general feature of these results is that the limiting process is no longer Poisson, but rather *cluster Poisson*, whereby each point of the Poisson process N_* in Theorem 4.4.1 is replaced by a cluster of secondary points generated by the Poisson point. Completely different behavior is possible for long-range dependent stationary sequences of regularly varying random vectors; an example is in Resnick and Samorodnitsky (2004).

Convergence of the empirical tail measure to the tail measure is often used to prove consistency of tail estimators; see, e.g., Resnick (2007). Theorem 4.4.2 also holds for certain non-i.i.d. stationary sequences of regularly varying random vectors; an example is in Resnick and Stărică (1995).

4.6 Exercises to Chapter 4

Exercise 4.6.1. *Prove the statements of Proposition 4.1.8 and Corollary 4.1.11 for real-valued subexponential random variables.*

Exercise 4.6.2. *Prove that the random variables in Example 4.1.18 are subexponential.*

Exercise 4.6.3. *Let X be a nonnegative regularly varying random variable. Define*

$$q(t) = \inf\{x > 0 : P(\|\mathbf{X}\| > x) \le t\}, \ 0 < t < 1 .$$

Then

$$\lim_{t \to 0} \frac{1}{t} P\big(\|\mathbf{X}\| > q(t)\big) = 1 .$$

Exercise 4.6.4. *Let X be a nonnegative regularly varying random variable with tail index α. Let $0 < p \le \alpha$. If $p = \alpha$, assume that $EX^p < \infty$. Then the upper truncated moment*

$$M_p(x) = E\big(X^p \mathbf{1}(X > x)\big), \ x \ge 0 ,$$

is regularly varying with exponent $p-\alpha$, and

$$\lim_{x\to\infty} \frac{E\big(X^p \mathbf{1}(X>x)\big)}{x^p P(X>x)} = \frac{\alpha}{\alpha-p},$$

with $\alpha/0 = +\infty$.

Exercise 4.6.5. *Prove Corollary 4.2.9.*

Exercise 4.6.6. *Give an example of a complete separable metric on $\overline{\mathbb{R}}_0^d$ that metrizes its topology.*

Exercise 4.6.7 (Basrak et al. (2002)). *Suppose that $\mathbf{X}$ takes values in $[0,\infty)^d$ and satisfies (4.42) for all vectors $\mathbf{b}$* with nonnegative coordinates. *Assume that $\alpha > 0$ is a noninteger. Replacing in the proof of Theorem 4.3.7 characteristic functions by Laplace transforms, prove that $\mathbf{X}$ is multivariate regularly varying.*

Exercise 4.6.8. *Let $X_1, X_2, \ldots$ be a sequence of one-dimensional random variables with balanced regularly varying tails with index $\alpha > 0$. Show that the limiting Poisson random measure N_* in Theorem 4.4.1 has the representation in distribution as*

$$N_* = \sum_{i=1}^{\infty} \delta_{(U_i,\, \theta_i \Gamma_i^{-1/\alpha})},$$

where (U_i) is a sequence of i.i.d. standard uniform random variables, (θ_i) is a sequence of i.i.d. random variables taking the value 1 with probability p, and the value -1 with probability $q = 1-p$ (in the notation of Definition 4.2.7), and (Γ_i) is a sequence of standard Poisson arrivals on $(0,\infty)$. All three sequences are assumed to be independent.

Exercise 4.6.9. *Let (k_n) be a sequence of positive numbers satisfying (4.59). For a sequence $(\mathbf{X}_1, \mathbf{X}_2 \ldots)$ of i.i.d. regularly varying random vectors in $\mathbb{R}^d$, with a tail index $\alpha > 0$ and tail measure μ, define random Radon measures on $[0,1]\times\overline{\mathbb{R}}_0^d$ by*

$$\tilde{M}_n = \frac{1}{k_n} \sum_{j=1}^{n} \delta_{(j/n,\, \mathbf{X}_j / a_{nk_n^{-1}})}, \quad n = 1, 2, \ldots .$$

Prove that $\tilde{M}_n \Rightarrow \lambda \times \mu$ weakly in $M_+^{\mathrm{R}}\big([0,1]\times\overline{\mathbb{R}}_0^d\big)$ endowed with the vague topology.

Chapter 5
Introduction to Long-Range Dependence

5.1 The Hurst Phenomenon

The history of long-range dependence as a concrete phenomenon believed to be important in its own right should be regarded as beginning in the 1960s with a series of papers by Benoit Mandelbrot and his coworkers, such as Mandelbrot (1965) and Mandelbrot and Wallis (1968). The cause was a need to explain an empirical finding by Hurst (1951, 1956) that studied the flow of water in the Nile. A particular data set studied by Hurst appears in Figure 5.1.

Many features of this data set are interesting (one of which is how long ago the data were collected). Harold Hurst, who was interested in the design of dams, looked at these data through a particular statistic. Given a sequence of n observations $X_1, X_2, \ldots, X_n$, define the partial sum sequence $S_m = X_1 + \ldots + X_m$ for $m = 0, 1, \ldots$ (with $S_0 = 0$). The statistic Hurst calculated is

$$\frac{R}{S}(X_1, \ldots, X_n) = \frac{\max_{0 \le i \le n}(S_i - \frac{i}{n}S_n) - \min_{0 \le i \le n}(S_i - \frac{i}{n}S_n)}{(\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2)^{1/2}}. \tag{5.1}$$

Note that S_n/n is the sample mean of the data. Therefore, $\max_{0 \le i \le n}(S_i - \frac{i}{n}S_n)$, for example, measures how far the partial sums rise above the straight line they would follow if all observations were equal (to the sample mean), and the difference between the maximum and the minimum of the numerator in (5.1) is the difference between the highest and lowest positions of the partial sums with respect to the straight line of uniform growth. It is referred to as the range of observations. The denominator of (5.1) is, of course, the sample standard deviation. The entire statistic in (5.1) has then been called the *rescaled range* or R/S *statistic*.

When Harold Hurst calculated the R/S statistic on the Nile data in Figure 5.1, he found that it grew as a function of the number n of observations, approximately as $n^{0.74}$.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_5

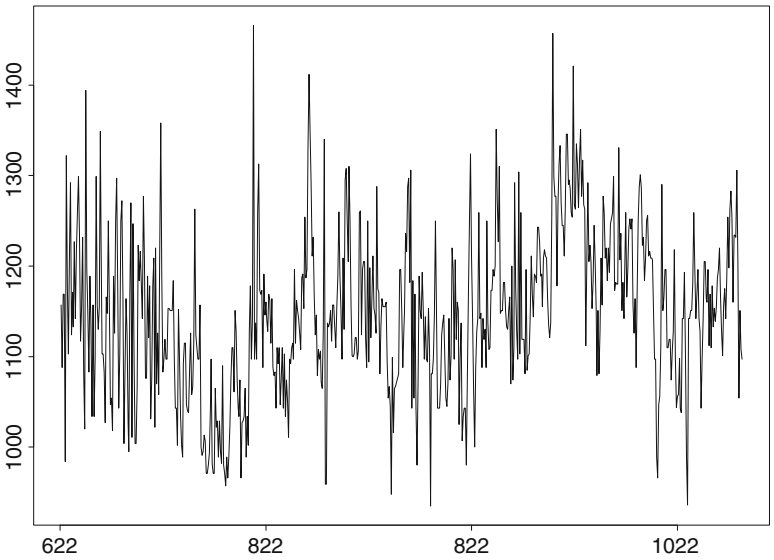

Fig. 5.1 Annual minima of the water level in the Nile for the years 622 to 1281, measured at the Roda gauge near Cairo

To see that this observation is interesting, let us suppose that $X_1, X_2, \ldots$ is a sequence of random variables. If we apply the R/S statistic to the first n observations $X_1, X_2, \ldots, X_n$ for increasing values of n, what would we expect the resulting sequence of values of the R/S statistic to be like, for the "usual" models of $X_1, X_2, \ldots$?

Example 5.1.1. Suppose that $X_1, X_2, \ldots$ is, in fact, a stationary sequence of random variables with a finite variance and a common mean μ. Define the centered partial sum process by

$$S^{(n)}(t) = S_{[nt]} - [nt]\mu, \ 0 \le t \le 1\,. \tag{5.2}$$

The classical functional central limit theorem (Donsker's theorem, invariance principle) says that if $X_1, X_2, \ldots$ are i.i.d., then

$$\frac{1}{\sqrt{n}} S^{(n)} \Rightarrow \sigma_* B \quad \text{weakly in } D[0,1], \tag{5.3}$$

where σ_*^2 is equal to the common variance σ^2 of the observations, and B is the standard Brownian motion on $[0,1]$ (Theorem 14.1 in Billingsley (1999)). Here $D[0,1]$ is the space of right continuous functions on $[0,1]$ having left limits equipped with the Skorokhod J_1 topology. In fact, the functional central limit theorem is known to hold for stationary processes with a finite variance that are much more

general than an i.i.d. sequence (with the limiting standard deviation σ_* not equal, in general, to the standard deviation of the X_i); see a survey by Merlevéde et al. (2006).

The function $f : D[0, 1] \to \mathbb{R}$ defined by

$$f(\mathbf{x}) = \sup_{0\le t\le 1} (x(t) - tx(1)) - \inf_{0\le t\le 1} (x(t) - tx(1)) ,$$

$\mathbf{x} = (x(t),\ 0 \le t \le 1) \in D[0, 1]$, is easily seen to be continuous. It is straightforward to check that the range of the first n observations (the numerator in the R/S statistic) is equal to $f(S^{(n)})$. Therefore, if the invariance principle (5.3) holds, then by the continuous mapping theorem, Theorem 10.2.4,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}(\text{the range of the first } n \text{ observations}) = f\left(\frac{1}{\sqrt{n}} S^{(n)}\right) \\
&\Rightarrow f(\sigma_* B) = \sigma_* \Big[\sup_{0\le t\le 1} (B(t) - tB(1)) - \inf_{0\le t\le 1} (B(t) - tB(1)) \Big] \\
&\qquad := \sigma_* \big[\sup_{0\le t\le 1} B_0(t) - \inf_{0\le t\le 1} B_0(t) \big],
\end{aligned}$$

where B_0 is a Brownian bridge on $[0, 1]$. Further, if the stationary process $X_1, X_2, \ldots$ (or its bilateral extension in Proposition 1.1.2) is ergodic, then by the pointwise ergodic theorem, Theorem 2.1.1 (or (2.8)), we have, with probability 1,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} (X_i - \frac{1}{n} S_n)^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \\
&\to E\big(X_1^2\big) - \Big(E\big(X_1\big)\Big)^2 = \sigma^2 .
\end{aligned}$$

Assuming, therefore, that the functional central limit theorem holds, and that the observations form an ergodic process, we see that

$$\frac{1}{\sqrt{n}} \frac{R}{S}(X_1, \ldots, X_n) \Rightarrow \frac{\sigma_*}{\sigma} \big[\sup_{0\le t\le 1} B_0(t) - \inf_{0\le t\le 1} B_0(t) \big]. \tag{5.4}$$

That is, the R/S statistic grows, distributionally, as the square root of the sample size.

The distributional $n^{0.5}$ rate of growth of the values of the R/S statistic obtained under, apparently quite reasonable, assumptions of Example 5.1.1 looks incompatible with the empirical $n^{0.74}$ rate of growth Hurst observed on the Nile data. Therefore, if one wants to construct a stochastic model of observations with a similar behavior of the R/S statistic to the one observed by Hurst, some of the "reasonable" assumptions of the example must be dropped.

The following example looks at what happens when the assumption of finite variance of the observations is dropped and replaced by an assumption of appropriately heavy tails.

Example 5.1.2. In order to be able to concentrate better on the effect of heavy tails, we will assume that the observations $X_1, X_2, \ldots$ are i.i.d. Assume that the balanced regular variation property of Definition 4.2.7 holds, and $0 < \alpha < 2$; this guarantees that the variance of the observations is infinite. In this case, we will apply the Poisson convergence result in Theorem 4.4.1 to understand the "size" of the R/S statistic.

We begin with a typical "truncation" step, needed because various sums of points are not continuous functionals of point processes in the topology of vague convergence. For $\epsilon > 0$, let

$$S_m^{(\epsilon)} = \sum_{j=1}^{m} X_j \mathbf{1}\big(|X_j| > \epsilon a_n\big), \ m = 0, 1, 2 \ldots ,$$

where, as usual, $a_n = \inf\{x > 0 : P(|X_1| > x) \le 1/n\}$, $n = 1, 2, \ldots$. Consider a modified version of the R/S statistic defined by

$$RS_n(\epsilon) = \frac{\max_{0\le i\le n}(S_i^{(\epsilon)} - \frac{i}{n}S_n^{(\epsilon)}) - \min_{0\le i\le n}(S_i^{(\epsilon)} - \frac{i}{n}S_n^{(\epsilon)})}{(\sum_{i=1}^{n} X_i^2 \mathbf{1}\big(|X_j| > \epsilon a_n\big))^{1/2}}. \tag{5.5}$$

Note that $RS_n(\epsilon) = g_\epsilon(N_n)$, where N_n is the point process in (4.52) and $g_\epsilon : M_+^{\mathrm{R}}\big([0,1] \times \overline{\mathbb{R}}_0^d\big) \to (0, \infty)$ is defined by

$$g_\epsilon(K) = \frac{R_\epsilon(K)}{(\int_{[0,1]\times(\mathbb{R}\setminus[-\epsilon,\epsilon])} y^2 K(ds, dy))^{1/2}},$$

with

$$R_\epsilon(K) = \sup_{0\le t\le 1}\left[\int_{[0,t]\times(\mathbb{R}\setminus[-\epsilon,\epsilon])} yK(ds, dy) - t\int_{[0,1]\times(\mathbb{R}\setminus[-\epsilon,\epsilon])} yK(ds, dy)\right]$$
$$- \inf_{0\le t\le 1}\left[\int_{[0,t]\times(\mathbb{R}\setminus[-\epsilon,\epsilon])} yK(ds, dy) - t\int_{[0,1]\times(\mathbb{R}\setminus[-\epsilon,\epsilon])} yK(ds, dy)\right].$$

According to Exercise 4.6.8, the law of the limiting Poisson process in Theorem 4.4.1 does not charge the set of the discontinuities of the function g_ϵ; see Exercise 5.5.1. Therefore, by the continuous mapping theorem (Theorem 10.2.4),

$$g_\epsilon(N_n) \Rightarrow g_\epsilon(N) \quad \text{in } [0, \infty) \text{ as } n \to \infty,$$

and we can represent the limit distributionally as

$$g_\epsilon(N) = \frac{\sup_{0\le t\le 1} Y_\epsilon(t) - \inf_{0\le t\le 1} Y_\epsilon(t)}{(\sum_{j=1}^{\infty} \Gamma_j^{-2/\alpha} \mathbf{1}(\Gamma_j < \epsilon^{-\alpha}))^{1/2}}\,.$$

Here

$$Y_\epsilon(t) = \sum_{j=1}^{\infty} \big(\mathbf{1}(U_j \le t) - t\big)\theta_j \Gamma_j^{-1/\alpha} \mathbf{1}(\Gamma_j < \epsilon^{-\alpha}),\ \ 0 \le t \le 1\,. \tag{5.6}$$

Recall that (U_i) is a sequence of i.i.d. standard uniform random variables, (θ_i) is a sequence of i.i.d. random variables taking the value 1 with probability p, and the value -1 with probability $q = 1 - p$, and (Γ_i) is a sequence of standard Poisson arrivals on $(0, \infty)$, with all three sequences being independent.

Corollary 3.4.2 says that Y_ϵ is an infinitely divisible process. Furthermore, if we take $0 < \epsilon_1 < \epsilon_2$ and use the same random ingredients in (5.6) for the two processes, Y_{ϵ_1} and Y_{ϵ_2}, then the difference $Y_{\epsilon_1} - Y_{\epsilon_2}$ can be written in the form

$$Y_{\epsilon_1}(t) - Y_{\epsilon_2}(t) = L_{\epsilon_1,\epsilon_2}(t) - tL_{\epsilon_1,\epsilon_2}(1),\ \ 0 \le t \le 1,$$

where

$$L_{\epsilon_1,\epsilon_2}(t) = \sum_{j=1}^{\infty} \mathbf{1}(U_j \le t)\theta_j \Gamma_j^{-1/\alpha} \mathbf{1}\big(\epsilon_2^{-\alpha} \le \Gamma_j < \epsilon_1^{-\alpha}\big),\ \ 0 \le t \le 1\,.$$

By Corollary 3.4.2, $L_{\epsilon_1,\epsilon_2}$ is a Lévy process without a Gaussian component, whose one-dimensional Lévy measure $\rho_{\epsilon_1,\epsilon_2}$ is the measure

$$m(dx) = \big(p\mathbf{1}(x > 0) + q\mathbf{1}(x < 0)\big)\alpha|x|^{-(\alpha+1)}\,dx$$

restricted to the set $(-\epsilon_2, -\epsilon_1) \cup (\epsilon_1, \epsilon_2)$, and whose local shift is

$$b = (2p - 1)\int_{\epsilon_2^{-\alpha}}^{\epsilon_1^{-\alpha}} [\![x^{-1/\alpha}]\!]\,dx\,;$$

see Example 3.2.3. It is, in fact, a compound Poisson Lévy process. It follows from the general properties of Lévy processes that $\sup_{0\le t\le 1} |L_{\epsilon_1,\epsilon_2}(t)| \to 0$ in probability as $\epsilon_2 \to 0$, uniformly in $0 < \epsilon_1 < \epsilon_2$; see Kallenberg (1974). Therefore, the same is true for $\sup_{0\le t\le 1} |Y_{\epsilon_1}(t) - Y_{\epsilon_2}(t)|$. We conclude that

$$Y_\epsilon(t) \to Y(t) := \sum_{j=1}^{\infty} \big(\mathbf{1}(U_j \le t) - t\big)\theta_j \Gamma_j^{-1/\alpha},\ \ 0 \le t \le 1,\ \ \text{as } \epsilon \to 0$$

in probability in the uniform topology in the space $D([0,1])$. Therefore, as $\epsilon \to 0$,

$$g_\epsilon(N) \to \frac{\sup_{0\le t\le 1}\sum_{j=1}^{\infty}\big(\mathbf{1}(U_j \le t) - t\big)\theta_j \Gamma_j^{-1/\alpha} - \inf_{0\le t\le 1}\sum_{j=1}^{\infty}\big(\mathbf{1}(U_j \le t) - t\big)\theta_j \Gamma_j^{-1/\alpha}}{(\sum_{j=1}^{\infty}\Gamma_j^{-2/\alpha})^{1/2}} \tag{5.7}$$

in probability. Notice that by the strong law of large numbers, $\Gamma_j/j \to 1$ as $j \to \infty$ with probability 1. This and the fact that $0 < \alpha < 2$ imply that the dominator on the right-hand side is finite and justifies the convergence.

Let $g(N)$ denote the random variable on the right-hand side of (5.7). If one shows that the statement

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty} P\left(\left|\frac{1}{\sqrt{n}}\frac{R}{S}(X_1,\dots,X_n) - RS_n(\epsilon)\right| > \lambda\right) = 0 \tag{5.8}$$

for every $\lambda > 0$ is true, then a standard weak convergence argument, e.g., in Theorem 3.2 in Billingsley (1999), allows us to conclude that

$$\frac{1}{\sqrt{n}}\frac{R}{S}(X_1,\dots,X_n) \Rightarrow g(N)\,. \tag{5.9}$$

Note that (5.9) means that even in the heavy-tailed case, the R/S statistic grows as the square root of the sample size.

The validity of (5.8) is verified in Exercise 5.5.2.

We conclude, therefore, as was done in Mandelbrot and Taqqu (1979), that infinite variance alone cannot explain the Hurst phenomenon. A different drastic departure from the assumptions leading to the square root of the sample size rate of growth of the R/S statistic was suggested in Mandelbrot (1965), and it had nothing to do with heavy tails. The idea was, instead, to take as a model a stationary process with a finite variance, but with correlations decaying so slowly as to invalidate the functional central limit theorem (5.3). The simplest model of that sort is *fractional Gaussian noise*, which is the increment process of fractional Brownian motion.

Let us begin with a fractional Brownian motion, or *FBM*, constructed in Example 3.5.1. This is a zero-mean Gaussian process $(B_H(t),\ t \ge 0)$ that is self-similar with exponent of self-similarity $H \in (0,1)$ and stationary increments. These properties imply that $B_H(0) = 0$ and $E(B_H(t) - B_H(s))^2 = \sigma^2|t-s|^{2H}$ for some $\sigma > 0$; see Section 8.2. Taking an appropriately high moment of the increment and using the Kolmogorov criterion in Theorem 10.7.7 allows us to conclude that a fractional Brownian motion has a continuous version, and we always assume that we are working with such a version.

A fractional Gaussian noise, or *FGN*, is a discrete step increment process of a fractional Brownian motion defined by $X_j = B_H(j) - B_H(j-1)$ for $j = 1, 2, \dots$. The stationarity of the increments of the FBM implies that this is a stationary Gaussian

process. Using the fact $ab = (a^2 + b^2 - (a-b)^2)/2$ and the incremental variance of the FBM, we easily see that

$$\mathrm{Cov}(X_{j+n}, X_j) = \frac{\sigma^2}{2}\Big[(n+1)^{2H} + |n-1|^{2H} - 2n^{2H}\Big] \tag{5.10}$$

for $j \geq 1$, $n \geq 0$. That is,

$$\rho_n := \mathrm{Corr}(X_{j+n}, X_j) \sim H(2H-1)n^{-2(1-H)} \quad \text{as } n \to \infty. \tag{5.11}$$

In particular, $\rho_n \to 0$ as $n \to \infty$. This implies that the FGN is a mixing, hence ergodic, process; see Example 2.2.8. Furthermore, by the self-similarity of the fractional Brownian motion, for every n,

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}B_H(n) = \sigma n^{2H}\,. \tag{5.12}$$

Suppose now that a set of observations $X_1, X_2, \ldots$ forms a fractional Gaussian noise as defined above, and let us consider the behavior of the R/S statistic on these observations. The ergodicity of the FGN implies that the denominator of the statistic converges a.s. to the standard deviation of the observations, σ; see Example 2.1.5. For the numerator of the R/S statistic, we notice that $S_i = B_H(i)$ for every i, and the self-similarity of the FBM gives us

$$\begin{aligned}
&\max_{0\leq i\leq n}(S_i - \frac{i}{n}S_n) - \min_{0\leq i\leq n}(S_i - \frac{i}{n}S_n)\\
&= \max_{0\leq i\leq n}(B_H(i) - \frac{i}{n}B_H(n)) - \min_{0\leq i\leq n}(B_H(i) - \frac{i}{n}B_H(n))\\
&\stackrel{d}{=} n^H\left[\max_{0\leq i\leq n}(B_H(\frac{i}{n}) - \frac{i}{n}B_H(1)) - \min_{0\leq i\leq n}(B_H(\frac{i}{n}) - \frac{i}{n}B_H(1))\right].
\end{aligned}$$

By the continuity of the sample paths of the fractional Brownian motion, we have

$$\begin{aligned}
&\max_{0\leq i\leq n}(B_H(\frac{i}{n}) - \frac{i}{n}B_H(1)) - \min_{0\leq i\leq n}(B_H(\frac{i}{n}) - \frac{i}{n}B_H(1))\\
&\to \sup_{0\leq t\leq 1}(B_H(t) - tB_H(1)) - \inf_{0\leq t\leq 1}(B_H(t) - tB_H(1))
\end{aligned}$$

with probability 1. That is, for the FGN,

$$n^{-H}\frac{R}{S}(X_1, \ldots, X_n) \Rightarrow \sup_{0\leq t\leq 1}(B_H(t) - tB_H(1)) - \inf_{0\leq t\leq 1}(B_H(t) - tB_H(1))\,,$$

and so the R/S statistic grows distributionally at the rate n^H as a function of the sample size. Therefore, selecting an appropriate H in the model will, finally, explain

the Hurst phenomenon. In particular, the exponent H of self-similarity of fractional Brownian motion is often referred to as a *Hurst parameter*.

This success of the fractional Gaussian noise model in explaining the Hurst phenomenon is striking. We have used the self-similarity of the fractional Brownian motion in the above computation, but it is not hard to see that a very important property of the fractional Gaussian noise is the unusually slow decay of correlations in (5.11), especially for high values of H (i.e., close to 1). For these values of H, the variance of the partial sums in (5.12) also increases unusually fast. Unlike the previous unsuccessful attempt to explain the Hurst phenomenon by introducing in the model unusually heavy tails (infinite variance in this case), the FGN model succeeds here by introducing unusually long memory. Particularly vivid terminology was introduced in Mandelbrot and Wallis (1968), in the context of weather and precipitation: unusually heavy tails were called the *Noah effect*, referring to the biblical story of Noah and extreme incidents of precipitation, while unusually long memory was called the *Joseph effect*, referring to the biblical story of Joseph and long stretches (seven years) of time greater than average and less than average precipitation. This success of the FGN brought the fact that memory of a certain length can make a big difference to the attention of many. The terms "long-range dependent process" and "long memory" came into being; they can already be found in the early papers by Mandelbrot and coauthors.

5.2 The Joseph Effect and Nonstationarity

The Joseph effect is clearly visible in Figure 5.2: in the left plot, where the observations are those of fractional Gaussian noise with Hurst parameter $H = 0.8$, there are long stretches of time (hundreds of observations) during which the observations tend to be on one side of the true mean 0. This is clearly not the case on the right plot of i.i.d. normal observations. Returning momentarily to the Nile data in Figure 5.1, we see evidence of the Joseph effect there as well.

Such behavior of the observations obviously seems to indicate lack of stationarity, and in general, the relationship between long-range dependence and nonstationarity is delicate in a number of ways. We have seen that the Joseph effect involves long stretches of time when the process tends to be above the mean, and long stretches of time when the process tends to be below the mean. Quoting a description in Mandelbrot (1983), page 251, of a fractional Gaussian noise with $H > 1/2$: "Nearly every sample looks like a 'random noise' superimposed upon a background that performs several cycles, whichever the sample's duration. However, these cycles are *not* periodic, that is, *cannot* be extrapolated as the sample lengthens."

This discussion shows that in application to real data, either stationary long memory models or appropriate nonstationary models can be used in similar situations. There is, obviously, no "right" or "wrong" way to go here, beyond the principle of parsimony.

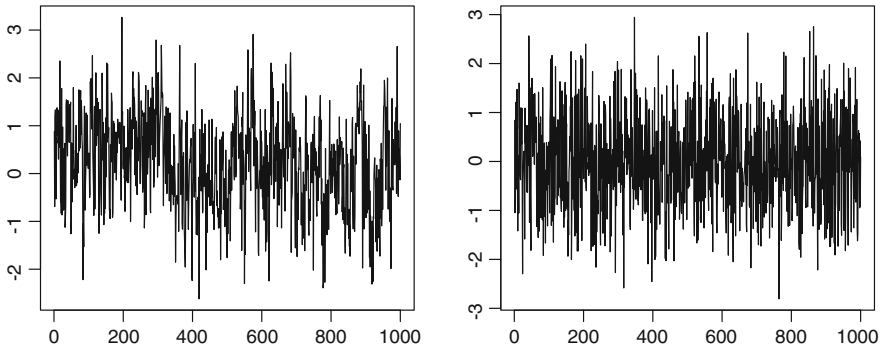

Fig. 5.2 Fractional Gaussian noise with $H = 0.8$ (left plot) and i.i.d. standard Gaussian random variables (right plot)

Among the first to demonstrate the difficulty of distinguishing between stationary long memory models and certain nonstationary models was the paper Bhattacharya et al. (1983), in which it was suggested that instead of fractional Gaussian noise or another model with long memory, the Hurst phenomenon can be explained by a simple nonstationary model as follows. Let $Y_1, Y_2, \ldots$ be a sequence of independent identically distributed random variables with a finite variance σ^2. Let $0 < \beta < 1/2$, choose $a \geq 0$, and consider the model

$$X_i = Y_i + (a + i)^{-\beta}, \; i = 1, 2, \ldots . \tag{5.13}$$

Clearly, the stochastic process $X_1, X_2, \ldots$ is nonstationary, for it contains a nontrivial drift. However, it is asymptotically stationary (as the time increases), and the drift can be taken to be very small to start with (by taking a to be large). This process has no memory at all, since the sequence $Y_1, Y_2, \ldots$ is i.i.d. It does, however, cause the R/S statistic to behave in the same way as if the sequence $X_1, X_2, \ldots$ were a fractional Gaussian noise, or another long-range dependent process.

To see why this is true, note that for this model, the numerator of the R/S statistic is bounded between

$$r_n - R_n^Y \leq \max_{0 \leq i \leq n} (S_i - \frac{i}{n} S_n) - \min_{0 \leq i \leq n} (S_i - \frac{i}{n} S_n) \leq r_n + R_n^Y ,$$

where

$$r_n = \max_{0 \leq i \leq n} (s_i - \frac{i}{n} s_n) - \min_{0 \leq i \leq n} (s_i - \frac{i}{n} s_n) ,$$

$$R_n^Y = \max_{0 \leq i \leq n} (S_i^Y - \frac{i}{n} S_n^Y) - \min_{0 \leq i \leq n} (S_i^Y - \frac{i}{n} S_n^Y) ,$$

and $S_m^Y = Y_1 + \ldots + Y_m$, $s_m = \sum_{j=1}^m (a + j)^{-\beta}$ for $m = 0, 1, 2, \ldots$.

Since s_m is a sum of a decreasing sequence of numbers, we see that $\min_{0\le i\le n}(s_i - \frac{i}{n}s_n) = 0$. On the other hand, by Theorem 10.5.6,

$$s_n \sim \frac{1}{1-\beta} n^{1-\beta} \quad \text{as } n \to \infty.$$

If we denote by i_n^* the value of i over which the maximum is achieved in $\max_{0\le i\le n}(s_i - \frac{i}{n}s_n)$, then we see that

$$i_n^* = \lfloor (s_n/n)^{-1/\beta} - a \rfloor \sim (1-\beta)^{1/\beta} n$$

as $n \to \infty$. Therefore,

$$\max_{0\le i\le n} (s_i - \frac{i}{n}s_n) = s_{i_n^*} - \frac{i_n^*}{n} s_n \sim \beta(1-\beta)^{1/\beta - 2}\, n^{1-\beta}\,,$$

so that

$$r_n \sim C_\beta\, n^{1-\beta},\ C_\beta = \beta(1-\beta)^{1/\beta-2}$$

as $n \to \infty$.

Recall that $Y_1, Y_2, \ldots$ are i.i.d. random variables with a finite variance. Therefore, the range R_n^Y of the first n observations from this sequence grows distributionally as $n^{1/2}$. We immediately conclude that

$$\frac{1}{n^{1-\beta}} \left[\max_{0\le i\le n} (S_i - \frac{i}{n}S_n) - \min_{0\le i\le n} (S_i - \frac{i}{n}S_n) \right] \to C_\beta$$

in probability as $n \to \infty$.

Similarly, for the denominator of the R/S statistic, we have a bound

$$D_n^Y - d_n \le \Big(\sum_{i=1}^n (X_i - \frac{1}{n}S_n)^2 \Big)^{1/2} \le D_n^Y + d_n\,,$$

where

$$D_n^Y = \Big(\sum_{i=1}^n (Y_i - \frac{1}{n}S_n^Y)^2 \Big)^{1/2}, \quad d_n = \Big(\sum_{i=1}^n \big((a+i)^{-\beta} - \frac{1}{n}s_n\big)^2 \Big)^{1/2}.$$

We know that $D_n^Y/n^{1/2} \to \sigma$ a.s. as $n \to \infty$, while an elementary computation using, for example, Theorem 10.5.6, leads to $d_n \sim C'_\beta\, n^{1/2-\beta}$ as $n \to \infty$ for some $0 < C'_\beta < \infty$. Therefore,

$$n^{-1/2} \Big(\sum_{i=1}^n (X_i - \frac{1}{n}S_n)^2 \Big)^{1/2} \to \sigma$$

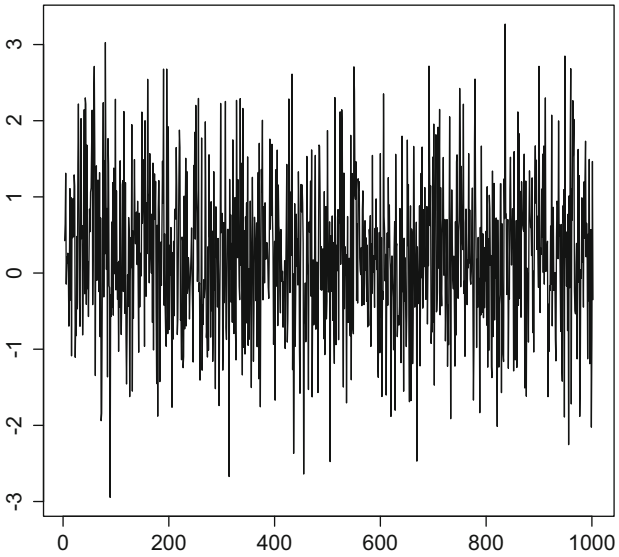

Fig. 5.3 Observations from the model (5.13) with standard normal noise, $a = 2$ and $\beta = 1/4$. No Joseph effect is visible

a.s., and we conclude that

$$\frac{1}{n^{1-\beta}}\frac{R}{S}(X_1, \ldots, X_n) \to \frac{C_\beta}{\sigma}$$

in probability as $n \to \infty$.

Therefore, for the i.i.d. model with small drift as in (5.13), the R/S statistic grows as $n^{1-\beta}$, the same rate as for the FGN with $H = 1 - \beta$, and so the R/S statistic cannot distinguish between these two models. Apart from fooling the R/S statistic, however, the model (5.13) is not difficult to tell apart from a stationary process with correlations decaying as in (5.11). Even visually, the observations from the model (5.13) do not appear to exhibit the Joseph effect, as the plot in Figure 5.3 indicates.

A very important class of nonstationary models that empirically resemble long-memory stationary models is that of *regime-switching models*. The name is descriptive, and it makes it clear where the lack of stationarity comes from. The fractional Gaussian noise also appears to exhibit different "regimes" (the Joseph effect), but the nonstationary regime-switching models are usually those with breakpoints, whose location changes with the sample size, in either a random or nonrandom manner.

One class of regime-switching models is obtained by taking a parametric model that would be stationary if its parameters were kept constant and then changing the parameters along a sequence of nonrandom time points, again chosen relative to the

sample size. Such a change can affect the mean and the variance (among many other things) of the process after breakpoints, and to many sample statistics this will look like long memory.

To see what might happen, consider a sample $X_1, \ldots, X_n$, where the observations come from $r \geq 2$ subsamples of lengths proportional to the overall sample size. That is, given fixed proportions $0 < p_i < 1$, $i = 1, \ldots, r$ with $p_1 + \ldots + p_r = 1$, the sample has the form

$$X_1^{(1)}, \ldots, X_{[np_1]}^{(1)}, X_{[np_1]+1}^{(2)}, \ldots, X_{[n(p_1+p_2)]}^{(2)}, \ldots, X_{[n(1-p_r)]+1}^{(r)}, \ldots, X_n^{(r)}, \tag{5.14}$$

where the ith subsample forms a stationary ergodic process with a finite variance, $i = 1, \ldots, r$. Since one of the common ways to try to detect long-range dependence is by looking for a slow decay of covariances and correlations, let us check the behavior of the sample covariance of the sample (5.14). Fix a time lag m and denote by $\hat{R}_m(n)$ the sample covariance at that lag based on the n observations in (5.14). Note that

$$\hat{R}_m(n) = \frac{1}{n}\sum_{j=1}^{n-m}(X_j - \bar{X})(X_{j+m} - \bar{X}) = A_m(n) + B_m(n)\,,$$

where $\bar{X} = (X_1 + \ldots + X_n)/n$ is the overall sample mean,

$$A_m(n) = \frac{1}{n}\sum_{j=1}^{n-m} X_j X_{j+m} - (\bar{X})^2\,,$$

and

$$B_m(n) = \frac{1}{n}\bar{X}\left(\sum_{j=1}^{m} X_j + \sum_{j=n-m+1}^{n} X_j\right) - \frac{m}{n}(\bar{X})^2\,.$$

By ergodicity, $\bar{X} \to \sum_{i=1}^{r} p_i\mu_i$, where μ_i is the mean of the ith subsample, $i = 1, \ldots, r$. Further, since m is fixed, $B_m(n) \to 0$ in probability as $n \to \infty$.

Finally, if I_i denotes the set of indices within $\{1, \ldots, n\}$ corresponding to the ith subsample, $i = 1, \ldots, r$, then by ergodicity,

$$\frac{1}{n}\sum_{j=1}^{n-m} X_j X_{j+m}$$

$$= \sum_{i=1}^{r} \frac{\mathrm{Card}(I_i \cap (I_i - m))}{n} \frac{1}{\mathrm{Card}(I_i \cap (I_i - m))} \sum_{j \in I_i \cap (I_i - m)} X_j^{(i)} X_{j+m}^{(i)}$$

$$+\frac{1}{n}\sum_{i=1}^{r}\sum_{\substack{j\in\{1,\ldots,n-m\}\\ j\in I_i, j+m\in I_{i+1}}} X_j X_{j+m}$$

$$\to \sum_{i=1}^{r} p_i E\left(X_1^{(i)} X_{1+m}^{(i)}\right) = \sum_{i=1}^{r} p_i (R_m^{(i)} + \mu_i^2)\,,$$

where $R_m^{(i)}$ is the covariance at lag m of the ith subsample. We conclude that

$$\hat{R}_m(n) \to \sum_{i=1}^{r} p_i \left(R_m^{(i)} + \mu_i^2\right) - \left(\sum_{i=1}^{r} p_i \mu_i\right)^2 \tag{5.15}$$

$$= \sum_{i=1}^{r} p_i R_m^{(i)} + \frac{1}{2}\sum_{i_1=1}^{r}\sum_{i_2=1}^{r} p_{i_1} p_{i_2} \left(\mu_{i_1} - \mu_{i_2}\right)^2$$

in probability as $n \to \infty$. What (5.15) indicates is that if there is regime-switching as we have described, and (some of) the mean values in different regimes are different, then the sample covariance function will tend to stabilize, at large, but fixed, lags at a positive value.

This is what is often observed in practice, and long memory is suspected. Of course, this regime-switching model is simply a deterministic way of mimicking the Joseph effect (recall Figure 5.2), and an example of this phenomenon can be seen in Figure 5.4, where $r = 4$, $p_1 = p_2 = p_3 = p_4 = 1/4$, and the four different stationary ergodic processes are all autoregressive processes of order 1, with normal innovations with the mean and the standard deviation both equal to 1. The autoregressive coefficients are $\psi_1 = 0.7$, $\psi_2 = 0.75$, $\psi_3 = 0.65$, $\psi_4 = 0.8$.

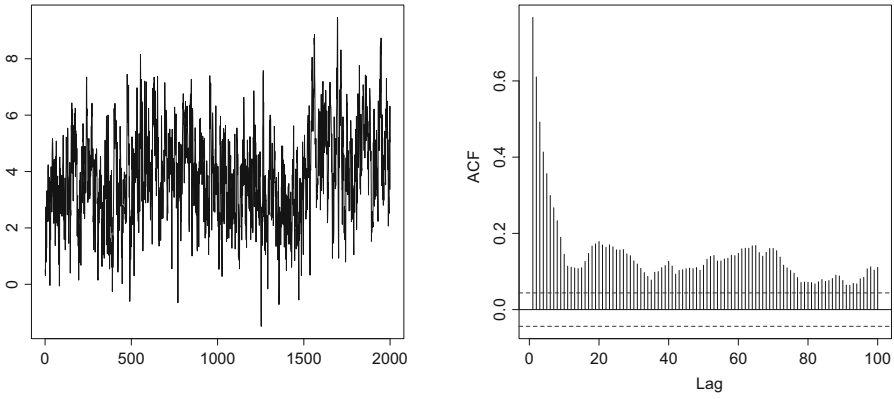

Fig. 5.4 Observations from a regime-switching AR(1) model (left plot) and their sample autocorrelation function (right plot)

5.3 Long Memory, Mixing, and Strong Mixing

The notion of memory in a stationary stochastic process is by definition related to the connections between certain observations and those occurring after an amount of time has passed. If $X_1, X_2, \ldots$ is the process, then the passage of time corresponds to the shifted process, $X_{k+1}, X_{k+2}, \ldots$, for a time shift k. In other words, the notion of memory is related to the connections between the process and its shifts. This makes the language of the ergodic theory of stationary processes, elements of which are outlined in Chapter 2, an attractive language for describing the memory of a stationary process.

We begin by observing that it is very natural to say that a nonergodic stationary process $\mathbf{X}$ *has infinite memory*. Indeed, a nonergodic process has the structure given in Proposition 2.1.6. That is, it is a mixture of the type

$$\left(X_n,\, n \in \mathbb{Z}\right) \stackrel{d}{=} \begin{cases} \left(Y_n,\, n \in \mathbb{Z}\right) & \text{with probability } p, \\ \left(Z_n,\, n \in \mathbb{Z}\right) & \text{with probability } 1-p, \end{cases}$$

where stationary processes $\left(Y_n,\, n \in \mathbb{Z}\right)$ and $\left(Z_n,\, n \in \mathbb{Z}\right)$ have different finite-dimensional distributions, and the choice with probability $0 < p < 1$ is made independently of the two stationary processes. This means that the result of a single "coin toss" (with probabilities p and $1-p$) will be "remembered forever." Therefore, it certainly makes sense to call stationary ergodic processes "processes with finite memory," and stationary nonergodic processes "processes with infinite memory."

It is very tempting to try to use another ergodic theoretical notion, stronger than ergodicity, such as weak mixing or mixing, for example, to define finite and short memory in a stationary process. Then ergodic stationary processes that lack this stronger property will be naturally called processes with long memory. If the property of mixing were used for this purpose, for example, then a long-range dependent process would be an ergodic but nonmixing process.

Such definitions of long-range dependence are possible, but they have not become standard, for reasons that will be discussed below. Before we do that, however, it is important to note that the approaches to memory of a stationary process via the ergodic theoretical properties of the corresponding shift transformation are very attractive from the following point of view. Let $\mathbf{X}$ be a stationary process, and let the process $\mathbf{Y}$ be derived from the process $\mathbf{X}$ by means of a point transformation $Y_n = g(X_n)$ for all n, where $g : \mathbb{R} \to \mathbb{R}$ is a measurable function. Clearly, $\mathbf{Y}$ is also a stationary process. It is intuitively clear that the process $\mathbf{X}$ "remembers at least as much" as the process $\mathbf{Y}$ does. If, in particular, g is a one-to-one map, and g^{-1} is also measurable, then this intuition says that the processes $\mathbf{X}$ and $\mathbf{Y}$ should have the "same length of memory": if one of them has long memory, then so should the other one.

This, apparently very natural, requirement has proved to be difficult to satisfy by many of the proposed definitions of long-range dependence. It is, however, automatic with ergodic theoretically based definitions. Indeed, it follows from

Corollary 2.2.5 and Proposition 2.2.14 that if $\mathbf{X}$ is mixing (respectively weak mixing), then the process $\mathbf{Y}$ with $Y_n = g(X_n)$ for all n is also mixing (respectively weak mixing). This would imply that short memory was preserved under a measurable map, and if the map is one-to-one, with a measurable inverse, then the map must preserve long memory as well.

It is instructive to record what the ergodic theoretically based notions of memory discussed above mean for stationary Gaussian processes. Let $\mathbf{X}$ be a (real-valued) stationary Gaussian process with covariance function R_k, $k \geq 0$ and spectral measure F on $(-\pi, \pi]$. That is, $R_k = \int_{(-\pi,\pi]} \cos(kx)\, F(dx)$ for $k \geq 0$. Then

- the process $\mathbf{X}$ is ergodic if and only if the spectral measure F is atomless;
- the process $\mathbf{X}$ is mixing if and only if $R_k \to 0$ as $k \to \infty$;

see Examples 2.2.8 and 2.2.18. The requirement that the covariance function vanish as the time lag increases, however, proved to be insufficient in dealing with long memory for Gaussian processes. Indeed, many "unusual" phenomena have been observed for Gaussian processes whose covariance functions vanish in the limit, but sufficiently slowly, as we have already seen in the example of fractional Gaussian noise. Therefore, the mixing property is not believed to be sufficiently strong to say that a stationary process with this property has short memory. A stronger requirement is needed.

For this purpose, strong mixing conditions, some of which are discussed in Section 2.3, have been used. A possible connection between strong mixing properties and lack of long memory (i.e., presence of short memory) has been observed, beginning with Rosenblatt (1956). We discuss results in this spirit in Comments to Chapter 9. Such results explain why the absence of one or another strong mixing condition (as opposed to ergodic-theoretical mixing) is sometimes taken as the definition of long-range dependence.

The strong mixing properties share with the ergodic-theoretical notions of ergodicity and mixing the following very desirable feature: if a process $\mathbf{Y}$ is derived from a process $\mathbf{X}$ by means of a one-to-one point transformation $Y_n = g(X_n)$ for all n, where $g : \mathbb{R} \to \mathbb{R}$ is a one-to-one function such that both g and g^{-1} are measurable, then the process $\mathbf{X}$ has long memory in the sense of lacking one of the strong mixing properties if and only if the process $\mathbf{Y}$ does; see Exercise 2.6.11.

The role that the strong mixing conditions play in eliminating the possibility of long-range dependence is real, but limited. Its effects are felt more in the behavior of the partial sums of a process than in, say, the behavior of the partial maxima.

Overall, the strong mixing conditions have not become standard definitions of absence of long-range dependence, i.e., of short memory. To some extent, this is due to the fact that the effect of strong mixing conditions is limited. More importantly, the strong mixing conditions are not easily related to the natural building blocks of many stochastic models and are difficult to verify, with the possible exception of Gaussian processes and Markov chains. Even in the latter cases, necessary and sufficient conditions are not always available, particularly for more complicated types of strong mixing.

5.4 Comments on Chapter 5

Comments on Section 5.2

The fact that the i.i.d. model with small drift as in (5.13) can be easily distinguished from the fractional Gaussian noise with $H > 1/2$ was shown in Künsch (1986) using the *periodogram*.

In Mikosch and Stărică (2016) and Mikosch and Stărică (2000a), the procedure of changing the parameters of an otherwise stationary model was applied to the short-memory GARCH(p, q) model, resulting in behavior resembling long-range dependence.

Various other regime-switching models mimicking long-range dependence are suggested in Diebold and Inoue (2001).

5.5 Exercises to Chapter 5

Exercise 5.5.1. *Show that the function g_ϵ in Example 5.1.2 is a function on $M_+^{\mathrm{R}}\left([0,1] \times \overline{\mathbb{R}}_0^d\right)$ that is continuous at all points at which the denominator in its definition does not vanish.*

Exercise 5.5.2. *In this exercise, we will check the validity of the statement (5.8). Write*

$$\frac{1}{\sqrt{n}} \frac{R}{S}(X_1, \ldots, X_n) = \frac{M_n - m_n}{D_n}, \quad RS_n(\epsilon) = \frac{M_n(\epsilon) - m_n(\epsilon)}{D_n(\epsilon)} .$$

(i) Use the maximal inequality in Theorem 10.7.4 to show that for some finite positive constant c,

$$P\left(\frac{1}{a_n} |M_n - M_n(\epsilon)| > \lambda\right) \le \frac{c}{\lambda} \frac{n^{1/2}}{a_n} \left(E\left(X_1^2 \mathbf{1}(|X_1| \le \epsilon a_n)\right)\right)^{1/2}$$

for each $\lambda > 0$ and $n = 1, 2, \ldots$. Next, use the estimate on the moments of truncated random variables in Proposition 4.2.3 and the fact that $0 < \alpha < 2$ to show that

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} P\left(\frac{1}{a_n} |M_n - M_n(\epsilon)| > \lambda\right) = 0$$

for every $\lambda > 0$. A similar argument proves that

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} P\left(\frac{1}{a_n} |m_n - m_n(\epsilon)| > \lambda\right) = 0$$

for every $\lambda > 0$.

(ii) Check that

$$|D_n - D_n(\epsilon)| \le \frac{|S_n|}{n^{1/2}} + \left(\sum_{i=1}^{n} X_i^2 \mathbf{1}\big(|X_i| \le \epsilon a_n\big) \right)^{1/2}$$

and conclude that for every $\lambda > 0$,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} P\left(\frac{1}{a_n} |D_n - D_n(\epsilon)| > \lambda \right) = 0 \,.$$

(iii) Use the truncation argument used in Example 5.1.2 and the already checked part of (5.8) to conclude that $a_n^{-1}(M_n - m_n)$ *converges weakly to the numerator of (5.7) and so the corresponding sequence of laws is tight.*

(iv) Show that

$$\lim_{M \to \infty} \limsup_{n \to \infty} \sup_{0 < \epsilon < 1} P\left(\frac{a_n}{D_n(\epsilon)} > M \right) = 0$$

and

$$\lim_{M \to \infty} \limsup_{n \to \infty} \sup_{0 < \epsilon < 1} P\left(\frac{a_n^2}{D_n D_n(\epsilon)} > M \right) = 0 \,.$$

(v) Put together the previous parts of the exercise to obtain (5.8).

Chapter 6
Second-Order Theory of Long-Range Dependence

6.1 Time-Domain Approaches

The second-order theory is the oldest and best developed among all points of view on long-range dependence. By necessity, it applies only to stationary processes with a finite variance. Let, therefore, $\mathbf{X}$ be a stationary process with a finite variance σ_X^2 and covariance function R_X. The time-domain second-order approaches to long-range dependence concentrate on two related issues.

Issue 1. How fast or slowly does the covariance function $R_X(n)$ decay as the lag n goes to ∞?

Issue 2. How fast or slowly does the partial sum variance $\mathrm{Var}S_n$ increase as the lag n goes to ∞?

Here, as usual, $S_n = X_1 + \ldots + X_n$ is the sum of the first n observations, $n = 1, 2, \ldots$.

The general idea is that short memory corresponds to a sufficiently fast rate of decay of the covariance function and a sufficiently slow rate of increase of the variance of the partial sums. Correspondingly, long-range dependence corresponds to a sufficiently slow rate of decay of the covariance function and a sufficiently fast rate of increase of the variance of the partial sums. Some intuition into this approach can be obtained by examining the corresponding rates for the fractional Gaussian noise in (5.11) and (5.12).

Notice that the terms "fast rate of decay/increase" and "slow rate of decay/increase" are vague. Furthermore, the two issues above are related, but not identical. One of our goals in the sequel is to understand the relationship between these issues.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_6

The reason the two issues are related is the following simple expression for the variance of the partial sum:

$$\mathrm{Var}S_n = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i, X_j) \tag{6.1}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} R_X(|i-j|) = nR_X(0) + 2\sum_{i=1}^{n-1}(n-i)R_X(i)\,.$$

The following elementary statement describes one way to make precise "fast rate of decay of covariances" and the implication that this has on the variance of the partial sums.

Proposition 6.1.1. *Assume that*

$$\sum_{i=0}^{\infty}|R_X(i)| < \infty\,. \tag{6.2}$$

Then

$$\lim_{n\to\infty}\frac{\mathrm{Var}S_n}{n} = R_X(0) + 2\sum_{i=1}^{\infty}R_X(i) \in [0,\infty)\,. \tag{6.3}$$

Proof. The claim follows immediately from (6.1), since the condition (6.2) allows us to use the Lebesgue dominated convergence theorem. □

The summability of covariances condition in (6.2) is the single most commonly used second-order definition of short memory. By Proposition 6.1.1, it leads to an at most linear rate of increase of the variance of the partial sums. Since the limit in (6.3) can happen to be equal to zero (as the example of fractional Gaussian noise with $0 < H < 1/2$ shows), the summability of covariances does not guarantee a linear rate of increase in the partial sum variance.

If the summability of covariances condition in (6.2) is taken as the definition of a short-memory second-order process, then the lack of this property, that is,

$$\sum_{i=0}^{\infty}|R_X(i)| = \infty\,, \tag{6.4}$$

should be taken as the definition of long-range dependence.

On the other hand, one can also define short memory in a second-order process by extending somewhat the conclusion of Proposition 6.1.1. Specifically, a process can be considered to have short memory if

$$\limsup_{n\to\infty}\frac{\mathrm{Var}S_n}{n} < \infty\,, \tag{6.5}$$

that is, if the variance of the partial sums grows at most linearly. If (6.5) is taken as the definition of short memory, then long-range dependence means absence of this property. This is a slightly modified version of the Allen short and long memory notions of Heyde and Yang (1997).

It is important to keep in mind that the Allen short memory (6.5) can hold even when the summability of covariances fails. By (6.1),

$$\frac{\mathrm{Var}S_n}{n} = R_X(0) + 2\frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{j} R_X(i)\,, \tag{6.6}$$

so the condition

$$\sum_{i=1}^{n} R_X(i) \text{ is a bounded function of } n \tag{6.7}$$

implies (6.5), whereas a stronger condition

$$\sum_{i=1}^{n} R_X(i) \to \gamma_X \in (-\infty, \infty) \text{ as } n \to \infty \tag{6.8}$$

implies even that

$$\lim_{n\to\infty} \frac{\mathrm{Var}S_n}{n} = R_X(0) + 2\gamma_X\,.$$

The following example shows that one or both of these conditions can hold when the covariances are not summable.

Example 6.1.2. The covariance function $R_X(i) = (-1)^i$, $i = 0, 1, 2, \ldots$, provides an example of the situation in which (6.7) holds but (6.2) fails.

Further, let R be any nonnegative nonsummable covariance function that is eventually decreasing to zero; the covariance function (5.10) of a fractional Gaussian noise with $0 < H < 1/2$ is one example of such a covariance function. Since the product of two nonnegative definite functions is a nonnegative definite function, $R_X(i) = (-1)^i R(i)$, $i = 0, 1, 2, \ldots$, is a covariance function that satisfies (6.8) but not (6.2).

Finally, additional insight into the Allen short memory condition (6.5) can be gained by noticing that we can have

$$0 < \liminf_{n\to\infty} \frac{\mathrm{Var}S_n}{n} < \limsup_{n\to\infty} \frac{\mathrm{Var}S_n}{n} < \infty\,,$$

starting, for instance, with Example 6.5.3.

The above example shows, among other things, that the lack of summability of covariances (6.4) is not, by itself, informative as far as the rate of growth of the variance of the partial sums is concerned. Much more informative is the following, more specific, assumption:

$$(R_X(n)) \text{ is a regularly varying sequence with exponent } d \in [-1, 0]; \tag{6.9}$$

if $d = -1$, assume, additionally, that (6.4) holds.

Theorem 6.1.3. *Assume that* (6.9) *holds. Then*

$$\lim_{n\to\infty} \frac{\mathrm{Var} S_n}{n} = \infty\,. \tag{6.10}$$

Further,

$$\lim_{n\to\infty} \frac{\mathrm{Var} S_n}{n^2 R_X(n)} = \frac{2}{(1+d)(2+d)}\,. \tag{6.11}$$

Proof. Let

$$h_n = \sum_{i=1}^{n} R_X(i),\ i = 1, 2, \ldots\,. \tag{6.12}$$

It follows by the definition of a regularly varying sequence and Theorem 10.5.6 that the sequence (h_n) is regularly varying with exponent $1 + d$ and

$$\lim_{n\to\infty} \frac{h_n}{nR_X(n)} = \frac{1}{1+d}\,. \tag{6.13}$$

We have

$$\frac{\mathrm{Var} S_n}{n} = R_X(0) + 2\frac{1}{n}\sum_{j=1}^{n-1} h_j\,.$$

Since (6.9) implies that $h_n \to \infty$ as $n \to \infty$, we obtain (6.10). Furthermore, (6.11) follows from (6.13) and Theorem 10.5.6. □

It follows from Theorem 6.1.3 that under the assumption (6.9) of regular variation of covariances, the variances of the partial sums grow at a rate faster than linear, so that (6.5) fails, implying the Allen long memory. In fact, (6.9) also implies that

$$(\mathrm{Var} S_n) \text{ is a regularly varying sequence with exponent } \gamma \in [1, 2]; \tag{6.14}$$

moreover, if $\gamma = 1$, then $\mathrm{Var} S_n/n \to \infty$ as $n \to \infty$.

Example 6.1.4. An example of the situation in which (6.9) holds, and hence (6.14) holds as well, is that of fractional Gaussian noise; see (5.11) and (5.12), and Example 6.2.8 below. On the other hand, (6.14) can hold without regular variation of covariances, as the covariance function $R_X(i) = (-1)^i + R_{\mathrm{FGN}}(i)$, $i = 0, 1, \ldots$, (with R_{FGN} the covariance function of a fractional Gaussian noise) demonstrates.

The message of Example 6.1.4 notwithstanding, we have the following partial converse result.

Proposition 6.1.5. *Suppose that the covariance function* (R_X) *is eventually monotone decreasing. If* (6.14) *holds with* $\gamma \in (1, 2)$*, then* (6.9) *holds as well with* $d = \gamma - 2$*.*

Proof. The eventual monotonicity of the covariance implies, in particular, that the covariance function is eventually positive. Therefore, the sequence (h_n) in (6.12) is eventually increasing. It follows from (6.14) that

$$\lim_{n\to\infty} \frac{\mathrm{Var} S_n}{\sum_{j=1}^n h_j} = 1\,;$$

in particular,

$$\text{the sequence } \Big(\sum_{j=1}^n h_j\Big) \text{ is regularly varying with exponent } \gamma.$$

By the second part of Exercise 10.9.8, we conclude that the sequence (h_j) is regularly varying with exponent $\gamma - 1$. Appealing, once again, to the eventual monotonicity of the covariance and to the second part of Exercise 10.9.8, we obtain the claim of the proposition. □

6.2 Spectral Domain Approaches

In Section 6.1, we saw certain second-order approaches to long-range dependence that concentrated either on the rate of decay of the covariance function of the process or on the rate of increase of the variance of the partial sums of the successive observations of the values of the process. A closely related second-order approach to long-range dependence concentrates on the behavior of the spectral measure of the process near the origin.

Let, therefore, F_X be the spectral measure of a stationary finite-variance process $\mathbf{X} = (X_1, X_2, \ldots)$. Recall from Section 1.2 that F_X is a symmetric measure on $(-\pi, \pi]$ such that

$$R_X(n) = \int_{(-\pi,\pi]} \cos nx\, F_X(dx),\ n = 0, 1, \ldots\,. \tag{6.15}$$

If a spectral density exists, we will denote it by f_X.

The first connection between the memory notions of Section 6.1 and the properties of the spectral measure is contained in the following immediate corollary of Proposition 1.2.3.

Corollary 6.2.1. *Suppose the property* (6.2) *of summability of correlations holds. Then the process has a bounded and continuous spectral density given by*

$$f_X(x) = \frac{1}{2\pi}\left(R_X(0) + 2\sum_{n=1}^{\infty} R_X(n)\cos nx\right), \ -\pi < x < \pi\,. \tag{6.16}$$

That is, the summability of covariances implies existence of a spectral density that is "nice" (bounded and continuous) everywhere. On the other hand, it turns out that it is mostly the behavior of the spectral measure near the origin that affects the rate of increase of the partial sum variance and, to a lesser extent, the rate of decay of the covariances. The following statement is our first result of this type.

Theorem 6.2.2. *Suppose that the spectral measure F_X has, in a neighborhood of the origin, a density f_X such that the limit*

$$f_X^*(0) = \lim_{x\to 0} f_X(x) \in [0,\infty] \ \textit{exists.}$$

Then

$$\lim_{n\to\infty} \frac{\mathrm{Var}S_n}{n} = 2\pi f_X^*(0)\,. \tag{6.17}$$

Proof. Suppose first that the spectral measure F_X puts no mass at the point π. We will use the following standard formulas for the sum of cosine functions and for the sum of sine functions:

$$\sum_{m=j}^{j+k} \cos mx = \frac{\sin\big((j+k+1)x\big) + \sin\big((j+k)x\big) - \sin(jx) - \sin\big((j-1)x\big)}{2\sin x} \tag{6.18}$$

and

$$\sum_{m=j}^{j+k} \sin mx = \frac{\cos(jx) + \cos\big((j-1)x\big) - \cos\big((j+k+1)x\big) - \cos\big((j+k)x\big)}{2\sin x}. \tag{6.19}$$

We begin by applying (6.18) to the inner sum on the right-hand side of (6.6). We have by (6.15), for every $j = 1, 2, \ldots$,

$$\sum_{m=1}^{j} R_X(m) = \sum_{m=1}^{j} \int_{(-\pi,\pi]} \cos mx\, F_X(dx) \tag{6.20}$$

$$= \int_{(-\pi,\pi]} \frac{\sin\big((j+1)x\big) + \sin(jx) - \sin x}{2\sin x} F_X(dx)$$

$$= \frac{1}{2}\int_{(-\pi,\pi]} \frac{\sin\big((j+1)x\big)}{\sin x} F_X(dx) + \frac{1}{2}\int_{(-\pi,\pi]} \frac{\sin(jx)}{\sin x} F_X(dx) - \frac{1}{2}R_X(0)\,.$$

Next, we apply (6.19) to the outer sum on the right-hand side of (6.6), which amounts to applying (6.19) twice to the right-hand side of (6.20). After a simplification, we obtain

$$\frac{\mathrm{Var}S_n}{n} = \frac{1}{n}R_X(0) + \int_{(-\pi,\pi]} \frac{1}{n}A_n(x)\, F_X(dx)\,,$$

where

$$A_n(x) = \frac{1 + 2\cos x + \cos(2x) - \cos\big((n-1)x\big) - 2\cos(nx) - \cos\big((n+1)x\big)}{2\sin^2 x}\,.$$

Let now $\delta \in (0,\pi)$ be such that in the interval $(-\delta,\delta)$, the spectral measure has density f_X as described in the theorem. It is elementary to check that the functions A_n are uniformly bounded on $(-\pi,\pi)\setminus(-\delta,\delta)$, which implies that

$$\lim_{n\to\infty} \int_{(-\pi,\pi)\setminus(-\delta,\delta)} \frac{1}{n}A_n(x)\, F_X(dx) = 0\,,$$

and the claim of the theorem will follow once we prove that

$$\lim_{n\to\infty} \int_{(-\delta,\delta)} \frac{1}{n}A_n(x) f_X(x)\, dx = 2\pi f_X^*(0)\,. \tag{6.21}$$

Let

$$b_n = \frac{1}{2n}\int_{(-\delta,\delta)} \frac{1-\cos(nx)}{\sin^2 x} f_X(x)\, dx,\ n = 1,2,\ldots\,.$$

Then as $n\to\infty$,

$$\begin{aligned} b_n &\sim \frac{1}{2n}\int_{(-\delta,\delta)} \frac{1-\cos(nx)}{x^2} f_X(x)\, dx \\ &= \frac{1}{2}\int_{-n\delta}^{n\delta} \frac{1-\cos(x)}{x^2} f_X(x/n)\, dx \\ &\to \frac{f_X^*(0)}{2}\int_{-\infty}^{\infty} \frac{1-\cos(x)}{x^2}\, dx, \end{aligned}$$

by the dominated convergence theorem. Since the integral obtained in the limit is equal to π by (10.49), and since

$$\int_{(-\delta,\delta)} \frac{1}{n} A_n(x) f_X(x)\, dx = b_{n-1} + 2b_n + b_{n+1} - \frac{1}{n} \int_{(-\delta,\delta)} \frac{2(1-\cos x) + (1-\cos(2x))}{\sin^2 x} f_X(x)\, dx\,,$$

we have proved (6.21) and hence also (6.17), under the assumption that the spectral measure F_X puts no mass at the point π.

Allowing for a point mass $p \geq 0$ of the spectral measure at the point π, write

$$F_X = \tilde{F}_X + p\delta_\pi\,,$$

where $\tilde{F}_X$ puts no mass at the point π. Note that $\tilde{F}_X$ coincides with F_X on $(-\pi, \pi)$, hence has a density in a neighborhood of the origin as described in the theorem. Write

$$\tilde{R}_X(n) = \int_{(-\pi,\pi]} \cos nx\, \tilde{F}_X(dx),\ n = 0, 1, \ldots\,,$$

so that

$$R_X(n) = \tilde{R}_X(n) + p(-1)^n,\ n = 0, 1, \ldots\,.$$

Then

$$\begin{aligned} \frac{\mathrm{Var} S_n}{n} &= R_X(0) + 2\frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{j} R_X(i) \\ &= \tilde{R}_X(0) + 2\frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{j} \tilde{R}_X(i) + p\left(1 + 2\frac{1}{n}\sum_{j=1}^{n-1}\sum_{i=1}^{j}(-1)^i\right) \\ &\to 2\pi f_X^*(0), \end{aligned}$$

since the expression in parentheses is easily seen to vanish in the limit. □

Remark 6.2.3. Note that if under the assumptions of Theorem 6.2.2, the spectral density vanishes at the origin, $f_X^*(0) = 0$, then $\mathrm{Var} S_n/n \to 0$ as $n \to \infty$, so that the variance of the partial sums grows at a slower than linear rate. This is the case, for example, for fractional Gaussian noise with $0 < H < 1/2$; see Example 1.2.4.

Remark 6.2.4. A common spectral domain definition of short memory requires the existence of a spectral density that is continuous at the origin. Indeed, by Theorem 6.2.2, in this case the variance of the partial sums of the process grows

at most linearly, so the Allen short memory condition (6.5) is satisfied. However, the theorem also demonstrates that the spectral measure can have arbitrarily "bad properties" outside a small neighborhood of the origin and still have a variance of the partial sums growing at most linearly if the assumptions of the theorem are satisfied in that small neighborhood of the origin (with $f_X^*(0) < \infty$).

An elementary modification of the proof of Theorem 6.2.2 gives us the following result.

Proposition 6.2.5. *Suppose that the spectral measure F_X has, in a neighborhood of the origin, a density f_X bounded from above by a constant M. Then*

$$\limsup_{n\to\infty} \frac{\mathrm{Var} S_n}{n} \le 2\pi M\,.$$

In particular, this proposition demonstrates that if a spectral density exists in a small neighborhood of the origin, then the Allen short memory condition (6.5) does not require its continuity at the origin. In fact, even the existence of a finite limit on the left-hand side of (6.17) does not require continuity of a version of the spectral density at the origin; see Problem 6.5.4.

What behavior of the spectral measure of the process leads to a violation of the Allen short memory condition, i.e., faster than linear rate of growth of the variance of the partial sums? We first check that such a violation is already guaranteed by "too much mass near the origin" of the spectral measure.

Proposition 6.2.6. *Suppose that*

$$\limsup_{\varepsilon\downarrow 0} \frac{F_X\big([0,\varepsilon]\big)}{\varepsilon} = \infty\,. \tag{6.22}$$

Then

$$\limsup_{n\to\infty} \frac{\mathrm{Var} S_n}{n} = \infty\,.$$

Proof. We may assume that the spectral measure F_X puts no mass at the point π. As in the proof of Theorem 6.2.2, it is enough to show that

$$\limsup_{n\to\infty} b_n = \infty\,,$$

where

$$b_n = \frac{1}{2n}\int_{(-\delta,\delta)} \frac{1-\cos(nx)}{\sin^2 x}\, F_X(dx),\ n = 1, 2, \dots\,,$$

for some $0 < \delta < \pi$. However, using the facts that $\sin x \sim x$ as $x \to 0$ and $1 - \cos x \geq cx^2$ for some $c > 0$ if $|x| \leq 1$, we have for n large enough,

$$b_n \geq \frac{1}{2n} \int_{[-1/n,1/n]} \frac{1 - \cos(nx)}{\sin^2 x} F_X(dx) \geq (c/2) n F_X\big([0, 1/n]\big)\,,$$

and this converges to infinity along some subsequence of the integers by (6.22). □

Obviously, if (6.22) holds with the actual limit replacing the limsup, the conclusion of the proposition will change accordingly to (6.10). Some situations in which this happens are the presence of an atom of the spectral measure at the origin and the equality $f_X^*(0) = \infty$ in Theorem 6.2.2. A fractal-like nature of the support of the spectral measure near the origin can easily lead to the same phenomenon; a simple example can be constructed by taking the Cantor set with its natural Hausdorff measure.

More precise information about the rate of growth of the variance of the partial sums of a process can be obtained under the assumption of existence, in a neighborhood of the origin, of a regularly varying spectral density.

Theorem 6.2.7. *Suppose that the spectral measure F_X has, in a neighborhood of the origin, a density f_X such that*

$$f_X(x) = |x|^{-(1+d)} L\big(1/|x|\big) \tag{6.23}$$

as $x \to 0$, where $-2 < d < 0$, and L is a slowly varying function. Then

$$\operatorname{Var} S_n \sim v_d\, n^{2+d} L(n) \quad \text{as } n \to \infty, \tag{6.24}$$

where

$$v_d = \begin{cases} \frac{4\Gamma(-d)\cos(\pi d/2)}{(1+d)(2+d)} & \text{if } d \neq -1 \\ 2\pi & \text{if } d = -1. \end{cases}$$

Proof. As in the proof of Theorem 6.2.2, let $\delta \in (0, \pi)$ be such that in the interval $(-\delta, \delta)$, the spectral measure has a density f_X as described in the theorem. If we set

$$a_n = \frac{1}{2} \int_{(-\delta,\delta)} \frac{1 - \cos(nx)}{x^2} f_X(x)\, dx, \ n = 1, 2, \ldots,$$

then

$$\operatorname{Var} S_n = (1 + o(1))\big(a_{n-1} + 2a_n + a_{n+1}\big) + O(1)$$

as $n \to \infty$. Therefore, the statement of the theorem will follow once we prove that

$$a_n \sim (v_d/4) n^{2+d} L(n) \quad \text{as } n \to \infty. \tag{6.25}$$

However, by (6.23),

$$a_n = \int_0^{\delta} \frac{1-\cos(nx)}{x^{3+d}} L(1/x)\,dx = n^{2+d}L(n)\int_0^{\delta n} \frac{1-\cos y}{y^{3+d}} \frac{L(n/y)}{L(n)}\,dy\,.$$

Since $L(n/y)/L(n) \to 1$ for every $y > 0$ as $n \to \infty$, and since

$$\int_0^{\infty} \frac{1-\cos y}{y^{3+d}}\,dy = \begin{cases} \frac{\Gamma(-d)\cos(\pi d/2)}{(1+d)(2+d)} & \text{if } d \neq -1, \\ \pi/2 & \text{if } d = -1, \end{cases}$$

by (10.49), we only need to justify the passage to the limit under the integral in this argument. To this end, choose $\varepsilon \in \big(0, \min(-d, 2+d)\big)$. By the Potter bounds of Corollary 10.5.8, for all n large enough we have

$$\frac{L(n/y)}{L(n)} \leq \begin{cases} (1+\varepsilon)y^{-\varepsilon} & \text{if } y < 1 \\ (1-\varepsilon)^{-1}y^{\varepsilon} & \text{if } y \geq 1 \end{cases} := h(y)\,.$$

Since the function

$$\frac{1-\cos y}{y^{3+d}} h(y),\ y > 0\,,$$

is integrable on $(0,\infty)$ by the choice of ε, we can use the Lebesgue dominated convergence theorem to justify the passage to the limit under the integral. □

The cases $-1 < d < 0$ and $d = -1$ and the function L does not remain bounded as its argument approaches infinity describe the situations relevant to our discussion of long-range dependence. Indeed, in these cases the variance of the partial sums grows faster than linearly fast, at least along a subsequence.

Concentrating on the case $-1 < d < 0$, the statements of Theorem 6.1.3 and Theorem 6.2.7 lead to an interesting observation. Consider the following three properties of a stationary finite-variance process $\mathbf{X} = (X_1, X_2, \ldots)$. Let L be a slowly varying function on $(0,\infty)$. The first property is a property of the covariance function of the process:

$$R_X(n) \sim n^d L(n),\ -1 < d < 0,\ n \to \infty\,. \tag{6.26}$$

The second property is a property of the spectral measure of the process: there is, in a neighborhood of the origin, a density f_X such that

$$f_X(x) \sim \frac{1}{2\Gamma(-d)\cos(\pi d/2)} |x|^{-(1+d)} L\big(1/|x|\big),\ -1 < d < 0,\ x \to 0\,. \tag{6.27}$$

The final property is a property of the variance of the partial sums of the process:

$$\operatorname{Var} S_n \sim \frac{2}{(1+d)(2+d)} n^{2+d} L(n),\ -1 < d < 0,\ n \to \infty\,. \tag{6.28}$$

Certain fractional Gaussian noises, for example, possess all three of these properties, as the following example shows.

Example 6.2.8. Let $(B_H(t),\ t \geq 0)$ be the fractional Brownian motion of Example 3.5.1. Recall that the increment process $X_j = B_H(j) - B_H(j-1)$ for $j = 1, 2, \ldots$ is a stationary zero-mean Gaussian process, called the fractional Gaussian noise (FGN). It has the covariance function

$$R_X(n) = \frac{\sigma^2}{2}\Big[(n+1)^{2H} + |n-1|^{2H} - 2n^{2H}\Big],\ n = 0, 1, \ldots,$$

where $\sigma > 0$ is a scale parameter, and $H \in (0, 1)$ the Hurst parameter, which is the parameter of self-similarity of the underlying fractional Brownian motion. Recall also that the FGN has a spectral density given by

$$f_X(x) = C(H)\sigma^2(1 - \cos x) \sum_{j=-\infty}^{\infty} |2\pi j + x|^{-(1+2H)},\ -\pi < x < \pi\,,$$

where

$$C(H) = \frac{H(1-2H)}{\Gamma(2-2H)\cos(\pi H)}\,.$$

Suppose that $1/2 < H < 1$. It follows from (5.11) that the FGN has the property (6.26) with $d = -2(1-H)$ and $L(x) = H(2H-1)\sigma^2$, $x > 0$. Further, it is clear that

$$f_X(x) \sim \frac{C(H)}{2}|x|^{1-2H} \ \text{ as } x \to 0,$$

whence the FGN has the property (6.27) as well. Finally, by the definition of the FGN,

$$\operatorname{Var} S_n = \operatorname{Var}\left(\sum_{j=1}^{n} X_j\right) = \operatorname{Var} B_H(n) = \sigma^2 n^{2H},\ n = 1, 2, \ldots\,.$$

Therefore, the FGN has the property (6.28) too.

All three properties (6.26), (6.27), and (6.28) have been taken as definitions of long-range dependence in a stationary process with a finite variance. In general, it follows from Theorem 6.1.3 that (6.26) implies (6.28), and it follows from Theorem 6.2.7 that (6.27) also implies (6.28).

On the other hand, Example 6.1.4 shows that the property (6.28) of the variance of the partial sums does not imply the property (6.26) of the covariances. Further, Exercise 6.5.5 contains an example of a spectral measure without the property (6.27). Nonetheless, the property (6.28) of the variance of the partial sums still holds.

It is, clear, therefore, that the property (6.28) is the weakest of the three properties. The exact relation between the properties (6.26) and (6.27), however, still needs to be clarified. We begin with examples showing that neither of these two properties implies the other.

We begin with an example of a process with a spectral density that is regularly varying at the origin but without a regularly varying covariance function.

Example 6.2.9. Let $0 < \varepsilon < \pi/2$. Let $-1 < d < 0$, and let g be a positive integrable function on $(0, \varepsilon)$ such that g is regularly at the origin with exponent $-(1 + d)$ of regular variation; one could take as g, for instance, the restriction of the spectral density of the fractional Gaussian noise to the interval $(0, \varepsilon)$. Define

$$f(x) = g(|x|)\mathbf{1}\big(0 < |x| < \varepsilon\big) + g(\pi - |x|)\mathbf{1}\big(\pi - \varepsilon < |x| < \pi\big),\ -\pi < x < \pi\,.$$

Then f is a nonnegative integrable symmetric function on $(-\pi, \pi)$ that satisfies the property (6.27) of regular variation at the origin. Let (R_n) be the covariance function of a process with spectral density equal to f. Then

$$\begin{aligned} R_n = 2\int_0^\pi \cos nx\, f(x)\, dx &= 2\int_0^\varepsilon \cos nx\, g(x)\, dx + 2\int_{\pi-\varepsilon}^\pi \cos nx\, g(\pi - x)\, dx \\ &= 2\Big(1 + (-1)^n\Big)\int_0^\varepsilon \cos nx\, g(x)\, dx\,. \end{aligned}$$

Therefore, the covariance function vanishes at all odd lags; hence it does not have the property (6.26) of regular variation.

In the next example, we demonstrate a covariance function that is regularly varying according to (6.26) and whose spectral density does not satisfy (6.27).

Example 6.2.10. Let $0 < d < 1$, and let g be the spectral density satisfying (6.27) such that the corresponding covariance function satisfies (6.26). One could use the spectral density and the covariance function of the fractional Gaussian noise, for instance. We will construct a nonnegative continuous integrable function g_1 on $(0, \pi)$ such that

$$\limsup_{x \downarrow 0} x^2 g_1(x) > 0 \tag{6.29}$$

and

$$\int_0^\pi \cos\, nx\, g_1(x)\, dx = o\left(\int_0^\pi \cos\, nx\, g(x)\, dx\right) \tag{6.30}$$

as $n \to \infty$. Once this is done, we will define a spectral density by

$$f(x) = g(x) + g_1(|x|),\ -\pi < x < \pi\,.$$

This function does not have the property (6.26), because every function with the latter property must satisfy, by (10.33),

$$f(x) = o\big(x^{-(1+d')}\big) \text{ as } x \downarrow 0$$

for every $d < d' < 1$, but the property (6.29) of g_1 rules this out. On the other hand, by (6.30), we see that

$$\lim_{n\to\infty} \frac{\int_{-\pi}^{\pi} \cos\, nx\, g(x)\, dx}{\int_{-\pi}^{\pi} \cos\, nx f(x)\, dx} = 1\,.$$

Since the sequence in the numerator of the ratio above is regularly varying by construction, the same is true for the sequence in the denominator, and the latter sequence is the sequence of the covariances corresponding to the spectral density f.

It remains to construct a function g_1 satisfying (6.29) and (6.30). Let

$$h(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1/2, \\ 2(1-x) & \text{if } 1/2 \le x \le\le 1. \end{cases}$$

Define

$$g_1(x) = 2^{2j} h\left(\frac{x - 2^{-j}}{2^{-2^j}}\right) \text{ if } 2^{-j} \le x \le 2^{-j} + 2^{-2^j},\ j = 0, 1, 2, \ldots,$$

and $g_1(x) = 0$ otherwise. Note that for $x_j = 2^{-j} + 2^{-2^j}/2$ (the midpoint of the jth interval), we have

$$g_1(x_j) = 2^{2j} \sim x_j^{-2} \text{ as } j \to \infty.$$

Therefore, g_1 satisfies (6.29). Next, notice that for every $0 < a < b < \pi$,

$$\begin{aligned} &\int_a^b \cos\, nx\, h\left(\frac{x-a}{b-a}\right) dx \\ &= \int_a^b \cos\, nx \left(\int_a^x \frac{1}{b-a} h'\left(\frac{y-a}{b-a}\right) dy\right) dx \\ &= \frac{1}{n(b-a)} \int_a^b h'\left(\frac{y-a}{b-a}\right) (\sin nb - \sin ny)\, dy \\ &= \frac{2}{n(b-a)} \int_a^b h'\left(\frac{y-a}{b-a}\right) \sin \frac{n(b-y)}{2} \cos \frac{n(b+y)}{2}\, dy\,. \end{aligned}$$

Using the facts that $|h'(z)| \le 2$ a.e. and $|\sin z| \le \min(1, |z|)$, we see that

$$\left| \int_a^b \cos\, nx\, h\left(\frac{x-a}{b-a}\right) dx \right| \le \frac{4}{n} \min\left(1, \frac{n(b-a)}{4}\right).$$

Therefore,

$$\left|\int_0^\pi \cos\, nx\, g_1(x)\, dx\right| = \left|\sum_{j=0}^{\infty} 2^{2j} \int_{2^{-j}}^{2^{-j}+2^{-2j}} \cos\, nx\, h\left(\frac{x-2^{-j}}{2^{-2j}}\right) dx\right|$$

$$\leq \frac{4}{n}\sum_{j=0}^{\infty} 2^{2j} \min\left(1, \frac{n}{4}\, 2^{-2^j}\right) = \frac{4}{n}\sum_{j\leq \log_2\log_2 n} + \frac{4}{n}\sum_{j>\log_2\log_2 n}.$$

Clearly, for $n > 1$,

$$\frac{4}{n}\sum_{j\leq \log_2\log_2 n} 2^{2j} \min\left(1, \frac{n}{4}\, 2^{-2^j}\right) \leq \frac{4}{n}\sum_{j\leq \log_2\log_2 n} 2^{2j}$$

$$\leq \frac{4}{3} n^{-1} 2^{2\log_2\log_2 n} = \frac{4}{3} n^{-1} \left(\log n\right)^2.$$

Further, for $n \geq 16$,

$$\frac{4}{n}\sum_{j>\log_2\log_2 n} 2^{2j} \min\left(1, \frac{n}{4}\, 2^{-2^j}\right) \leq \frac{4}{n}\sum_{j>\log_2\log_2 n} 2^{2j}\, \frac{n}{4}\, 2^{-2^j}$$

$$= \sum_{j>\log_2\log_2 n} 2^{2j-2^j} \leq 2\cdot 2^{2\log_2\log_2 n - 2^{\log_2\log_2 n}} = 2n^{-1}\left(\log n\right)^2.$$

Since by construction, the sequence of the integrals on the right-hand side of (6.30) is regularly varying with exponent $d \in (-1, 0)$, we see that the function g_1 satisfies (6.30).

Examples 6.2.9 and 6.2.10 notwithstanding, under certain regularity assumptions, conditions (6.26) and (6.27) do, in a sense, imply each other. This is exhibited in the result below. A main assumption in this result is that the slowly varying functions involved belong to the Zygmund class; see Definition 10.5.1. Membership in this class is not preserved under tail equivalence, and therefore, in this theorem it is no longer enough to assume the respective condition (6.26) or(6.27) as an asymptotic equivalence.

Theorem 6.2.11. (a) *Assume that the covariances of a stationary second-order process* $\mathbf{X}$ *satisfy*

$$R_X(n) = n^d L(n),\ n = 1, 2, \ldots, \tag{6.31}$$

$-1 < d < 0$, *and the slowly varying function* L *belongs to the Zygmund class. Then the process* $\mathbf{X}$ *has a spectral density that satisfies* (6.27).

(b) *Assume that a stationary second-order process* $\mathbf{X}$ *has a spectral density such that*

$$f_X(x) = \frac{1}{2\Gamma(-d)\cos(\pi d/2)} |x|^{-(1+d)} L\big(1/|x|\big),\ 0 < x < \pi, \tag{6.32}$$

$-1 < d < 0$, and the slowly varying function L belongs to the Zygmund class. Assume, further, that the spectral density has bounded variation on the interval (ε, π) for every $0 < \varepsilon < \pi$. Then the covariance function of $\mathbf{X}$ satisfies (6.26).

Remark 6.2.12. It will be seen from the proof of the theorem that in fact, only a part of the Zygmund class assumptions is needed for each part of the theorem. For part (a), one needs to assume only that the function $\left(x^d L(x),\ x > 0\right)$ is eventually nonincreasing. For part (b), one needs to assume only that the function $\left(x^{1+d} L(x),\ x > 0\right)$ is eventually nondecreasing.

Before embarking on the proof of the theorem, we establish several preliminary results.

Lemma 6.2.13. *For every nonnegative nonincreasing sequence (a_n) and arbitrary sequence (c_n),*

$$\left| \sum_{n=1}^{k} a_n c_n \right| \le a_1 \max_{n \le k} |s_n|,\ k = 1, 2, \ldots, \tag{6.33}$$

where $s_n = c_1 + \ldots + c_n$, $n = 1, 2, \ldots$. In particular, for every nonnegative nonincreasing sequence (a_n) and $0 < x < \pi$,

$$\left| \sum_{n=m_1}^{m_2} a_n \cos nx \right| \le 2(\sin x)^{-1} a_{m_1} \tag{6.34}$$

and

$$\left| \sum_{n=m_1}^{m_2} a_n \sin nx \right| \le 2(\sin x)^{-1} a_{m_1} \tag{6.35}$$

for every $m_2 > m_1 \ge 1$. Consequently, if $a_n \downarrow 0$, then the sums $\sum_n a_n \cos nx$ and $\sum_n a_n \sin nx$ converge.

Proof. Let $b_k = a_k$, $b_j = a_j - a_{j+1}$, $j = 1, \ldots, k-1$. Since all b_j are nonnegative, we see that

$$\left| \sum_{n=1}^{k} a_n c_n \right| = \left| \sum_{n=1}^{k} c_n \sum_{j=n}^{k} b_j \right| = \left| \sum_{j=1}^{k} b_j s_j \right|$$

$$\le \sum_{j=1}^{k} b_j |s_j| \le \sum_{j=1}^{k} b_j \max_{n \le k} |s_n| = a_1 \max_{n \le k} |s_n|,$$

proving (6.33).

To verify the statements (6.34) and (6.35), one uses (6.33) and the summation formulas (6.18) and (6.19). □

Lemma 6.2.14. *For* $-1 < d < 0$, *let*

$$h(x) = \sum_{n=1}^{\infty} n^d \cos nx, \ 0 < x < \pi .$$

Then

$$h(x) \sim \Gamma(1+d) \sin\bigl(\pi |d|/2\bigr)\, x^{-(d+1)} \ \text{ as } x \downarrow 0. \tag{6.36}$$

Proof. The function h is finite-valued by Lemma 6.2.13. To show (6.36), let us start with the obvious fact that the (main branch of the) function

$$g(z) = (1-z)^{-(d+1)}$$

is analytic in $\{z \in \mathbb{C} : |z| < 1\}$. Specifically, in this domain, the Taylor expansion

$$(1-z)^{-(d+1)} = \sum_{n=0}^{\infty} \frac{(1+d)(2+d)\dots(n+d)}{n!} z^n := \sum_{n=0}^{\infty} c_n(d) z^n$$

holds. In particular, for every $0 < x < \pi$ and $0 < r < 1$,

$$\bigl(1 - re^{ix}\bigr)^{-(d+1)} = \sum_{n=0}^{\infty} c_n(d) r^n e^{inx} . \tag{6.37}$$

We claim that it is legitimate to let $r \to 1$ in (6.37) to obtain

$$\bigl(1 - e^{ix}\bigr)^{-(d+1)} = \sum_{n=0}^{\infty} c_n(d) e^{inx} . \tag{6.38}$$

Since the left-hand side of (6.37) clearly converges to the left-hand side of (6.38), we need to show only that the right-hand side of (6.38) makes sense and is the limit of the right-hand side of (6.37). Since

$$c_{n+1}(d) = \frac{n+1+d}{n+1} c_n(d) < c_n(d)$$

for $n = 0, 1, \dots$, Lemma 6.2.13 applies and shows that the series on the right-hand side of (6.38) converges. Next, by Stirling's formula for the gamma function,

$$c_n(d) = \frac{1}{n!} \left(\frac{\Gamma(2+d)}{\Gamma(1+d)} \frac{\Gamma(3+d)}{\Gamma(2+d)} \cdots \frac{\Gamma(n+1+d)}{\Gamma(n+d)} \right) \tag{6.39}$$

$$
\begin{aligned}
&= \frac{1}{\Gamma(1+d)} \frac{\Gamma(n+1+d)}{n!} \\
&= \frac{1}{\Gamma(1+d)} \frac{e^{-(n+1+d)}(n+1+d)^{n+1+d} n^{-1/2}\big(1+O(1/n)\big)}{e^{-(n+1)}(n+1)^{n+1} n^{-1/2}\big(1+O(1/n)\big)} \\
&= \frac{n^d}{\Gamma(1+d)}\big(1+O(1/n)\big)
\end{aligned}
$$

as $n \to \infty$. In particular, by Lemma 6.2.13,

$$
\begin{aligned}
\left| \sum_{n=N+1}^{\infty} c_n(d) e^{inx} \right| &= \left| \sum_{n=N+1}^{\infty} c_n(d) \cos nx + i \sum_{n=N+1}^{\infty} c_n(d) \sin nx \right| \\
&\le 4(\sin x)^{-1} c_{N+1}(d) r^{N+1} \le 4(\sin x)^{-1} c_{N+1}(d) \to 0
\end{aligned}
$$

as $N \to \infty$ uniformly in $0 < r \le 1$. Since for every fixed finite N,

$$
\sum_{n=0}^{N} c_n(d) r^n e^{inx} \to \sum_{n=0}^{N} c_n(d) e^{inx}
$$

as $r \to 1$, the identity (6.38) follows.

Since for $0 < x < \pi$,

$$
1 - e^{ix} = 2\sin(x/2) e^{i(\pi - x)/2},
$$

we can take the real parts of both sides of the identity (6.38) to obtain

$$
\begin{aligned}
\sum_{n=1}^{\infty} c_n(d) \cos nx &= \operatorname{Re}\Big(\big(2\sin(x/2)\big)^{-(d+1)} e^{i(x-\pi)(d+1)/2} \Big) \\
&= \big(2\sin(x/2)\big)^{-(d+1)} \cos\big((x-\pi)(d+1)/2\big) \\
&\sim x^{-(d+1)} \cos\big(\pi(d+1)/2\big) \\
&= x^{-(d+1)} \sin\big(\pi|d|/2\big)
\end{aligned}
$$

as $x \to 0$. Now the claim (6.36) follows, since by (6.39),

$$
\begin{aligned}
\sum_{n=1}^{\infty} c_n(d) \cos nx &= \sum_{n=1}^{\infty} \frac{n^d}{\Gamma(1+d)}\big(1+O(1/n)\big) \cos nx \\
&= \frac{1}{\Gamma(1+d)} \sum_{n=1}^{\infty} n^d \cos nx + O(1)
\end{aligned}
$$

as $x \to 0$. □

Proof of Theorem 6.2.11. We begin by proving part (a). Define a function f_X on $(-\pi, \pi)$ by (6.16). We need to prove that this function is actually well defined, that it is (a version of) the spectral density of the process $\mathbf{X}$, and finally, that it satisfies (6.27).

We will assume that the covariance function R_X is eventually positive. The case in which it is eventually negative can be treated in a similar way. First of all, since the slowly varying function belongs to the Zygmund class, the covariance function R_X is eventually nonincreasing, and hence by Lemma 6.2.13, the infinite series in (6.16) converges, and hence the function f_X is well defined.

In order to prove that the function f_X is the spectral density of the process $\mathbf{X}$, we need to show that it is integrable over $(-\pi, \pi)$ and further, that for every $m = 0, 1, \ldots,$

$$\int_{-\pi}^{\pi} \cos mx f_X(x)\, dx = R_X(m)\,. \tag{6.40}$$

We will prove the integrability and (6.40) for $m = 0$. The proof for general m is similar and is left as an exercise. By (6.16), we need to show only that the integral

$$\int_{-\pi}^{\pi} \left(\sum_{n=1}^{\infty} R_X(n) \cos nx \right) dx$$

is well defined and is actually equal to zero. Since

$$\int_{-\pi}^{\pi} \cos nx\, dx = 0$$

if $n \geq 1$, it is enough to show that

$$\lim_{N\to\infty} \int_0^{\pi} \left| \sum_{n=N}^{\infty} R_X(n) \cos nx \right| dx = 0\,. \tag{6.41}$$

To this end, we may assume that N is so large that for $n \geq N$, the covariance function is positive and nonincreasing. Write

$$\int_0^{\pi} \left| \sum_{n=N}^{\infty} R_X(n) \cos nx \right| dx$$

$$\leq \int_0^{\pi} \sum_{n=N}^{\infty} R_X(n) \mathbf{1}\big(n \leq 1/\sin x\big)\, dx + \int_0^{\pi} \left| \sum_{n \geq \max(N,\, 1/\sin x)} R_X(n) \cos nx \right| dx\,.$$

By Fubini's theorem,

$$\int_0^\pi \sum_{n=N}^{\infty} R_X(n)\mathbf{1}\big(n \le 1/\sin x\big)\,dx \tag{6.42}$$

$$= 2\sum_{n=N}^{\infty} R_X(n)\arcsin(1/n) \sim 2\sum_{n=N}^{\infty} \frac{1}{n} R_X(n) \to 0$$

as $N \to \infty$ by (6.31). Next, by Lemma 6.2.13 and the choice of N,

$$\int_0^\pi \left| \sum_{n \ge \max(N, 1/\sin x)} R_X(n)\cos nx \right| dx$$

$$\le 2\int_0^\pi \frac{1}{\sin x} \max\left(N, \frac{1}{\sin x}\right)^d L\Big(\max\big(N, 1/\sin x\big)\Big)\,dx\,.$$

Let $0 < \delta < -d$. By (10.33), we can further bound the above expression by

$$\le C\int_0^\pi \frac{1}{\sin x} \max\left(N, \frac{1}{\sin x}\right)^{d+\delta} dx\,,$$

where C is a finite constant. Since the integrand is bounded from above by

$$\big(\sin x\big)^{-1-d+\delta}\,,$$

which is an integrable function, the dominated convergence theorem implies that

$$\int_0^\pi \left| \sum_{n \ge \max(N, 1/\sin x)} R_X(n)\cos nx \right| dx \to 0$$

as $N \to \infty$. Together with (6.42), this implies (6.41).

It remains to prove that the spectral density (6.16) satisfies (6.27). Let $0 < x < \pi$. For a small $0 < \theta < 1$, we write

$$\sum_{n=1}^{\infty} R_X(n)\cos nx = \sum_{n \le \theta/x} + \sum_{\theta/x < n \le \theta^{-1}/x} + \sum_{n > \theta^{-1}/x} := T_1(x) + T_2(x) + T_3(x)\,. \tag{6.43}$$

It follows by Theorem 10.5.6 that as $x \downarrow 0$,

$$|T_1(x)| \le \sum_{n \le \theta/x} n^d |L(n)| \sim (1+d)^{-1}\big(\theta/x\big)\big(\theta/x\big)^d L\big(\theta/x\big)$$

$$\sim \theta^{1+d}(1+d)^{-1}x^{-(1+d)}L(1/x)$$

by the slow variation of L. Therefore,

$$\lim_{\theta\to 0}\limsup_{x\downarrow 0}\frac{T_1(x)}{x^{-(1+d)}L(1/x)} = 0\,. \tag{6.44}$$

Next, for $x > 0$ so small that $n^d L(n)$ is nonincreasing on $n > \theta^{-1}/x$, we have by Lemma 6.2.13,

$$\begin{aligned}|T_3(x)| &= \left|\sum_{n>\theta^{-1}/x} n^d L(n)\cos nx\right| \\ &\le 2\big(\lfloor\theta^{-1}/x\rfloor + 1\big)^d L\big(\lfloor\theta^{-1}/x\rfloor + 1\big)(\sin x)^{-1} \\ &\sim 2\theta^{-d}x^{-(1+d)}L(1/x)\end{aligned}$$

as $x \downarrow 0$. Therefore,

$$\lim_{\theta\to 0}\limsup_{x\downarrow 0}\frac{T_3(x)}{x^{-(1+d)}L(1/x)} = 0\,. \tag{6.45}$$

Further, write

$$\begin{aligned}T_2(x) &= L(1/x)\sum_{\theta/x<n\le\theta^{-1}/x} n^d\cos nx + \sum_{\theta/x<n\le\theta^{-1}/x} n^d\big(L(n) - L(1/x)\big)\cos nx \\ &:= T_{2,1}(x) + T_{2,2}(x)\,.\end{aligned}$$

It follows from Lemma 6.2.14 and (6.44) and (6.45) with $L \equiv 1$ that for every $0 < \varepsilon < 1$, there is $\theta_\varepsilon \in (0,1)$ such that for all $0 < \theta \le \theta_\varepsilon$,

$$\begin{aligned}(1-\varepsilon)\Gamma(1+d)\sin\big(\pi|d|/2\big) &\le \liminf_{x\downarrow 0}\frac{T_{21}(x)}{x^{-(1+d)}L(1/x)} \\ &\le \limsup_{x\downarrow 0}\frac{T_{21}(x)}{x^{-(1+d)}L(1/x)} \le (1+\varepsilon)\Gamma(1+d)\sin\big(\pi|d|/2\big)\,.\end{aligned} \tag{6.46}$$

Note also that

$$\begin{aligned}|T_{2,2}(x)| &\le \max_{\theta/x<n\le\theta^{-1}/x}\big|L(n) - L(1/x)\big|\sum_{\theta/x<n\le\theta^{-1}/x} n^d \\ &\le L(1/x)\max_{\theta/x<n\le\theta^{-1}/x}\left|\frac{L(n)}{L(1/x)} - 1\right|\left(\frac{\theta^{-1}}{x}\right)^{1+d}\,.\end{aligned}$$

Therefore, by Proposition 10.5.5, for each $0 < \theta < 1$,

$$\lim_{x \downarrow 0} \frac{T_{2,2}(x)}{x^{-(1+d)} L(1/x)} = 0 \,. \tag{6.47}$$

Putting all the pieces together, we conclude by (6.44), (6.45), (6.46), and (6.47) that

$$\lim_{x \downarrow 0} \big(x^{-(1+d)} L(1/x) \big)^{-1} \sum_{n=1}^{\infty} R_X(n) = \Gamma(1+d) \sin\big(\pi |d|/2\big) \,.$$

The property (6.27) now follows using (6.16) and the identity (10.48) of the gamma function.

We now prove part (b) of the theorem. By (6.15), we need to prove that

$$\int_0^{\pi} x^{-(1+d)} L(1/x) \cos nx \, dx \sim \Gamma(-d) \cos(\pi d/2) n^d L(n) \quad \text{as } n \to \infty. \tag{6.48}$$

Note, first, that in the case $L \equiv 1$, we already know that(6.48) holds, since by (10.47) and (10.48),

$$\begin{aligned} \int_0^{\pi} x^{-(1+d)} \cos nx \, dx &= n^d \int_0^{n\pi} x^{-(1+d)} \cos x \, dx \\ &\sim n^d \int_0^{\infty} x^{-(1+d)} \cos x \, dx = n^d \, \Gamma(-d) \cos(\pi d/2) \,. \end{aligned}$$

In the general case, we proceed similarly to the steps taken in the proof of part (a). For a small $0 < \theta < 1$, we write

$$\int_0^{\pi} x^{-(1+d)} L(1/x) \cos nx \, dx = \int_0^{\theta/n} + \int_{\theta/n}^{\theta^{-1}/n} + \int_{\theta^{-1}/n}^{\pi} := I_1(n) + I_2(n) + I_3(n) \,.$$

By Theorem 10.5.9, as $n \to \infty$,

$$\begin{aligned} |I_1(n)| &\le \int_0^{\theta/n} x^{-(1+d)} L(1/x) \, dx \\ &\sim |d|^{-1} (\theta/n)(\theta/n)^{-(1+d)} L(n/\theta) \sim \theta^{-d} |d|^{-1} n^d L(n) \end{aligned}$$

by the slow variation of L. Therefore,

$$\lim_{\theta \to 0} \limsup_{n \to \infty} \frac{I_1(n)}{n^d L(n)} = 0 \,. \tag{6.49}$$

Next, let $\gamma > 0$ be so small that the function $x^{1+d}L(x)$ is nondecreasing on $(1/\gamma, \infty)$. Since the spectral density is assumed to be of bounded variation on (γ, π), there is a signed measure m_f on (γ, π) such that

$$x^{-(1+d)}L(1/x) = \gamma^{-(1+d)}L(1/\gamma) + m_f\big((\gamma, x]\big) \text{ a.e. on } (\gamma, \pi).$$

If we write

$$I_3(n) = \int_{\theta^{-1}/n}^{\gamma} + \int_{\gamma}^{\pi} := I_{3,1}(n) + I_{3,2}(n)\,,$$

then by Fubini's theorem,

$$\begin{aligned} I_{3,2}(n) &= \int_{\gamma}^{\pi} \Big(\gamma^{-(1+d)}L(1/\gamma) + m_f\big((\gamma, x]\big)\Big) \cos nx\, dx \\ &= -\frac{1}{n}\gamma^{-(1+d)}L(1/\gamma)\sin n\gamma - \frac{1}{n}\int_{\gamma}^{\pi} \sin ny\, m_f(dy)\,, \end{aligned}$$

so that

$$|I_{3,2}(n)| \le \frac{1}{n}\Big(\gamma^{-(1+d)}|L(1/\gamma)| + \|m_f\|\Big)\,.$$

Therefore, for every $0 < \theta < 1$,

$$\lim_{n\to\infty} \frac{I_{3,2}(n)}{n^d L(n)} = 0\,. \tag{6.50}$$

We treat the term $I_{3,1}(n)$ in a similar manner. By the choice of γ, the function $x^{-(1+d)}L(1/x)$ is nonincreasing on $(0, \gamma)$, so there is a positive measure $\tilde{m}_f$ on $(0, \gamma)$, finite away from the origin, such that

$$x^{-(1+d)}L(1/x) = \gamma^{-(1+d)}L(1/\gamma) + \tilde{m}_f\big((x, \gamma)\big) \text{ a.e. on } (0, \gamma).$$

Once again, by Fubini's theorem,

$$\begin{aligned} I_{3,1}(n) &= \int_{\theta^{-1}/n}^{\gamma} \Big(\gamma^{-(1+d)}L(1/\gamma) + \tilde{m}_f\big((x, \gamma)\big)\Big) \cos nx\, dx \\ &\le \gamma^{-(1+d)}L(1/\gamma)\left|\int_{\theta^{-1}/n}^{\gamma} \cos nx\, dx\right| + \left|\int_{(\theta^{-1}/n, \gamma)} \left(\int_{\theta^{-1}/n}^{z} \cos nx\, dx\right) \tilde{m}_f(dz)\right| \\ &\le \left(\gamma^{-(1+d)}L(1/\gamma) + \int_{(\theta^{-1}/n, \gamma)} \tilde{m}_f(dz)\right) \sup_{z > \theta^{-1}/n} \left|\int_{\theta^{-1}/n}^{z} \cos nx\, dx\right| \\ &\le \big(\theta^{-1}/n\big)^{-(1+d)}L(\theta n)\big(2/n\big) = 2\theta^{1+d}n^d L(\theta n)\,. \end{aligned}$$

Using again the slow variation of L, we conclude that

$$\lim_{\theta\to 0}\limsup_{n\to\infty}\frac{I_{3,1}(n)}{n^d L(n)} = 0\,. \tag{6.51}$$

Finally, we treat the term $I_2(n)$, and we do it in the same way we treated the term $T_2(x)$ in the proof of part (a). Write

$$\begin{aligned} I_2(n) &= L(n)\int_{\theta/n}^{\theta^{-1}/n} x^{-(1+d)}\cos nx\,dx + \int_{\theta/n}^{\theta^{-1}/n} x^{-(1+d)}\big(L(1/x)-L(n)\big)\cos nx\,dx \\ &= I_{2,1}(n) + I_{2,2}(n)\,. \end{aligned}$$

Since (6.48) holds in the case $L \equiv 1$, it follows from (6.49), (6.50), and (6.51) that for every $0 < \varepsilon < 1$, there is $\theta_\varepsilon \in (0,1)$ such that for all $0 < \theta \le \theta_\varepsilon$,

$$(1-\varepsilon)\Gamma(-d)\cos(\pi d/2) \le \liminf_{n\to\infty}\frac{I_{2,1}(n)}{n^d L(n)} \tag{6.52}$$

$$\le \limsup_{n\to\infty}\frac{I_{2,1}(n)}{n^d L(n)} \le (1+\varepsilon)\Gamma(-d)\cos(\pi d/2),$$

while as in the proof of part (a), using Proposition 10.5.5, we see that for each $0 < \theta < 1$,

$$\lim_{\theta\to 0}\limsup_{n\to\infty}\frac{I_{2,2}(n)}{n^d L(n)} = 0\,. \tag{6.53}$$

Now the statement (6.48) follows from (6.49), (6.50), (6.51), (6.52), and (6.53). □

6.3 Pointwise Transformations of Gaussian Processes

We have already discussed in Section 5.3 that it is useful to understand the effect of pointwise transformations on a stationary process that has, from some point of view, long memory. Does the transformed process have, from the same point of view, long memory or not? This question turns out to be particularly interesting when we work with the second-order notions of long memory and the initial long-range dependent process is centered Gaussian.

In this section, we will work with a stationary zero-mean Gaussian process $\mathbf{X}$ and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function. Then $Y_n = g(X_n)$, $n = 1, 2, \ldots$, is again a stationary process. Assuming that the function g is such that $EY_n^2 < \infty$, what is the effect of the transformation on the covariance function and on the spectral measure of the process?

The key tool in answering this question is *Hermite polynomials*. For $n \geq 0$, define a function of a real variable x by

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2},\ x \in \mathbb{R}\,. \tag{6.54}$$

It is elementary to check that $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$. Furthermore, it is not hard to check by induction the following explicit representation:

$$H_n(x) = \sum_{m=0}^{[n/2]} \frac{n!}{m!(n-2m)!} (-2)^{-m} x^{n-2m},\ x \in \mathbb{R}\,; \tag{6.55}$$

see Exercise 6.5.7. Therefore, each H_n is a polynomial of degree n. It is called the *nth Hermite polynomial*.

The following proposition describes the first- and second-order properties of Hermite polynomials evaluated at a normalized and centered Gaussian vector.

Proposition 6.3.1. *(i) Let X be a standard normal random variable. Then for every $n \geq 1$, $E\big(H_n(X)\big) = 0$.*

(ii) Let (X, Y) be a zero-mean unit-variance Gaussian vector with $\mathrm{Corr}(X, Y) = \rho$. Then for $n, m \geq 0$,

$$E\big(H_n(X) H_m(Y)\big) = \begin{cases} 0 & \text{if } n \neq m, \\ n!\, \rho^n & \text{if } n = m. \end{cases} \tag{6.56}$$

Proof. Note that for $n \geq 1$,

$$\begin{aligned} E\big(H_n(X)\big) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} H_n(x)\, dx \\ &= (-1)^n \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d^n}{dx^n} e^{-x^2/2}\, dx \\ &= (-1)^n \frac{1}{\sqrt{2\pi}} \left. \frac{d^{n-1}}{dx^{n-1}} e^{-x^2/2} \right|_{-\infty}^{\infty} = 0\,, \end{aligned}$$

proving part (i). Next, consider the function

$$\varphi(s, x) = e^{sx - s^2/2},\ s, x \in \mathbb{R}\,. \tag{6.57}$$

Equivalently,

$$\varphi(s, x) = e^{-(s-x)^2/2} e^{x^2/2},\ s, x \in \mathbb{R}\,. \tag{6.58}$$

By (6.58),

$$\begin{aligned}\frac{\partial^n}{\partial s^n}\varphi(s,x) &= e^{x^2/2}\frac{\partial^n}{\partial y^n}e^{-y^2/2}\Big|_{y=s-x}\\ &= e^{x^2/2}(-1)^n e^{-(s-x)^2/2}H_n(s-x)\,.\end{aligned}$$

In particular,

$$\frac{\partial^n}{\partial s^n}\varphi(s,x)\bigg|_{s=0} = (-1)^n H_n(-x)\,. \tag{6.59}$$

Furthermore, let (X,Y) be the Gaussian vector in the proposition. By (6.57), for every real s,t,

$$\begin{aligned}E\big(\varphi(s,X)\varphi(t,Y)\big) &= E\Big(\exp\{sX - s^2/2\}\exp\{tY - t^2/2\}\Big) \\ &= \exp\{-(s^2+t^2)/2\}E\Big(\exp\{sX+tY\}\Big)\\ &= \exp\{-(s^2+t^2)/2\}\exp\{(s^2+2st\rho+t^2)/2\}\\ &= e^{st\rho}\,,\end{aligned} \tag{6.60}$$

where we used the fact that $sX+tY$ has the zero-mean normal distribution with variance $s^2+2st\rho+t^2$. Since an easy calculus exercise shows that for $n,m\ge 0$,

$$\frac{\partial^{n+m}}{\partial s^n\partial t^m}e^{st\rho} = e^{st\rho}\sum_{k=0}^{n\wedge m} k!\binom{n}{k}\binom{m}{k}\rho^{n+m-k}s^{m-k}t^{n-k}\,,$$

it follows that

$$\frac{\partial^{n+m}}{\partial s^n\partial t^m}e^{st\rho}\bigg|_{s=t=0} = \begin{cases}0 & \text{if } n\ne m,\\ n!\,\rho^n & \text{if } n=m.\end{cases}$$

On the other hand, by (6.60) and (6.59),

$$\begin{aligned}\frac{\partial^{n+m}}{\partial s^n\partial t^m}e^{st\rho}\bigg|_{s=t=0} &= \frac{\partial^{n+m}}{\partial s^n\partial t^m}E\big(\varphi(s,X)\varphi(t,Y)\big)\bigg|_{s=t=0}\\ &= E\left(\frac{\partial^{n+m}}{\partial s^n\partial t^m}\varphi(s,X)\varphi(t,Y)\bigg|_{s=t=0}\right)\\ &= E\left(\frac{\partial^n}{\partial s^n}\varphi(s,X)\bigg|_{s=0}\frac{\partial^m}{\partial t^m}\varphi(t,Y)\bigg|_{t=0}\right)\end{aligned}$$

$$= E\Big((-1)^n H_n(-X)\,(-1)^m H_m(-Y)\Big)$$
$$= (-1)^{n+m} E\big(H_n(X)H_m(Y)\big)\,,$$

since the good integrability properties of a normal random variable allow taking the derivatives inside the expectation. Note further that on the last step, we used the fact that $(-X,-Y) \stackrel{d}{=} (X,Y)$.

The two resulting expressions for the mixed partial derivative prove the claim of part (ii) of the proposition. □

An immediate conclusion from the proposition is the fact that for a standard normal random variable X,

$$\mathrm{Var}\big(H_n(X)\big) = n!,\ n = 1, 2, \ldots\,. \tag{6.61}$$

Proposition 6.3.1 has a very important consequence for square integrable functions of a standard normal random variable.

Proposition 6.3.2. *Let μ_G be the law of the standard normal random variable on $\mathbb{R}$. The Hermite polynomials $(H_n,\ n = 0, 1, \ldots)$ form an orthogonal basis in $L^2\big(\mathbb{R}, \mu_G\big)$.*

Proof. Since orthogonality of the Hermite polynomials follows from Proposition 6.3.1, it remains to show that every function $f \in L^2\big(\mathbb{R}, \mu_G\big)$ such that

$$\int_{-\infty}^{\infty} f(x)H_n(x)\varphi(x)\,dx = 0 \ \text{ for every } n = 0, 1, \ldots, \tag{6.62}$$

where φ is the standard normal density, must be the null function.

Indeed, since for every $n = 0, 1, \ldots$, the function of the real variable x^n can be written as a finite linear combination of $H_0, \ldots, H_n$, every function f satisfying (6.62) must also satisfy

$$\int_{-\infty}^{\infty} x^n f(x)\varphi(x)\,dx = 0 \ \text{ for every } n = 0, 1, \ldots. \tag{6.63}$$

We claim that f is also orthogonal to the complex exponentials: for every $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} e^{itx} f(x)\varphi(x)\,dx = 0\,. \tag{6.64}$$

To see this, we use the Taylor expansion of the complex exponential. Because each power function is, by (6.63), orthogonal to f, it is enough to check that

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} f(x) \sum_{n=N}^{\infty} \frac{i^n t^n x^n}{n!}\varphi(x)\,dx = 0\,.$$

Since the integrand converges pointwise to zero as $N \to \infty$, it is enough to exhibit an integrable dominating function. Such a function is provided by the estimate

$$\left| f(x) \sum_{n=N}^{\infty} \frac{i^n t^n x^n}{n!} \right| \le |f(x)| \sum_{n=N}^{\infty} \frac{|t|^n |x|^n}{n!}$$
$$\le |f(x)| \sum_{n=0}^{\infty} \frac{|t|^n |x|^n}{n!} = |f(x)| e^{|tx|} .$$

The just established equation (6.64) says that the signed measure with density $f(x)\varphi(x)$, $x \in \mathbb{R}$, with respect to the Lebesgue measure has vanishing Fourier transform. Therefore, the measure is the zero measure. Since the Gaussian density never vanishes, the function f is the null function. □

The following corollary forms the basis of our subsequent analysis of square integrable functions of centered and normalized stationary Gaussian processes.

Corollary 6.3.3. *Let X be a standard normal random variable. For every function $g \in L^2\big(\mathbb{R}, \mu_G\big)$, there is a unique expansion in L^2,*

$$g(X) = \sum_{n=0}^{\infty} \frac{a_n(g)}{n!} H_n(X) . \tag{6.65}$$

The coefficients in the expansion are given by

$$a_n(g) = E\big(H_n(X) g(X)\big), \; n = 0, 1, \ldots . \tag{6.66}$$

The expansion (6.65) is called the *Hermite expansion* of $g(X)$. Suppose that (X, Y) is a zero-mean unit-variance Gaussian vector with $\mathrm{Corr}(X, Y) = \rho$, and g, h are two functions in $L^2\big(\mathbb{R}, \mu_G\big)$. Then $g(X)$ and $h(Y)$ are random variables with finite variance, each of which has its Hermite expansion (6.65) with corresponding sequences of coefficients given by (6.66). By Proposition 6.3.1, we conclude that

$$E\big(g(X)h(Y)\big) = \sum_{n=0}^{\infty} \frac{a_n(g) a_n(h)}{n!} \rho^n \tag{6.67}$$

and

$$\mathrm{Cov}\big(g(X), h(Y)\big) = \sum_{n=1}^{\infty} \frac{a_n(g) a_n(h)}{n!} \rho^n . \tag{6.68}$$

Now we are in a position to describe the effect of a pointwise transformation on the second-order characteristics of a centered and normalized stationary Gaussian process.

Theorem 6.3.4. *Let* $\mathbf{X}$ *be a stationary centered Gaussian process with unit variance, and let* $g \in L^2(\mathbb{R}, \mu_G)$. *Then* $Y_n = g(X_n)$, $n = 1, 2, \ldots$, *is a stationary process with finite variance.*

If R_X *is the covariance function of the process* $\mathbf{X}$, *then the covariance function of the process* $\mathbf{Y}$ *is given by*

$$R_Y(n) = \sum_{k=1}^{\infty} \frac{a_k(g)^2}{k!} R_X(n)^k \,. \tag{6.69}$$

If F_X *is the spectral measure of the process* $\mathbf{X}$, *then the spectral measure of the process* $\mathbf{Y}$ *is given by*

$$F_Y = \sum_{k=1}^{\infty} \frac{a_k(g)^2}{k!} F_X^{*k,f} \,, \tag{6.70}$$

where $F_X^{*k,f}$ *is the folded kth convolution power of* F_X, $k = 1, 2, \ldots$, *defined by*

$$F_X^{*k,f}(A) = F_X \times \ldots \times F_X\Big(\big\{(x_1, \ldots, x_k) \in (-\pi, \pi]^k : x_1 + \ldots + x_k \in A \ mod\ 2\pi\big\}\Big) \,,$$

A a Borel subset of $(-\pi, \pi]$.

Proof. The only part of the theorem that has not yet been proved is the expression (6.70) for the spectral measure of the process $\mathbf{Y}$. However, the latter statement follows immediately from (6.69), since for every $n = 0, 1, \ldots$ and $k = 1, 2, \ldots$,

$$\int_{(-\pi,\pi]} e^{inx} F_X^{*k}(dx) = \left(\int_{(-\pi,\pi]} e^{inx} F_X(dx)\right)^k = R_X(n)^k \,.$$

□

Given a function $g \in L^2(\mathbb{R}, \mu_G)$, let

$$k_g = \inf\big\{k \geq 1 : a_k(g) \neq 0\big\} \,. \tag{6.71}$$

The number k_g is called the *Hermite index* or *Hermite rank* of the function g. For example, the Hermite index of the nth Hermite polynomial H_n is equal to n for $n \geq 1$ and to infinity for $n = 0$. We can restate the conclusions of Theorem 6.3.4 as

$$R_Y(n) = \sum_{k=k_g}^{\infty} \frac{a_k(g)^2}{k!} R_X(n)^k \tag{6.72}$$

and

$$F_Y = \sum_{k=k_g}^{\infty} \frac{a_k(g)^2}{k!} F_X^{*k,f} \,. \tag{6.73}$$

The importance of the notion of the Hermite index becomes clear in the following result.

Theorem 6.3.5. *Under the assumptions of Theorem 6.3.4, suppose additionally that* $k_g < \infty$ *and* $R_X(n) \to 0$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} \frac{R_Y(n)}{\big(R_X(n)\big)^{k_g}} = \frac{a_{k_g}(g)^2}{k_g!}\,.$$

Proof. The claim is immediate from (6.72) and the L^2 convergence of the Hermite expansion (6.65). □

In particular, if the Hermite index of the function is equal to 1, then the covariance function of the process $\big(Y_n = g(X_n),\, n = 1, 2, \ldots\big)$, decays to zero at the same rate as the covariance function of the original process $\mathbf{X}$, while if $k_g > 1$, then the covariance function R_Y converges to zero, since the lag goes to infinity faster than the covariance function R_X does.

Suppose now that the covariance function $(R_X(n))$ satisfies the assumption (6.9), i.e., it is regularly varying with exponent $d \in [-1, 0]$, and the covariances are nonsummable. Then it follows from Theorem 6.3.5 that the covariance function $(R_Y(n))$ is regularly varying with exponent $k_g d$. If $k_g = 1$, then this covariance function still satisfies the assumption (6.9). On the other hand, if $k_g > 1$, then the covariance function $(R_Y(n))$ may or may not satisfy the assumption (6.9), and the covariances may or may not be nonsummable. The crucial factor is whether $k_g d \geq -1$.

From a certain point of view, the situation may appear to be reasonable, since we expect the transformed process $\big(Y_n = g(X_n),\, n = 1, 2, \ldots\big)$ to "remember as much as the process $\mathbf{X}$ does or less." If one measures the memory in terms of the rate of decay of the covariance function, then the above analysis showing that the covariance function R_Y converges to zero at the same rate as the covariance function R_X or faster seems to carry the same message.

If a function g is, however, a one-to-one function, then arguably, the process $\big(Y_n = g(X_n),\, n = 1, 2, \ldots\big)$ "remembers exactly as much as the process $\mathbf{X}$ does." It is an uncomfortable fact that there exist one-to-one functions $g \in L^2\big(\mathbb{R}, \mu_G\big)$ with Hermite index larger than 1. This is demonstrated by the following example.

Example 6.3.6. There exists $a > 0$ such that

$$ae^{-a} \int_a^\infty x e^x e^{-x^2/2}\, dx = \int_0^a x^2 e^{-x^2/2}\, dx\,. \tag{6.74}$$

To see this, note that the expressions on the left-hand side and the right-hand side of (6.74) are respectively $O(a)$ and $O(a^3)$ as $a \downarrow 0$, so for small a, the expression on the left-hand side is larger. On the other hand, the expression on the left-hand side vanishes in the limit as $a \to \infty$, while the expression on the right-hand side

converges to a positive limit. Thus for large a, the right-hand side is larger. Since both expressions are continuous functions of a, there is $a > 0$ satisfying (6.74).

Choose such an $a > 0$ and define

$$g(x) = \begin{cases} -\frac{1}{a}x \text{ if } 0 \le x < a \\ e^{x-a} \text{ if } x \ge a \end{cases} .$$

Set $g(x) = -g(-x)$ for $x < 0$.

Notice that the function g is one-to-one and measurable. Furthermore, the inverse function g^{-1} is measurable as well. Clearly, $g \in L^2\big(\mathbb{R}, \mu_G\big)$. Since the function g is odd, it is immediate that

$$a_k(g) = E\big(H_k(X)g(X)\big) = 0 \text{ for all even } k,$$

with X a standard normal random variable. Furthermore, since the number a satisfies (6.74),

$$a_1(g) = E\big(H_1(X)g(X)\big) = E\big(Xg(X)\big) = 0\,.$$

Therefore, the Hermite rank of the function g is at least 3.

If $\mathbf{X}$ is a stationary centered Gaussian process with unit variance whose covariance function $(R_X(n))$ satisfies the assumption (6.9) and whose exponent of regular variation satisfies $d \in [-1, -1/3)$, and g is the function in Example 6.3.6, then by Theorem 6.3.5, the covariance function $(R_Y(n))$ of the process $\big(Y_n = g(X_n),\, n = 1, 2, \ldots\big)$ is regularly varying with exponent $k_g d < -1$. That is, the original process $\mathbf{X}$ has long-range dependence according to its rate of decay of covariances and nonsummability of its covariances, while the transformed process $\mathbf{Y}$ has short memory according to either of the two criteria.

The phenomenon discussed above is also visible in the spectral domain. The following theorem is a version of Theorem 6.3.5 for spectral densities.

Theorem 6.3.7. *Under the assumptions of Theorem 6.3.4, suppose additionally that $k_g < \infty$ and that the process $\mathbf{X}$ has a spectral density f_X satisfying (6.27) with some $d \in (-1, 0)$ and bounded outside of every neighborhood of the origin. Then the process $\mathbf{Y}$ has a spectral density f_Y. This density can be chosen so that the following properties hold:*

(i) If $k_g d > -1$, then

$$\lim_{x \downarrow 0} \frac{f_Y(x)}{x^{k_g - 1} f_X(x)^{k_g}} = \frac{a_{k_g}(g)^2}{k_g!} c(k_g, d)\,, \tag{6.75}$$

where the constant $c(k, d)$ is given by the integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |z_1|^{-(1+d)} \ldots |z_{k-1}|^{-(1+d)} \big|1 - (z_1 + \ldots + z_{k-1})\big|^{-(1+d)}\, dz_1 \ldots dz_{k-1} \tag{6.76}$$

and is finite if $kd > -1$.

(ii) If $k_g d < -1$, then the density f_Y is bounded.

Proof. It follows from (6.73) that under the assumptions of the theorem, the spectral measure F_Y has density given by

$$f_Y(x) = \sum_{k=k_g}^{\infty} \frac{a_k(g)^2}{k!} f_X^{*k,f}(x)\,, \tag{6.77}$$

where

$$f_X^{*k,f}(x) = \sum_{n\in\mathbb{Z}:\, x+2\pi n\in(-\pi k,\pi k)} f_X^{*k}(x+2\pi n),\ -\pi < x < \pi\,,$$

and f_X^{*k} is the usual kth convolution power of f_X. We will prove parts (i) and (ii) of the theorem for the version of the spectral density given in (6.77).

As a first observation, it is easy to see that the constant in (6.76) is indeed finite if $kd > -1$, simply by computing the iterated integrals one by one. Next, we proceed with two claims. First of all,

$$\lim_{x\downarrow 0} \frac{f_X^{*k,f}(x)}{x^{k-1} f_X(x)^k} = c(k,d) < \infty \ \text{ if } kd > -1. \tag{6.78}$$

Further,

$$f_X^{*k,f} \ \text{ is bounded if } kd < -1. \tag{6.79}$$

Indeed, we may assume that (6.27) holds as an exact equality on $(0,\pi)$ and not only as an asymptotic equivalence. Extend the slowly varying function L to the entire half-line $(0,\infty)$ by setting it equal to zero outside of the interval $(1/\pi,\infty)$, so that we may regard (6.27) as equality on $\mathbb{R}$. Note that for $0 < x < \pi$,

$$f_X^{*k}(x) = \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} f_X(y_1)\dots f_X(y_{k-1}) f_X\big(x - y_1 - \dots - y_{k-1}\big)\, dy_1 \dots dy_{k-1} \tag{6.80}$$

$$= f_X(x)^k x^{k-1}$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{f_X(xz_1)}{f_X(x)} \dots \frac{f_X(xz_{k-1})}{f_X(x)} \frac{f_X\big(x(1 - z_1 - \dots - z_{k-1})\big)}{f_X(x)}\, dz_1 \dots dz_{k-1}\,.$$

Suppose that $kd > -1$. As $x \downarrow 0$, the integrand on the right-hand side converges, by the regular variation, to the integrand in (6.76). Furthermore, by the Potter bounds of Corollary 10.5.8 and boundedness of f_X outside of every neighborhood of the origin, for every $0 < \varepsilon < 1$ and all positive x small enough, we have

$$\frac{f_X(xz)}{f_X(x)} \le C\big(|z|^{-(1+d+\varepsilon)} + |z|^{-(1+d-\varepsilon)}\big)$$

for some finite C and all $z \in \mathbb{R}$. Since we can choose ε so small that $k(d-\varepsilon) > -1$ and $d+\varepsilon < 0$, the finiteness of the constant in (6.76) gives us a dominating function, and we conclude that

$$\lim_{x \downarrow 0} \frac{f_X^{*k}(x)}{x^{k-1} f_X(x)^k} = c(k,d)\,.$$

In order to verify that this implies (6.78), we need to check only that f_X^{*k} is bounded outside of every neighborhood of the origin. To see this, suppose that $|x| > \varepsilon$ for some $\varepsilon > 0$. Then at least one of the arguments of the k functions f_X appearing in the definition of $f_X^{*k}(x)$ in (6.80) has absolute value larger than ε/k. Therefore, the value of f_X at that point is bounded from above by some finite number, say a, depending only on ε and k. Since the integral in (6.80) over that part of the domain where a particular f_X term is bounded by a is itself bounded by a (since f_X is a probability density), we conclude that $f_X^{*k,f}(x) \le ka$, which proves the required boundedness outside of every neighborhood of the origin.

Next we check (6.79), so suppose that $kd < -1$. Choose $\varepsilon > 0$ so small that $k(d+\varepsilon) < -1$ and $d+\varepsilon < 0$. Using once again the Potter bounds and boundedness of f_X outside of every neighborhood of the origin, we see that there is a finite constant C such that $f_X(x) \le C|x|^{-(1+d+\varepsilon)}$ for $|x| < \pi$. Therefore, we can also bound f_X by a constant times the density of the symmetric gamma random variable with the shape $-(d+\varepsilon)$ and scale 1, i.e., of $W_1 - W_2$, where W_1 and W_2 and independent gamma random variables with the shape $-(d+\varepsilon)$ and scale 1. It is an elementary probability exercise to check that if $k\gamma > 1$, then the sum of k i.i.d. symmetric gamma random variables with shape γ has bounded density. Therefore, (6.79) follows.

We need one more fact about a folded convolution. Let f and g be two probability densities on $(-\pi,\pi)$ such that $g(x) \le M$ for all x. We claim that the folded convolution of f and g is bounded by M as well. Indeed, suppose, for instance, that $-\pi < x \le 0$. The value of the folded convolution of f and g at the point x is

$$\begin{aligned} & f * g(x) + f * g(x+2\pi) \\ = & \int_{-\pi}^{x+\pi} f(y) g(x-y)\,dy + \int_{x+\pi}^{\pi} f(y) g(x+2\pi-y)\,dy \\ \le & M \int_{-\pi}^{x+\pi} f(y)\,dy + M \int_{x+\pi}^{\pi} f(y)\,dy = M\,. \end{aligned}$$

Suppose now that $k_g d < -1$. By (6.79), $f_X^{*k_g,f}$ is bounded. For all $k > k_g$, $f_X^{*k,f}$ is the folded convolution of $f_X^{*k_g,f}$ and $f_X^{*k-k_g,f}$ and hence is bounded by the same constant. Since the coefficients in (6.77) are summable, the statement of part (ii) follows.

Suppose, finally, that $k_g d > -1$. If $-1/d$ is not an integer, then the statement of part (i) also follows from what has already been proved, because we can write

$$f_Y(x) = \sum_{k=k_g}^{\lfloor -1/d \rfloor} \frac{a_k(g)^2}{k!} f_X^{*k,f}(x) + \sum_{k=\lfloor -1/d \rfloor + 1}^{\infty} \frac{a_k(g)^2}{k!} f_X^{*k,f}(x)\,.$$

We already know that the second sum on the right-hand side is bounded. The first sum on the right-hand side is a finite sum and, by (6.78), the first term in that sum dominates, as $x \to 0$, the rest of the terms. Therefore, the statement of part (i) follows from (6.78).

An extra step is needed if $-1/d$ is an integer. We leave it to Exercise 6.5.8. □

The constant $c(k, d)$ in (6.76) can be computed explicitly; see Exercises 6.5.9 and 6.5.10.

We see, therefore, that if a stationary Gaussian process $\mathbf{X}$ has spectral density f_X satisfying (6.27) with $d \in (-1, 0)$ (and is bounded outside of the origin), and $k_g = 1$, then the spectral density of the process $\mathbf{Y}$ also satisfies (6.27), with the same value of d. If, however, $k_g > 1$, then the spectral density of $\mathbf{Y}$ may still satisfy (6.27), but with a larger d, or it may even be bounded, depending on the value of $k_g d$. Since Example 6.3.6 shows that there exist one-to-one functions g with Hermite index $k_g > 1$, we may construct examples of processes $\mathbf{X}$ that have long-range dependence according to the behavior of the spectral density at the origin, but for which the process $\mathbf{Y}$ has short memory according to the same definition even though the transformation is via a one-to-one function.

6.4 Comments on Chapter 6

Comments on Section 6.1

Much of the emphasis on the second-order point of view on long-range dependence originated with B. Mandelbrot. A power-like decay of covariances appears in Mandelbrot (1965) as an attribute of a model needed to account for the Hurst phenomenon. Both slow decay of covariances and fast rate of increase of the variances of the partial sums were explicitly introduced as crucial features of long-range dependent processes in Mandelbrot and Wallis (1968). A significant number of papers appeared in the subsequent 10–15 years discussing concrete models having regularly varying nonsummable covariances and partial sums whose variances grow faster than linearly fast; the early ones include Davydov (1970), Taqqu (1975), Rosenblatt (1976). The survey Mandelbrot and Taqqu (1979) uses persisting correlations as synonymous with long memory.

Comments on Section 6.2

A very early paper exhibiting the importance of slowly vanishing correlations and pointing out the concurrent singularity at the origin of the spectral density is Rosenblatt (1961). The exact relationship between properties (6.26) and (6.27) generated some confusion, with imprecise statements having been published. Part (i) of Theorem 6.2.11 is in Theorem (2–6) in Chapter V of Zygmund (1968).

Comments on Section 6.3

The importance of the Hermite rank for the second-order characteristics of pointwise transformations of stationary Gaussian processes was pointed out in Taqqu (1975), but a special case of this phenomenon is mentioned in Rosenblatt (1961).

6.5 Exercises to Chapter 6

Exercise 6.5.1. *Let* $\mathbf{X} = (X_n,\, n \in \mathbb{Z})$ *be an infinite moving average process of Section 1.4. Assume that the noise variables* $(\varepsilon_n,\, n \in \mathbb{Z})$ *have zero mean and finite variance. Prove that if* $\sum |\varphi_j| < \infty$, *then the covariances of the process* $\mathbf{X}$ *are absolutely summable, i.e., that* (6.2) *holds.*

Exercise 6.5.2. *Let* m *be a finite (not necessarily symmetric) measure on* $(-\pi, \pi]$. *Prove that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_{(-\pi,\pi]} e^{ijx}\, m(dx) = m(\{0\})\,.$$

Exercise 6.5.3. *Suppose that the spectral measure of a stationary finite-variance process* $\mathbf{X}$ *is given by*

$$F = \sum_{k=1}^{\infty} 2^{-k!} \left(\delta_{2^{-k!}(\pi/2)} + \delta_{-2^{-k!}(\pi/2)}\right)\,.$$

Show that $\mathrm{Var} S_n/n \to c > 0$ *as* $n \to \infty$ *over the subsequence* $n = 2^{m!}$, *while* $\mathrm{Var} S_n/n \to 0$ *as* $n \to \infty$ *over the subsequence* $n = 4 \cdot 2^{m!}$.

Exercise 6.5.4. *Suppose that the spectral measure* F_X *has, in a neighborhood of the origin, the density* f_X *given by*

$$f_X(x) = 1 + \cos(1/|x|)\,.$$

Show that (6.17) *holds with* $f_X^*(0) = 1$.

Exercise 6.5.5. *Suppose that the spectral measure* F_X *has, in a neighborhood of the origin, the density* f_X *given by*

$$f_X(x) = \big(1 + \cos(1/|x|)\big) \frac{1}{2\Gamma(-d)\cos(\pi d/2)} |x|^{-(1+d)} L\big(1/|x|\big),\ -1 < d < 0\,,$$

where L *is a function that is slowly varying at infinity. Show that* (6.28) *holds.*

Exercise 6.5.6. *Prove* (6.40) *for* $m \neq 0$.

Exercise 6.5.7. *(i) Prove, directly from the definition* (6.54) *of the Hermite polynomials, the following relations:*

$$H_n'(x) = nH_{n-1}(x),\ x \in \mathbb{R},\ n = 1, 2, \ldots \tag{6.81}$$

and

$$H_{n+1}(x) = xH_n(x) - H_n'(x),\ x \in \mathbb{R},\ n = 0, 1, \ldots\,. \tag{6.82}$$

(ii) Use part (i) to prove by induction the representation (6.55) of the Hermite polynomials.

Exercise 6.5.8. *Complete the proof of Theorem 6.3.7 in the case that $-1/d$ is an integer.*

Exercise 6.5.9. *Let f_1 and f_2 be symmetric probability density functions supported by the interval $(-\pi, \pi)$. Assume that f_1 and f_2 are bounded when their argument is bounded away from zero and that*

$$f_i(x) = x^{-(1+d_i)} L_i(1/x) \ \text{ as } x \downarrow 0$$

for $i = 1, 2$, some $-1 < d_i < 0$, and a slowly varying function L_i. Suppose that $d_1 + d_2 > -1$. Then the probability density function given by the folded convolution $f_1(*, \mathrm{f}) f_2$, is bounded when its argument is bounded away from the origin, and it satisfies

$$f_1(*, \mathrm{f}) f_2(x) \sim C(d_1, d_2) x^{-(1+d_1+d_2)} L_1(1/x) L_2(1/x) \ \text{ as } x \downarrow 0,$$

where

$$\begin{aligned} C(d_1, d_2) &= B(-d_1 -, d_2) + B(1 + d_1 + d_1, -d_1) + B(1 + d_1 + d_1, -d_2) \\ &= \frac{1}{\Gamma(-d_1 - d_2)\Gamma(1 + d_1)\Gamma(1 + d_2)} \, \frac{\pi^2}{2 \sin(\frac{\pi d_1}{2}) \sin(\frac{\pi d_2}{2}) \cos(\frac{\pi (d_1 + d_2)}{2})}. \end{aligned}$$

Exercise 6.5.10. *Use Exercise 6.5.9 to show that the constant $c(k, d)$ in (6.76) is equal to*

$$c(k, d) = 2^{k-1} \frac{\big(\Gamma(-d) \cos(\pi d/2)\big)^k}{\Gamma(-kd) \cos(\pi kd/2)},$$

$kd > -1$.

Chapter 7
Fractionally Differenced and Fractionally Integrated Processes

7.1 Fractional Integration and Long Memory

The adjective "fractional" appears frequently in the names of processes related to long-range dependence; two immediate examples are the fractional Brownian motion of Example 3.5.1 and the fractional Gaussian noise introduced in Section 5.1. The term "fractional" carries a connotation of "unusual," which is one explanation for this appearance. There is, however, a deeper connection, and it is related to the issues of stationarity and nonstationarity in the context of long-range dependence discussed in Section 5.2.

Let $\mathbf{X} = (\ldots, X_{-1}, X_0, X_1, \ldots)$ be a stationary process. Clearly, the differenced process $\mathbf{Y}$ with $Y_n = X_n - X_{n-1}$ for $n \in \mathbb{Z}$ is also stationary. The usual notation is $\mathbf{Y} = (I - B)\mathbf{X}$, where I is the identity operator on the space $\mathbb{R}^{\mathbb{Z}}$ of sequences $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$, and B is the *backward shift operator* on the same space:

$$B(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) = (\ldots, x_{-2}, x_{-1}, x_0, x_1, \ldots);$$

note that B is the inverse of the shift operator θ from Section 2.1. A common notation for the differencing operator is $\Delta = I - B$.

Does there exist (perhaps on an enlarged probability space, but in this chapter we will disregard this point) for every stationary process $\mathbf{Y}$, a stationary process $\mathbf{X}$ for which $\mathbf{Y} = (I - B)\mathbf{X}$? The answer is no, as the following example shows.

Example 7.1.1. Let $\mathbf{Y} = (\ldots, Y_{-1}, Y_0, Y_1, \ldots)$ be a sequence of i.i.d. random variables such that $P(Y_0 \neq 0) > 0$. If there existed a stationary process $\mathbf{X}$ such that $\mathbf{Y} = (I - B)\mathbf{X}$, then we would have

$$S_n := Y_1 + \ldots + Y_n = X_n - X_0 \text{ for any } n = 1, 2, \ldots.$$

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_7

The stationarity of the process $\mathbf{X}$ implies that the sequence of the laws of the differences on the right-hand side of this equation is tight, whence so must be the sequence of the laws of the sums on the left-hand side. In particular, there must exist a sequence $n_k \to \infty$ such that the sequence (S_{n_k}) converges in distribution. Note, however, that by independence, we have for the characteristic functions

$$Ee^{i\theta S_n} = \left(Ee^{i\theta Y_0}\right)^n, \ \theta \in \mathbb{R}, \ n = 1, 2, \ldots .$$

The assumption $P(Y_0 \neq 0) > 0$ implies that there are arbitrarily small positive θ such that $\left|Ee^{i\theta Y_0}\right| < 1$, and at every such point, $Ee^{i\theta S_{n_k}} \to 0$. This precludes convergence in distribution of the sequence (S_{n_k}).

We conclude that there is no stationary process $\mathbf{X}$ such that $\mathbf{Y} = (I - B)\mathbf{X}$.

According to the example, not every stationary process $\mathbf{Y}$ can be written in the form $\mathbf{Y} = (I-B)\mathbf{X}$ for some stationary process $\mathbf{X}$. If, however, $\mathbf{Y}$ can be represented in this form, we can write $\mathbf{X} = (I - B)^{-1}\mathbf{Y}$ and call the process $\mathbf{X}$ an integrated process (specifically, $\mathbf{Y}$ that has been integrated). Obviously, an integrated process, if it exists, is not uniquely determined: one can add the same random variable to each X_n, as long as doing so preserves stationarity. For example, adding a random variable independent of $\mathbf{X}$ always works.

It is intuitive that the differencing operator on stationary processes, $\Delta = I - B$, makes the memory in the process "less positive, more negative," simply because of alternating plus and minus signs attached to the same random variables. For example, the left plot of Figure 7.1 is a realization of an AR(1) process $X_n = 0.5X_{n-1} + Z_n$, with the standard Gaussian noise (Z_n); it has positive dependence due to the positive autoregressive coefficient. The right plot of the figure shows the result of differencing this sample. Increased negative dependence is easily visible.

Similarly, if it is possible to "integrate" a stationary process (i.e., to apply the inverse operator $\Delta^{-1} = (I-B)^{-1}$) and obtain a stationary process, we would expect the integrated process to have "more positive" memory than the original stationary

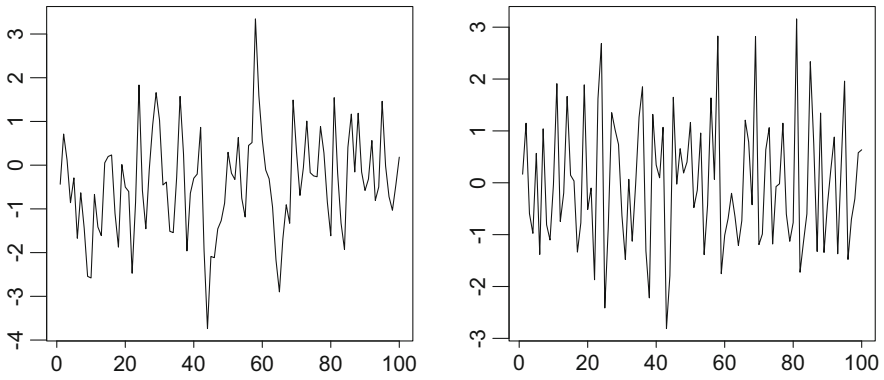

Fig. 7.1 An autoregressive Gaussian process (left plot) and its first difference (right plot)

process. Long memory, when present, is usually "of the positive kind" (we will address this point in the sequel); so, starting with some stationary process, one could try to obtain a process with long-range dependence by integrating that initial process "as many times as possible."

The difficulty with this program is that, as we saw in Example 7.1.1, many stationary processes cannot be integrated even once while preserving stationarity. Therefore, the desire to integrate a process "as many times as possible" must, in many cases, refer to a fractional number of times. This turns out to be possible, and it leads to a class of models known as *fractionally integrated processes*. The starting point is the generalized binomial formula, which was already used in Lemma 6.2.14,

$$(1-z)^d = \sum_{j=0}^{\infty} (-1)^j \binom{d}{j} z^j, \tag{7.1}$$

for complex z with $|z| < 1$, where

$$\binom{d}{j} = \frac{d(d-1)\dots(d-j+1)}{j!}.$$

If d is a nonnegative integer, then (7.1) is just the classical binomial formula and a sum with finitely many terms; otherwise, it is an infinite sum, and then it can be rewritten in the form

$$(1-z)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} z^j. \tag{7.2}$$

Notice that the coefficients of z^j in the sum (7.2) satisfy, by (6.39),

$$\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \frac{j^{-(d+1)}}{\Gamma(-d)}\Big(1 + O(1/j)\Big) \tag{7.3}$$

as $j \to \infty$. If $-1 < d < 0$, the coefficients are nonnegative and decreasing.

Now let d be a real number that is not a nonpositive integer. Given a stationary process $\mathbf{Y}$, one can formally define the process $\mathbf{X} = \Delta^{-d}\mathbf{Y}$ by expanding $\Delta^{-d} = (I - B)^{-d}$ in powers of the backward shift operator B as in (7.2) by formally identifying the identity operator with the unity and the shift operator B with z. That is, we define

$$X_n = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} Y_{n-j} \tag{7.4}$$

(as long as the sum converges, say in probability) for $n = \ldots, -1, 0, 1, 2, \ldots$. If $d > 0$, we view the process $\mathbf{X}$ as an integrated process $\mathbf{Y}$, and if $d < 0$, we view it as a differenced process $\mathbf{Y}$.

Since our goal is to construct processes with as a long memory as possible, we will be interested in the "integrated" case. We will consider the case of a "true fractional integration" $0 < d < 1$. The simplest way to perform integration with $d \geq 1$ is to start (if possible) with the usual, "nonfractional," integration, and then perform additional fractional integration as necessary.

It is clear that if the series in (7.4) converges in probability, then the resulting process $\mathbf{X}$ is stationary. However, convergence of the infinite series in (7.4) requires restrictions on the initial process $\mathbf{Y}$.

Remark 7.1.2. Since $(1 - z)^{-d}(1 - z)^d = 1$ (for z in the interior of the unit circle in the complex plane), the logic we used to define a fractionally integrated process suggests that the subsequent corresponding fractional differencing should undo the fractional integration, so that with $\mathbf{X}$ defined by (7.4), one recovers the original process $\mathbf{Y}$ via

$$Y_n = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} X_{n-j} \tag{7.5}$$

for $n = \ldots, -1, 0, 1, 2, \ldots$. This turns out to be true under appropriate assumptions on the process $\mathbf{Y}$.

We would like to draw the attention of the reader to the fact that if $0 < d < 1$ and both (7.4) and (7.5) hold, then for every real number x and for every n,

$$Y_n = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} (X_{n-j} + x)$$

as well, because by (7.2) and (7.3),

$$\sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = 0\,.$$

If we view a fractionally integrated process $\mathbf{Y}$ as a stationary process whose fractional difference in (7.5) coincides with the process $\mathbf{Y}$, then this process is not uniquely defined. In fact, exactly as in the case of the integration and differencing of order 1, we can add the same random variable to each X_n as long as doing so preserves stationarity. The definition used in (7.4) gives one version of such a process.

7.2 Fractional Integration of Second-Order Processes

In this section, we assume that the process $\mathbf{Y}$ is a stationary zero-mean finite-variance process with variance σ^2 and correlation function ρ, and we consider first the situation in which this covariance function is absolutely summable.

Theorem 7.2.1. *Let* $\mathbf{Y}$ *be a stationary zero-mean finite-variance process with covariance function* R *satisfying* (6.2). *Let* $0 < d < 1/2$.

(a) *The series* (7.4) *converges in* L^2, *and the resulting process* $\mathbf{X}$ *is a second-order stationary process, with covariance function given by*

$$R_n^* = \sum_{m=-\infty}^{\infty} b_{n-m} R_m, \ n \in \mathbb{Z}, \tag{7.6}$$

with

$$b_k = \sum_{i=-k\vee 0}^{\infty} a_i a_{i+k}, \ k \in \mathbb{Z}.$$

Furthermore, the inversion formula (7.5) *holds, with the sum converging in* L^2.

(b) *The process* $\mathbf{X}$ *has a spectral density* f^* *that is continuous outside of the origin, satisfying*

$$f^*(x) \sim \Big(\frac{1}{2\pi} \sum_{m=-\infty}^{\infty} R_m\Big) x^{-2d} \ \text{ as } x \downarrow 0. \tag{7.7}$$

(c) *Assume that*

$$\Psi_n := \sum_{m=n}^{\infty} R_m = o\Big(n^{-(1-2d)}\Big) \quad \text{as } n \to \infty. \tag{7.8}$$

Then the covariance function R^* *of the process* $\mathbf{X}$ *satisfies*

$$R_n^* \sim \Big(\frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sum_{m=-\infty}^{\infty} R_m\Big) n^{-(1-2d)} \tag{7.9}$$

as $n \to \infty$.

Proof. Denoting the jth coefficient in (7.4) by a_j, we note that for $m, k \geq 1$,

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 = R_0 \sum_{j=m+1}^{m+k} a_j^2 + 2 \sum_{j=m+1}^{m+k} a_j \sum_{i=j+1}^{m+k} a_i R_{i-j}. \tag{7.10}$$

Since the sequence (a_j) is nonnegative and decreasing, we conclude that

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 \le \left(R_0 + 2\sum_{n=1}^{\infty} |R_n|\right) \sum_{j=m+1}^{m+k} a_j^2 .$$

For $0 < d < 1/2$, the sum $\sum_j a_j^2$ converges by (7.3), and so the series (7.4) converges in L^2.

We prove now that the inversion formula (7.5) holds. For $m = 1, 2, \ldots$, consider the finite sum

$$\begin{aligned} Y_n^{(m)} &= \sum_{j=0}^{m} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} X_{n-j} \qquad (7.11) \\ &= \sum_{k=0}^{m} \left(\sum_{j=0}^{k} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \frac{\Gamma(k-j+d)}{\Gamma(k-j+1)\Gamma(d)} \right) Y_{n-k} \\ &+ \sum_{k=m+1}^{\infty} \left(\sum_{j=0}^{m} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \frac{\Gamma(k-j+d)}{\Gamma(k-j+1)\Gamma(d)} \right) Y_{n-k} . \end{aligned}$$

The following property of the generalized binomial coefficients can be easily verified from (7.1):

$$\sum_{j=0}^{k} \binom{d}{j} \binom{-d}{k-j} = 0 \quad \text{for } k \ge 1. \qquad (7.12)$$

Therefore, the first sum on the right-hand side of (7.11) is identically equal to Y_n, and it remains to prove that the second sum on the right-hand side of (7.11) converges to 0 in L^2 as $m \to \infty$. To this end, we will show that there is a finite constant c such that

$$\left| \sum_{j=0}^{m} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \frac{\Gamma(k-j+d)}{\Gamma(k-j+1)\Gamma(d)} \right| \le c k^{d-1}$$

for all $1 \le m \le k$. Once this is proved, a calculation similar to (7.10) will give us the claim we need. By (7.12), it is enough to prove that

$$\left| \sum_{j=m+1}^{k} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \frac{\Gamma(k-j+d)}{\Gamma(k-j+1)\Gamma(d)} \right| \le c k^{d-1} . \qquad (7.13)$$

Allowing a finite constant c to change from line to line, we can bound the left-hand side of (7.13) by a term $o(k^{-1})$ plus

$$c \sum_{j=m+1}^{k-1} j^{-(d+1)}(k-j)^{d-1} \le c \sum_{j=1}^{k-1} j^{-(d+1)}(k-j)^{d-1} ,$$

after which (7.13) follows easily, for example by splitting the sum into the part with $j \le k/2$ and its complement, and using Karamata's theorem.

For part (b) of the theorem, recall that the absolute summability of the correlations of the process $\mathbf{Y}$ implies that the latter process has a continuous spectral density f, given by (6.16). By Exercise 1.6.5, we know that the fractionally integrated process $\mathbf{X}$ has also a spectral density, f^*, given by $f^*(x) = \left|\sum_{m=0}^{\infty} a_m e^{imx}\right|^2 f(x)$, with the infinite sum in the expression for the density converging in $L^2\big((-\pi,\pi], f(x)dx\big)$. However, by (6.38), $\left|\sum_{m=-\infty}^{\infty} a_m e^{imx}\right|^2 = |1-e^{ix}|^{-2d}$ for every $x \in (-\pi,\pi]$ different from $x = 0$ and $x = \pi$. Therefore, a version of the spectral density f^* is given by

$$f^*(x) = |1-e^{ix}|^{-2d} f(x), \quad x \in (-\pi,\pi) . \tag{7.14}$$

This function is clearly continuous outside of the origin, and

$$f^*(x) \sim x^{-2d} f(0) = \Big(\frac{1}{2\pi} \sum_{m=-\infty}^{\infty} R_m\Big) x^{-2d}$$

as $x \downarrow 0$ by (6.16).

For part (c) of the theorem, notice that the covariance function of the process $\mathbf{X}$ is given by

$$R_n^* = \lim_{M\to\infty} \sum_{i=0}^{M}\sum_{j=0}^{M} a_i a_j R_{n+i-j} = \lim_{M\to\infty} \sum_{m=n-M}^{n+M} b_{n-m}^{(M)} R_m ,$$

where

$$b_k^{(M)} = \sum_{i=-k\vee 0}^{(M-k)\wedge M} a_i a_{i+k} .$$

Since $0 < d < 1/2$, it follows that the numbers $b_k^{(M)}$ are uniformly bounded (by $\sum_0^\infty a_i^2$). Since the correlations of the process $\mathbf{X}$ are absolutely summable, it follows by the dominated convergence theorem that (7.6) holds. Notice that $b_k = b_{-k}$.

It follows from (7.3) that

$$b_k \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{-(1-2d)} \quad \text{as } k \to \infty; \tag{7.15}$$

see Exercise 7.5.1. Clearly, the statement (7.9) will follow from (7.6) and (7.15) once we check that

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} \left| \sum_{m=-\infty}^{-M} b_{n-m} R_m \right| = 0 \tag{7.16}$$

and

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} \left| \sum_{m=M}^{\infty} b_{n-m} R_m \right| = 0\,. \tag{7.17}$$

The statement (7.16) follows from (7.15) and summability of the covariances of the process $\mathbf{Y}$, since by the monotonicity of the coefficients (a_i),

$$\left| \sum_{m=-\infty}^{-M} b_{n-m} R_m \right| \le b_n \sum_{m=-\infty}^{-M} |R_m|\,.$$

In order to prove (7.17), we note that by (7.3),

$$a_k - a_{k+1} = \frac{1-d}{k+1} a_k \sim \frac{1-d}{\Gamma(d)} k^{-(2-d)}. \tag{7.18}$$

It follows that

$$g_k := b_k - b_{k+1} \sim \frac{\Gamma(2-2d)}{\Gamma(d)\Gamma(1-d)} k^{-2(1-d)} \quad \text{as } k \to \infty, \tag{7.19}$$

see once again Exercise 7.5.1. Using summation by parts, we see that in the notation of (7.8),

$$\sum_{m=M}^{\infty} b_{n-m} R_m = b_{n-M+1} \Psi_M + \sum_{m=M}^{\infty} g_{n-m} \Psi_m\,.$$

By (7.15), for $c > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} b_{n-M+1} \Psi_M = \lim_{M\to\infty} c \Psi_M = 0\,.$$

Next, we write

$$\sum_{m=M}^{\infty} g_{n-m} \Psi_m = \sum_{M \le m \le n/2} + \sum_{m > n/2} := S_n^{(1)}(M) + S_n^{(2)}(M)\,.$$

By (7.19) and the assumption (7.8), we see that for some constant c (which may change from appearance to appearance) and large n,

$$\begin{aligned}\left|S_n^{(1)}(M)\right| &\le cn^{-2(1-d)} \sum_{m=M}^{[n/2]} |\Psi_m| \\ &\le cn^{-2(1-d)} \sum_{m=M}^{[n/2]} m^{-(1-2d)} \le cn^{-2(1-2d)}\,,\end{aligned}$$

and so for all $M > 0$,

$$\lim_{n\to\infty} n^{1-2d} S_n^{(1)}(M) = 0\,.$$

Finally, by the assumption (7.8), we have

$$\left|S_n^{(2)}(M)\right| \le o(1)n^{-(1-2d)} \sum_{m=-\infty}^{\infty} |g_m|\,.$$

Using the fact that $g_{-k} = -g_{k-1}$, we see that the sum on the right-hand side is finite by (7.19). Therefore, for all $M > 0$,

$$\lim_{n\to\infty} n^{1-2d} S_n^{(2)}(M) = 0\,,$$

and (7.17) follows. □

Note that if

$$\sum_{m=-\infty}^{\infty} \rho_m \neq 0\,, \tag{7.20}$$

then the fractionally integrated process $\mathbf{X}$ has the second-order properties (6.26) and (6.27) associated with long-range dependence. From this point of view, fractional integration achieves its goal of producing a long-range dependent model.

Remark 7.2.2. In the situation of Theorem 7.2.1, we can guarantee only a possibility of a fractionally integrated process for $0 < d < 1/2$. One can interpret this fact as follows: a sufficiently negative dependence in the process $\mathbf{Y}$ could potentially permit even a nonfractional (order-1) integration, and it is intuitive that the more negative the dependence in the process $\mathbf{Y}$, the higher the order of fractional integration one can perform on $\mathbf{Y}$. The assumptions of the theorem, however, do not imply any negative dependence in the process $\mathbf{Y}$. By the same token, if the process $\mathbf{Y}$ had a significant positive dependence, then the extent of possible fractional integration would be lower. In particular, we should not be able to integrate a very positively dependent process at all. This is, however, not allowed in Theorem 7.2.1 through

the assumption of absolute summability of covariances. This explains, intuitively, the "middle" upper bound of $1/2$ on the order of partial integration in the theorem. In the remaining part of this section, we will confirm this intuition.

We begin with the situation of positive dependence in the process $\mathbf{Y}$.

Theorem 7.2.3. *Let* $\mathbf{Y}$ *be a stationary zero-mean finite-variance process with covariance function* R.

(a) *Suppose that for some* $0 < \theta \leq 1$ *and* $a > 0$,

$$|R_n| \leq an^{-\theta}, \ n \geq 1 . \tag{7.21}$$

Then for every $0 < d < \theta/2$, *the series* (7.4) *converges in* L^2, *and the resulting process* $\mathbf{X}$ *is a second-order stationary process with covariance function given by* (7.6).

(b) *Suppose that* R_n *is regularly varying with exponent* $-\theta \in (-1, 0)$. *Then for every* $0 < d < \theta/2$, *the series* (7.4) *converges in* L^2, *and the resulting process* $\mathbf{X}$ *is a second-order stationary process with covariance function* R^* *satisfying*

$$R_n^* \sim a(d, \theta) n^{2d} R_n \tag{7.22}$$

as $n \to \infty$, *with*

$$a(d, \theta) = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \left[\frac{\Gamma(\theta-2d)\Gamma(1-\theta)}{\Gamma(1-2d)} + \frac{\Gamma(2d)\Gamma(1-\theta)}{\Gamma(1+2d-\theta)} + \frac{\Gamma(\theta-2d)\Gamma(2d)}{\Gamma(\theta)} \right] .$$

Proof. Starting with (7.10), and using (7.3) and (7.21), we see that

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j} \right)^2 \leq R_0 \sum_{j=m+1}^{\infty} a_j^2 + C \sum_{j=m+1}^{\infty} j^{d-1} \sum_{i=j+1}^{\infty} i^{d-1} (i-j)^{-\theta} .$$

The first sum on the right-hand side converges to zero as $m \to \infty$ for any $0 < d < 1/2$, while the second sum is bounded by

$$C \sum_{j=m+1}^{\infty} j^{2d-\theta-1} \to 0$$

as $m \to \infty$ if $0 < d < \theta/2$. This proves the L^2 convergence of the series (7.4).

The same argument as in the proof of Theorem 7.2.1 shows that (7.6) still holds. The covariances of the process $\mathbf{Y}$ are no longer guaranteed to be summable, but (7.15) and (7.21) allow us to use the dominated convergence theorem.

Under the regular variation assumption of part (b) of the theorem, note that by Corollary 10.5.8, (7.21) still holds if we replace θ by $\theta - \varepsilon$ for any $0 < \varepsilon < \theta$. Choosing ε so small that $d < (\theta - \varepsilon)/2$, we see that part (a) of the theorem applies,

and $\mathbf{X}$ is a stationary process with covariance function (R_n^*) given by (7.6). Write, using the fact that both (b_n) and (R_n) are even functions,

$$R_n^* = \sum_{m=1}^{\infty} b_{n+m}R_m + \sum_{m=0}^{n} b_{n-m}R_m + \sum_{m=1}^{\infty} b_m R_{n+m}\,.$$

By part (b) of Theorem 10.5.10 with $\alpha = -(1-2d)$, $\beta = -\theta$,

$$\sum_{m=1}^{\infty} b_{n+m}R_m \sim \left(\frac{\Gamma(\theta-2d)\Gamma(1-\theta)}{\Gamma(1-2d)}\right) nb_nR_n,\ n\to\infty\,.$$

By part (a) of Theorem 10.5.10 with $\alpha = -\theta$, $\beta = -(1-2d)$,

$$\sum_{m=0}^{n} b_{n-m}R_m \sim \left(\frac{\Gamma(2d)\Gamma(1-\theta)}{\Gamma(1+2d-\theta)}\right) nb_nR_n,\ n\to\infty\,.$$

Finally, by part (b) of Theorem 10.5.10 with $\alpha = -\theta$, $\beta = -(1-2d)$,

$$\sum_{m=1}^{\infty} b_m R_{n+m} \sim \left(\frac{\Gamma(\theta-2d)\Gamma(2d)}{\Gamma(\theta)}\right) nb_nR_n,\ n\to\infty\,.$$

In combination with (7.15), this proves (7.22). □

A possible way of defining a second-order notion of negative dependence is through a vanishing sum of covariances. Assume that the covariance function R of $\mathbf{Y}$ satisfies (6.2) and

$$\sum_{n=-\infty}^{\infty} R_n = 0\,. \tag{7.23}$$

The following relates the rate of convergence to zero in (7.23) to the degree to which a process can be fractionally integrated.

Theorem 7.2.4. *Let* $\mathbf{Y}$ *be a stationary zero-mean finite-variance process with an absolutely summable covariance function* R *satisfying* (7.23)*. Suppose that for some* $0 < \theta \le 1$ *and* $a > 0$*,*

$$\left|\sum_{j=n}^{\infty} R_j\right| \le an^{-\theta},\ n \ge 1\,. \tag{7.24}$$

Then for every $0 < d < (1+\theta)/2$, the series (7.4) converges in L^2, and the resulting process $\mathbf{X}$ is a second-order stationary process with covariance function given by

$$R_n^* = \sum_{j=1}^{n} W_j\big(h_{j+n-1} - h_{n-j}\big) + \sum_{j=n+1}^{\infty} W_j\big(h_{j+n-1} + h_{j-n-1}\big),\ n \in \mathbb{Z}\,. \tag{7.25}$$

Here for $k \in \mathbb{Z}$,

$$W_k = \sum_{j=k}^{\infty} R_j\,,$$

$$h_k = \sum_{i=0}^{\infty} a_i\big(a_{i+k} - a_{i+k-1}\big)\,.$$

Proof. We rewrite (7.10) as

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 = \left(\sum_{i=m+1}^{m+k} a_i^2\right) R_0 + 2\sum_{l=1}^{k-1}\left(\sum_{i=m+1}^{m+k-l} a_i a_{i+l}\right) R_l\,.$$

For fixed m, k, let

$$g_l = \sum_{i=m+1}^{m+k-l} a_i a_{i+l} - \sum_{i=m+1}^{m+k-l+1} a_i a_{i+l-1},\ l = 1, \ldots, k-1\,.$$

Then summation by parts and (7.23) give us the expression

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 = 2\sum_{j=1}^{k-1} g_j W_{j+1} - 2a_{m+1}a_{m+k}W_k\,. \tag{7.26}$$

The second term on the right-hand side of (7.26) clearly goes to zero as $m \to \infty$, uniformly in k, and we proceed to show that so does the first term. Observe that by (7.3) and (7.18),

$$\begin{aligned} |g_j| &\le |a_{m+k-j+1}a_{m+k}| + \sum_{i=m+1}^{m+k-j} |a_i(a_{i+j-1} - a_{i+j})| \\ &\le c(m+k-j+1)^{d-1}(m+k)^{d-1} + c\sum_{i=m+1}^{m+k-j} i^{d-1}(i+j)^{d-2}\,. \end{aligned}$$

Therefore, by (7.24) and elementary calculations,

$$\begin{aligned}\left|\sum_{j=1}^{k-1} g_j W_{j+1}\right| &\le c(m+k)^{d-1}\sum_{j=1}^{k-1} j^{-\theta}(m+k-j+1)^{d-1} + c\sum_{j=1}^{k-1} j^{-\theta}\sum_{i=m+1}^{m+k-j} i^{d-1}(i+j)^{d-2}\\ &\le c(m+k)^{d-1}(m+k)^{d-\theta} + c\sum_{i=m+1}^{m+k-1} i^{d-1}\sum_{j=1}^{\infty} j^{-\theta}(i+j)^{d-2}\\ &\le c(m+k)^{2d-1-\theta} + c\sum_{i=m+1}^{m+k-1} i^{d-1} i^{-\theta+d-1} \le c(m+1)^{2d-1-\theta}\,.\end{aligned}$$

Since $d < (1+\theta)/2$, this goes to zero as $m \to \infty$, and the proof of the L^2 convergence is complete. Therefore, the covariance function of the process $\mathbf{X}$ satisfies

$$R_n^* = \lim_{M\to\infty} \frac{1}{2}\left[E\left(\sum_{i=0}^{M} a_i\big(Y_{-i}+Y_{n-i}\big)\right)^2 - 2E\left(\sum_{i=0}^{M} a_i Y_{-i}\right)^2\right].$$

Notice that for a fixed n, the process $\big(Z_i^{(n)} = Y_i + Y_{n+i}\big)$ has summable covariance that moreover, adds up to zero. Therefore, (7.26) applies, and we obtain

$$R_n^* = \lim_{M\to\infty}\left[\sum_{j=0}^{M} h_j^{(M)}\big(W_{j+1}^{(n)} - 2W_{j+1}\big) - a_0 a_M\big(W_{M+1}^{(n)} - 2W_{M+1}\big)\right]$$

with

$$W_k^{(n)} = \sum_{j=k}^{\infty} E\big(Z_0^{(n)} Z_k^{(n)}\big),\ k = 1, 2, \ldots,$$

$$h_k^{(M)} = \sum_{i=0}^{M-k} a_i a_{i+k} - \sum_{i=0}^{M-k+1} a_i a_{i+k-1},\ k = 0, 1, \ldots, M\,.$$

Since

$$E\big(Z_0^{(n)} Z_k^{(n)}\big) = 2R_k + R_{k-n} + R_{k+n}\,,$$

we see that

$$R_n^* = \lim_{M\to\infty}\sum_{j=0}^{M} h_j^{(M)}\big(W_{j-n+1} + W_{j+n+1}\big)\,.$$

Writing

$$h_k^{(M)} = \sum_{i=0}^{M-k} a_i\big(a_{i+k} - a_{i+k-1}\big) - a_M a_{M-k+1} \to h_k \ \text{ as } M \to \infty,$$

by (7.3), (7.18), and (7.24), the dominated convergence theorem applies, whence

$$R_n^* = \sum_{j=0}^{\infty} h_j\big(W_{j-n+1} + W_{j+n+1}\big),\ n \in \mathbb{Z}\,.$$

Since it follows from the symmetry of the covariance function and (7.23) that for $k = 0, 1, \ldots$ we have $W_{-k} = -W_{k+1}$, (7.25) now follows by a simple rearrangement of the terms in the sum. □

7.3 Fractional Integration of Processes with Infinite Variance

To what extent can one fractionally integrate a stationary process that does not have a finite second moment? We begin with the example of an i.i.d. sequence.

Example 7.3.1. Let (Y_n) be a sequence of i.i.d. random variables, which for simplicity, we will assume to be symmetric. Let

$$p_{\mathrm{cr}} = \sup\big\{p \geq 0 :\ E|Y_0|^p < \infty\big\} \in [0, \infty]\,.$$

We are interested in the case in which the random variables do not have a finite second moment, so assume that $p_{\mathrm{cr}} \leq 2$. We claim that the fractionally integrated process $\mathbf{X}$ in (7.4) is well defined if $0 < d < 1 - 1/p_{\mathrm{cr}}$, and is not well defined if $d > 1 - 1/p_{\mathrm{cr}}$.

Indeed, suppose that $0 < d < 1 - 1/p_{\mathrm{cr}}$, and choose $0 < p < p_{\mathrm{cr}}$ such that $0 < d < 1 - 1/p$. Notice that by assumption, $0 < p < 2$. It follows from (7.3) that

$$\sum_{j=0}^{\infty} \left| \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \right|^p < \infty\,,$$

and the fact that the series (7.4) converges a.s. follows from Theorem 1.4.1. On the other hand, let $0 < d < 1$ satisfy $d > 1 - 1/p_{\mathrm{cr}}$. For every $c > 0$,

$$\sum_{j=0}^{\infty} P\big(|Y_0| > cj^{1-d}\big) = \sum_{j=0}^{\infty} P\big((|Y_0|/c)^{1/(1-d)} > j\big) = \infty,$$

since $E|Y|^{1/(1-d)} = \infty$. Now the fact that the series (7.4) diverges follows from (7.3) and the three series theorem, Theorem 10.7.6.

Example 7.3.1 shows that the extent to which an i.i.d. sequence can be fractionally integrated shrinks as the tails become heavier, and if the first moment is infinite, that such a sequence cannot be fractionally integrated at all. As in the finite-variance case, the extent of fractional integration is also affected by the dependence in the process $\mathbf{Y}$.

One possibility to measure the dependence in a stationary process in which the second moment is infinite is by the rate of growth of the moments of the partial sums, of order smaller than 2. This rate of growth is affected both by the memory and by the marginal tails.

Let $S_n^{(Y)} = Y_1 + \ldots + Y_n$, $n = 0, 1, \ldots$, be the partial-sum sequence of the process $\mathbf{Y}$, and suppose that $E|Y|^p < \infty$ for some $p > 0$. We will concentrate on conditions of the following form: for some $\theta > 0$ and $c > 0$,

$$E|S_n^{(Y)}|^p \le cn^\theta, \quad n = 1, 2, \ldots . \tag{7.27}$$

Note that it is always possible to choose

$$\theta = \max(1, p)\,, \tag{7.28}$$

and that this choice is, in general, the best possible; see Problem 7.5.3. On the other hand, if the random variables $\mathbf{Y}$ are symmetric and i.i.d., then it is always possible to choose

$$\theta = p/2\,, \tag{7.29}$$

and this choice is, once again, the best possible in general; see Problem 7.5.4. If the process $\mathbf{Y}$ satisfies (7.27) with θ smaller than the value given in (7.29), then we can view it as an indication of negative dependence. On the other hand, if the smallest possible value of θ for which (7.27) holds exceeds the value given in (7.29), then we can view it as an indication of positive dependence. The following proposition confirms that negative dependence can increase the order of the fractional integration one can perform on a process $\mathbf{Y}$.

Proposition 7.3.2. *Let a stationary process $\mathbf{Y}$ be such that for some $0 < p < 2$, $E|Y|^p < \infty$. Suppose that (7.27) holds for some $\theta > 0$. Then for every $0 < d < 1$ such that*

$$d < 2 - \frac{1}{\min(1, p)} - \frac{\theta}{p}\,, \tag{7.30}$$

the series (7.4) converges in L^p, and the resulting process $\mathbf{X}$ is a well-defined stationary process.

Proof. We consider the sum (7.4) for $n = 0$. For notational simplicity, we reverse the time in the process $\mathbf{Y}$. Note that marginally, the partial sums of the time-reversed process have the same law as those of the original process, so condition (7.27) does not change. The summation by parts gives us, for $m, k \geq 1$,

$$\sum_{j=m+1}^{m+k} a_j Y_j = \sum_{j=m}^{m+k-1} (a_j - a_{j+1}) S_j^{(Y)} + a_{m+k} S_{m+k}^{(Y)} - a_m S_m^{(Y)} . \tag{7.31}$$

Note, first of all, that by (7.27) and (7.3),

$$E\left|a_n S_n^{(Y)}\right|^p \leq c n^{p(d-1)} n^{\theta} \to 0$$

as $n \to \infty$, since $d < 1 - \theta/p$. Therefore, the last two terms on the right-hand side of (7.31) converge to zero in L^p as $m \to \infty$, uniformly in $k \geq 1$. We now consider the first term on the right-hand side of (7.31). Suppose first that $p \geq 1$. We use Theorem 10.7.1 with $q = 1$ to obtain

$$\begin{aligned} \left(E \left| \sum_{j=m}^{m+k-1} (a_j - a_{j+1}) S_j^{(Y)} \right|^p \right)^{1/p} &\leq \sum_{j=m}^{m+k-1} \left(E|S_j|^p \right)^{1/p} |a_j - a_{j+1}| \\ &\leq c \sum_{j=m}^{m+k-1} j^{-(2-d)} j^{\theta/p} \to 0 \end{aligned}$$

as $m \to \infty$, because $d < 1 - \theta/p$. If $0 < p < 1$, we proceed similarly, using the bound

$$\left| \sum_{j=m}^{m+k-1} (a_j - a_{j+1}) S_j^{(Y)} \right|^p \leq \sum_{j=m}^{m+k-1} |a_j - a_{j+1}|^p |S_j^{(Y)}|^p .$$

This shows the L^p convergence, and stationarity is obvious. □

Note that if $p \leq 1/2$, then no $d > 0$ will satisfy the constraint (7.30), regardless of how small $\theta > 0$ is. It turns out that in the case of very heavy tails, much of our intuition about fractional integration turns out to be misleading, as the following example demonstrates. It shows that in the case of very heavy tails, a stationary stochastic process that can be integrated completely (with $d = 1$) may fail to allow any fractional integration.

Example 7.3.3. Let $\mathbf{Z}$ be a sequence of i.i.d. symmetric random variables in the domain of attraction of an α-stable law with $0 < \alpha < 2$, and let $Y_n = Z_n - Z_{n-1}$, $n \in \mathbb{Z}$. Then the process $\mathbf{Y}$ has marginal tails that are regularly varying with exponent $-\alpha$. see, e.g., Feller (1971). Furthermore, the partial sums of this process satisfy $S_n^{(Y)} = Z_n - Z_0$, which shows that (7.27) holds with any $0 < p < \alpha$ and $\theta = 0$.

Depending on the exact tails of the process $\mathbf{Z}$, (7.27) may also hold with $p = \alpha$, but it cannot hold with $p > \alpha$. In particular, if $0 < \alpha \leq 1/2$, then Proposition 7.3.2 does not give any positive value of d for possible fractional integration. It turns out that this is, in fact, the case: the process $\mathbf{Y}$ cannot be fractionally integrated, in spite of the fact that it can be integrated with $d = 1$.

To see this, suppose that the process $\mathbf{Y}$ can be fractionally integrated for some $0 < d < 1$, so that the series (7.4) converges in probability. Then the expression on the right-hand side of (7.31) should converge to zero in probability as $m \to \infty$, uniformly in k. Since $a_n \to 0$ as $n \to \infty$, the last two terms on the right-hand side of (7.31) converge to zero in probability, so that the same must be true for the first term. Since $S_n^{(Y)} = Z_n - Z_0$ for all n, we conclude that the series

$$\sum_{j=1}^{\infty} (a_j - a_{j+1}) Z_j$$

converges in probability. This is a series of independent random variables, so that by the Itô–Nisio theorem (Theorem 10.7.6), this series must converge a.s. This is, however, impossible, since by the Borel–Cantelli lemma, $|Z_j|$ exceeds j^2 infinitely often, while $j^2 |a_j - a_{j+1}| \to \infty$ as $j \to \infty$ by (7.18).

7.4 Comments on Chapter 7

Most of the discussion related to convergence of the series (7.4) and, in the case of finite variance, to the behavior of the resulting covariance function, depends on little more than the asymptotic order of the magnitude of the coefficients in the infinite series (7.4) and their differences. The specific choice of the coefficients in (7.4), with their relation to the generalized binomial formula, is attractive both because of its intuitive meaning and because of parsimony arising from dependence on a single parameter $0 < d < 1$.

In practice, one often begins with an i.i.d. process $\mathbf{Y}$ or a stationary ARMA model of Example 1.4.4; see Brockwell and Davis (1987). A stationary ARMA process $\mathbf{Y}$ has exponentially fast decaying correlations, and Theorem 7.2.1 applies. The resulting fractionally integrated processes $\mathbf{X}$ are typically called ARI(integrated)MA processes or, more explicitly, fractional ARIMA (or FARIMA, alternatively ARFIMA) models, and were originally introduced in Granger and Joyeux (1980) and Hosking (1981). Such models are popular because they are specified by the finitely many parameters of the ARMA process $\mathbf{Y}$, in addition to the single parameter d of fractional integration.

Fractionally integrated processes, especially FARIMA models, have found numerous applications in economics and econometrics; two examples are Crato and Rothman (1994) and Gil-Alana (2004). In this area of application one would like to combine fractional integration with the so-called *clustering of*

volatility, or *conditional heteroscedasticity*. The standard (but nonfractional) model with clustering of volatility is the Generalized AutoregRessive Conditionally Heteroscedastic (or GARCH) process, introduced in Engle (1982) in its original (nongeneralized) form and generalized in Bollerslev (1986). A possible way of introducing clustering of volatility into a fractionally integrated model is to start with a process $\mathbf{Y}$ in (7.4) that has the clustering of volatility property, for example with a GARCH process. This approach is suggested in Hauser and Kunst (1998). Even earlier on, an alternative model was suggested in Baillie et al. (1996). This model directly combines fractional differencing/integration with the recursion for computation of the conditional variance of each subsequent observation, and has become known as a fractionally integrated GARCH (or FIGARCH) model. This model has proved difficult to analyze; even existence of a stationary version of the model that has desired properties is an issue. Recent progress has been made in Douc et al. (2008); see also Zaffaroni (2004).

7.5 Exercises to Chapter 7

Exercise 7.5.1. *Prove* (7.15) *and* (7.19). *You may need to use the fact that for* $0 < a < b$,

$$\int_0^\infty y^{a-1}(1+y)^{-b}\,dy = \frac{\Gamma(a)\Gamma(b-a)}{\Gamma(b)}\,.$$

Exercise 7.5.2. *In Theorem 7.2.4, assume that* S_n *is regularly varying with exponent* $-\theta \in (-1,0)$. *State and prove a version of part (b) of Theorem 7.2.3 that corresponds to this negatively dependent case.*

Exercise 7.5.3. *Prove that* (7.27) *holds for some* $c > 0$ *if one chooses* θ *as in* (7.28). *Prove also that for every* θ *strictly smaller than the value given* (7.28), *there exists a symmetric stationary process* $\mathbf{Y}$ *with* $E|Y|^p < \infty$ *such that* (7.27) *fails, no matter what* $c > 0$ *is.*

Exercise 7.5.4. *Suppose that* $\mathbf{Y}$ *consists of symmetric i.i.d. random variables, and* $E|Y|^p < \infty$. *Prove that* (7.27) *holds for some* $c > 0$ *if one chooses* θ *as in* (7.29). *Prove also that for every* θ *strictly smaller than the value given in* (7.29), *there exists a symmetric i.i.d. process* $\mathbf{Y}$ *with* $E|Y|^p < \infty$ *such that* (7.27) *fails, no matter what* $c > 0$ *is.*

Chapter 8
Self-Similar Processes

8.1 Self-Similarity, Stationarity, and Lamperti's Theorem

The notion of self-similarity was already introduced in Section 3.5. Recall that a stochastic process $\big(X(t),\ t \geq 0\big)$ is called *self-similar* if for some $H \in \mathbb{R}$,

$$\big(X(ct),\ t \geq 0\big) \stackrel{d}{=} \big(c^H X(t),\ t \geq 0\big) \tag{8.1}$$

in the sense of equality of finite-dimensional distributions for every $c > 0$. The number H is the exponent of self-similarity, and we will often say simply that the process is H-self-similar. The point $t = 0$ in the definition (8.1) of self-similarity is clearly "separate." If $\big(X(t),\ t > 0\big)$ is a stochastic process satisfying the condition

$$\big(X(ct),\ t > 0\big) \stackrel{d}{=} \big(c^H X(t),\ t > 0\big) \tag{8.2}$$

for all $c > 0$, then extending the time domain of the process by setting $X(0) = 0$ will always result in a process on $[0, \infty)$ satisfying the full condition (8.1). We will use either of the definitions, (8.1) or (8.2), as appropriate. Sometimes it is even appropriate to define self-similarity for stochastic processes indexed by the entire real line, as opposed to its nonnegative half.

Self-similarity is an invariance property with respect to certain simultaneous transformations of time and space. The following simple result connects the property of self-similarity to the property of stationarity.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_8

Proposition 8.1.1. (a) *Let* $\big(X(t),\, t > 0\big)$ *be H-self-similar. Then the process* $\big(Y(t),\, t \in \mathbb{R}\big)$ *defined by*

$$Y(t) = e^{-tH} X(e^t),\ t \in \mathbb{R}\,,$$

is stationary.

(b) *Let* $\big(Y(t),\, t \in \mathbb{R}\big)$ *be a stationary process, and let* $H \in \mathbb{R}$. *Then the process* $\big(X(t),\, t > 0\big)$ *defined by*

$$X(t) = t^H Y(\log t),\ t > 0,$$

is H-self-similar.

Proof. Both statements are immediate consequences of the definitions of self-similarity and stationarity. □

The transformations of self-similar processes into stationary processes and conversely, as described in Proposition 8.1.1, are sometimes called the *Lamperti transformations*. The following corollary is an immediate consequence of the proposition and Theorem 1.3.3.

Corollary 8.1.2. *Every measurable self-similar stochastic process* $\big(X(t),\, t > 0\big)$ *is continuous in probability.*

Note that appending the point $t = 0$ by setting $X(t) = 0$ can destroy the continuity in probability while preserving the measurability of the process.

Self-similar processes are ubiquitous in many areas of probability because they arise in common functional limit theorems. In fact, in many cases they turn out to be the only possible weak limits. The generic situation may be described as follows. Let $\big(U(t),\, t \geq 0\big)$ be a stochastic process. Let (a_n) be a sequence of positive numbers increasing to infinity. Many limit theorems in probability theory are of the form

$$\Big(\frac{1}{a_n} U(nt),\, t \geq 0\Big) \Rightarrow \big(Y(t),\, t \geq 0\big) \tag{8.3}$$

as $n \to \infty$, at least in terms of convergence of finite-dimensional distributions and, more frequently, in terms of weak convergence on an appropriate space of functions.

Example 8.1.3. Let $X_1, X_2, \ldots$ be a sequence of random variables. Let $S_0 = 0$ and $S_n = X_1 + \ldots + X_n$ for $n = 1, 2, \ldots$. If we define a stochastic process by $U(t) = S_{\lfloor t \rfloor},\ t \geq 0$, then a typical functional law of large numbers or a functional central limit theorem is a result of the type (8.3). Notice that this formulation does not allow certain types of centering in a theorem.

Example 8.1.4. Let $X_1, X_2, \ldots$ be a sequence of nonnegative random variables. Let $M_0 = 0$ and $M_n = \max(X_1, \ldots, X_n)$ for $n = 1, 2, \ldots$. If we define a stochastic process by $U(t) = M_{\lfloor t \rfloor},\ t \geq 0$, then a typical functional extremal limit theorem is a

result of the type (8.3). Once again, this formulation does not allow certain types of centering in a theorem.

It turns out that under reasonably weak assumptions, all possible limiting processes $\mathbf{Y}$ in (8.3) are self-similar. Together with Example 8.1.4 and Example 8.1.3, this explains the appearance of self-similar processes in functional central and extremal limit theorems. The first result of this type was proved by Lamperti (1962) (in a slightly different form, allowing extra centering). In particular, statements such as our Theorems 8.1.5 and 8.1.6 below are often called *Lamperti's theorems*.

Theorem 8.1.5. *Let $(U(t),\, t \geq 0)$ be a stochastic process and $a_n \uparrow \infty$ a sequence of positive numbers. Assume that (8.3) holds in terms of convergence of finite-dimensional distributions. Assume further that the limiting process $\mathbf{Y}$ satisfies $P(Y(1) \neq 0) > 0$ and is continuous in law. That is, for every $k = 1, 2, \ldots$, if $\big(s_1^{(m)}, \ldots, s_k^{(m)}\big) \to (t_1, \ldots, t_k)$, then $\big(Y(s_1^{(m)}), \ldots, Y(s_k^{(m)})\big) \Rightarrow \big(Y(t_1), \ldots, Y(t_k)\big)$. Then $\mathbf{Y}$ is H-self-similar for some $H \geq 0$, and the sequence (a_n) is regularly varying with exponent H.*

Proof. Let $k = 1, 2, \ldots$. Note that

$$\frac{1}{a_{kn}} U(n) = \frac{1}{a_{kn}} U\big(kn/k\big) \Rightarrow Y(1/k)\,.$$

If $l = \liminf a_n/a_{kn}$ and $L = \limsup a_n/a_{kn}$, then the above and the obvious expression

$$\frac{1}{a_{kn}} U(n) = \frac{1}{a_n} U(n)\, \frac{a_n}{a_{kn}}$$

show that

$$Y(1/k) \stackrel{d}{=} l\, Y(1) \stackrel{d}{=} L\, Y(1)\,.$$

Since $P(Y(1) \neq 0) > 0$, it follows that $l = L$, and hence the limit

$$\varphi(k) = \lim_{n \to \infty} \frac{a_{kn}}{a_n} \tag{8.4}$$

exists. Note that if $\varphi(k) = \infty$, then

$$\frac{1}{a_{kn}} U(kn) = \frac{1}{a_n} U(kn)\, \frac{a_n}{a_{kn}} \to 0$$

in probability. Since the left-hand side above also converges weakly to $Y(1)$, this contradicts the nondegeneracy assumption on $Y(1)$. Hence φ is a finite nondecreasing function, with values in $[1, \infty)$, and for every $t > 0$ and $k = 1, 2, \ldots$,

$$Y(kt) \stackrel{d}{=} \varphi(k)\, Y(t)\,.$$

This implies that for all $k_1, k_2 = 1, 2, \ldots$,

$$Y(1) \stackrel{d}{=} \frac{\varphi(k_1)}{\varphi(k_2)} Y\big(k_2/k_1\big)\,.$$

Since $\mathbf{Y}$ is continuous in law and $P(Y(1) \neq 0) > 0$, we conclude that

$$\lim_{k_2/k_1 \to 1} \frac{\varphi(k_1)}{\varphi(k_2)} = 1\,. \tag{8.5}$$

Next, we claim that

$$\lim_{n\to\infty} \frac{a_n}{a_{n+1}} = 1\,. \tag{8.6}$$

By the monotonicity of the sequence (a_n), it is enough to prove only the appropriate statement about the lower limit. To this end, choose $k_1, k_2 = 1, 2, \ldots, k_2 > k_1$. Then for large n,

$$n+1 \le \left\lfloor \frac{k_2}{k_1}\, n \right\rfloor, \quad \text{which implies that} \quad \frac{a_n}{a_{n+1}} \ge \frac{a_n}{a_{\lfloor (k_2/k_1)\, n \rfloor}}.$$

Let $jk_1 \le n < (j+1)k_1$ for $j = 1, 2, \ldots$. Then, by the monotonicity of the sequence (a_n), the above inequality implies

$$\frac{a_n}{a_{n+1}} \ge \frac{a_{jk_1}}{a_{(j+1)k_2}}\,.$$

However, for large j, $(j+1)k_2 \le j(k_2+1)$, and so for large n,

$$\frac{a_n}{a_{n+1}} \ge \frac{a_{jk_1}}{a_{j(k_2+1)}} \to \frac{\varphi(k_1)}{\varphi(k_2+1)}$$

as n, and hence j, increase to infinity. Therefore, for every $k_1, k_2 = 1, 2, \ldots, k_2 > k_1$,

$$\liminf_{n\to\infty} \frac{a_n}{a_{n+1}} \ge \frac{\varphi(k_1)}{\varphi(k_2+1)}\,.$$

Setting $k_2 = k_1 + 1$ and letting $k_1 \to \infty$, we can use (8.5) to obtain (8.6).

Next we show that for every $k_1, k_2 = 1, 2, \ldots$,

$$\lim_{n\to\infty} \frac{a_{\lfloor (k_2/k_1)\, n \rfloor}}{a_n} = \frac{\varphi(k_2)}{\varphi(k_1)}\,. \tag{8.7}$$

Indeed, as before, letting $jk_1 \le n < (j+1)k_1$ for $j = 1, 2, \ldots$, we can use monotonicity of the sequence (a_n) and (8.6) to obtain

$$\frac{a_{\lfloor (k_2/k_1)\, n \rfloor}}{a_n} \le \frac{a_{(j+1)k_2}}{a_{jk_1}} = \frac{a_{jk_2}}{a_{jk_1}} \frac{a_{jk_2+k_2}}{a_{jk_2}} \to \frac{\varphi(k_2)}{\varphi(k_1)}$$

and

$$\frac{a_{\lfloor (k_2/k_1)\, n\rfloor}}{a_n} \geq \frac{a_{jk_2}}{a_{(j+1)k_1}} = \frac{a_{jk_2}}{a_{jk_1}} \frac{a_{jk_1}}{a_{jk_1+k_1}} \to \frac{\varphi(k_2)}{\varphi(k_1)},$$

which proves (8.7). In particular, the ratio $\varphi(k_2)/\varphi(k_1)$ cannot decrease as long as k_2/k_1 does not decrease.

Let $c > 0$. We will check now that the limit

$$\psi(c) = \lim_{n\to\infty} \frac{a_{\lfloor cn\rfloor}}{a_n} \tag{8.8}$$

exists. Indeed, by (8.7) and monotonicity of the sequence (a_n),

$$\lim_{k_2/k_1\uparrow c} \frac{\varphi(k_2)}{\varphi(k_1)} \leq \liminf_{n\to\infty} \frac{a_{\lfloor cn\rfloor}}{a_n} \leq \limsup_{n\to\infty} \frac{a_{\lfloor cn\rfloor}}{a_n} \leq \lim_{k_2/k_1\downarrow c} \frac{\varphi(k_2)}{\varphi(k_1)}\,.$$

Moreover, it follows from (8.5) that the two limits above coincide. Therefore, (8.8) follows. Clearly, ψ is a positive nondecreasing function, and $\psi(k) = \varphi(k)$ for $k = 1, 2, \ldots$. Furthermore, for every $c_1, c_2 > 0$,

$$\begin{aligned}\psi(c_1c_2) &= \lim_{n\to\infty} \frac{a_{\lfloor c_1c_2n\rfloor}}{a_n} \geq \limsup_{n\to\infty} \frac{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor}}{a_n} \\ &= \lim_{n\to\infty} \frac{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor}}{a_{\lfloor c_2n\rfloor}} \frac{a_{\lfloor c_2n\rfloor}}{a_n} = \psi(c_1)\psi(c_2),\end{aligned}$$

and

$$\begin{aligned}\psi(c_1c_2) &\leq \liminf_{n\to\infty} \frac{a_{\lfloor c_1(\lfloor c_2n\rfloor+1)\rfloor}}{a_n} \leq \liminf_{n\to\infty} \frac{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor+\lceil c_1\rceil}}{a_n} \\ &= \lim_{n\to\infty} \frac{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor}}{a_{\lfloor c_2n\rfloor}} \frac{a_{\lfloor c_2n\rfloor}}{a_n} \frac{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor+\lceil c_1\rceil}}{a_{\lfloor c_1\lfloor c_2n\rfloor\rfloor}} = \psi(c_1)\psi(c_2),\end{aligned}$$

where in the last step we used (8.6). We conclude that

$$\psi(c_1c_2) = \psi(c_1)\psi(c_2) \text{ for all } c_1, c_2 > 0. \tag{8.9}$$

Since ψ is a monotone, hence measurable, function, it follows from (8.9) that $\psi(c) = c^H$ for some $H \geq 0$; see, e.g., Theorem 1.1.9 in Bingham et al. (1987). Setting $h(t) = a_{[t]}$ for $t \geq 1$, we may use an argument identical to that used to prove (8.9) to show that for every $c > 0$,

$$\lim_{t\to\infty} \frac{h(ct)}{h(t)} = \lim_{t\to\infty} \frac{a_{[ct]}}{a_{[t]}} = \psi(c) = c^H,$$

which shows that h, and hence the sequence (a_n), are regularly varying with exponent H.

The final step is to prove the self-similarity of the process $\mathbf{Y}$. We will show that for every $c > 0$,

$$\Big(\frac{1}{a_{\lfloor cn\rfloor}} U(cnt),\, t \geq 0\Big) \Rightarrow \big(Y(t),\, t \geq 0\big)\,. \tag{8.10}$$

Once this is established, the self-similarity of $\mathbf{Y}$ will follow from the obvious facts that

$$\Big(\frac{1}{a_n} U(cnt),\, t \geq 0\Big) \Rightarrow \big(Y(ct),\, t \geq 0\big)\,,$$

and

$$\Big(\frac{1}{a_n} U(cnt),\, t \geq 0\Big) = \Big(\frac{a_{\lfloor cn\rfloor}}{a_n} \frac{1}{a_{\lfloor cn\rfloor}} U(cnt),\, t \geq 0\Big) \Rightarrow \big(c^H Y(t),\, t \geq 0\big)\,.$$

To prove (8.10), let $k_1, k_2 = 1, 2, \ldots$. By taking the limit over the subsequence $n = jk_1, j = 1, 2, \ldots$, we see that (8.10) holds for every rational $c = k_2/k_1$. Let (c_m) be a sequence of positive rational numbers converging to c. By the continuity in law of the process $\mathbf{Y}$, we see that

$$\Big(Y\big((c/c_m)t\big),\, t \geq 0\Big) \Rightarrow \big(Y(t),\, t \geq 0\big)$$

as $m \to \infty$. Choose now $t_1, \ldots, t_j \geq 0$. The regular variation of the function $h(t) = a_{\lfloor t\rfloor}$, $t \geq 1$, implies that for every $\lambda > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\bigg(\bigg\| \frac{1}{a_{\lfloor cn\rfloor}} \Big(U(cnt_1), \ldots, (cnt_j)\Big) - \frac{1}{a_{\lfloor c_m n\rfloor}} \Big(U(cnt_1), \ldots, (cnt_j)\Big)\bigg\| > \lambda\bigg) = 0\,.$$

By Theorem 3.2 in Billingsley (1999), we conclude that

$$\frac{1}{a_{\lfloor cn\rfloor}} \Big(U(cnt_1), \ldots, (cnt_j)\Big) \Rightarrow \big(Y(ct_1), \ldots, Y(ct_j)\big)$$

is $\mathbb{R}^j$, and so (8.10) follows. □

The second version of Lamperti's theorem strengthens the assumption (8.3) to continuous scaling. As a result, fewer a priori assumptions are required on the limiting process $\mathbf{Y}$.

Theorem 8.1.6. *Let* $(U(t),\, t \geq 0)$ *be a stochastic process, and* $(A(\lambda),\, \lambda > 0)$ *a positive measurable function such that* $A(\lambda) \to \infty$ *as* $\lambda \to \infty$. *Assume that*

$$\Big(\frac{1}{A(\lambda)} U(\lambda t),\, t \geq 0\Big) \Rightarrow \big(Y(t),\, t \geq 0\big) \tag{8.11}$$

as $\lambda \to \infty$, *in terms of convergence of finite-dimensional distributions, to a limiting process* $\mathbf{Y}$ *such that* $P(Y(1) \neq 0) > 0$. *Then* $\mathbf{Y}$ *is* H*-self-similar for some* $H \geq 0$, *and the function* A *is regularly varying with exponent* H.

The proof is a simplification of the proof of Theorem 8.1.5, and is left as an exercise (Exercise 8.7.1).

Self-similar processes arising as limits in a situation such as (8.3) often have additional invariance properties. We begin with the property of stationary increments; see Definition 1.1.8. The first statement is obvious.

Proposition 8.1.7. *Suppose that* (8.3) *holds and the process* $\mathbf{U}$ *has stationary increments. Then so does the process* $\mathbf{Y}$.

The next proposition addresses the situation of Example 8.1.3. Note that Proposition 8.1.7 does not directly apply in this case.

Proposition 8.1.8. *Let* $X_1, X_2, \ldots$ *be a stationary stochastic process, and define* $U(t) = S_{\lfloor t \rfloor} = X_1 + \ldots + X_{\lfloor t \rfloor}$, $t \geq 0$. *Let* $a_n \to \infty$ *be a sequence of positive numbers, and assume that* (8.3) *holds. Then the process* $\mathbf{Y}$ *has stationary increments.*

Proof. Fix $s \geq 0$, $m \geq 1$, and positive $t_1, \ldots, t_m$. Notice that by stationarity,

$$\begin{aligned}
&\frac{1}{a_n}\Big(U\big(n(t_1+s)\big) - U(ns), \ldots, U\big(n(t_m+s)\big) - U(ns)\Big) \\
=&\frac{1}{a_n}\Big(X_{\lfloor ns \rfloor + 1} + \ldots + X_{\lfloor nt_1 + ns \rfloor}, \ldots, X_{\lfloor ns \rfloor + 1} + \ldots + X_{\lfloor nt_m + ns \rfloor}\Big) \\
\stackrel{d}{=}&\frac{1}{a_n}\Big(X_1 + \ldots + X_{\lfloor nt_1 + ns \rfloor - \lfloor ns \rfloor}, \ldots, X_1 + \ldots + X_{\lfloor nt_m + ns \rfloor - \lfloor ns \rfloor}\Big) \\
=&\frac{1}{a_n}\Big(U\big(nt_1\big), \ldots, U\big(nt_m\big)\Big) + \frac{1}{a_n}\Big(X_{\lfloor nt_1 \rfloor + 1}\theta_{n,1}, X_{\lfloor nt_m \rfloor + 1}\theta_{n,m}\Big),
\end{aligned}$$

where $\theta_{n,1}, \ldots, \theta_{n,m}$ are numbers taking values in $\{0, 1\}$. Since $a_n \to \infty$, it follows from (8.3) that the expression on the left-hand side converges weakly to $\big(Y(t_1+s) - Y(s), \ldots, Y(t_m+s) - Y(s)\big)$, while the expression on the right-hand side converges weakly to $\big(Y(t_1) - Y(0), \ldots, Y(t_m) - Y(0)\big)$. Therefore, the two limits have the same laws, which proves the stationary of the increments of the process $\mathbf{Y}$. □

An invariance property analogous to the stationarity of the increments, but more appropriate to the case of the partial maxima described in Example 8.1.4, is given in the following definition.

Definition 8.1.9. A stochastic process $\big(X(t),\, t \ge 0\big)$ has stationary max-increments if for every $s \ge 0$, there exists, perhaps on an enlarged probability space, a stochastic process $\big(X^{(s)}(t),\, t \ge 0\big)$ such that

$$\begin{aligned} \big(X^{(s)}(t),\, t \ge 0\big) &\stackrel{d}{=} \big(X(t),\, t \ge 0\big)\,, \\ \big(X(t+s),\, t \ge 0\big) &\stackrel{d}{=} \big(X(s) \vee X^{(s)}(t),\, t \ge 0\big)\,. \end{aligned} \tag{8.12}$$

The following statement is analogous to Proposition 8.1.8.

Proposition 8.1.10. *Let $X_1, X_2, \ldots$ be nonnegative random variables forming a stationary stochastic process, and define $U(t) = M_{\lfloor t \rfloor} = \max(X_1, \ldots, X_{\lfloor t \rfloor})$, $t \ge 0$. Let $a_n \to \infty$ be a sequence of positive numbers, and assume that (8.3) holds. Then the process $\mathbf{Y}$ is continuous in probability and has stationary max-increments.*

Proof. First of all, it is not difficult to check that with $A(\lambda) = a_n$ if $n - 1 \le \lambda < n$, $n = 1, 2, \ldots$, the process $\mathbf{U}$ satisfies the continuous scaling assumption (8.11) (Exercise 8.7.3). If the limiting process $\mathbf{Y}$ is the zero process, then the claim of the proposition is trivial. If it is not the zero process, then Theorem 8.1.6 applies, and we conclude that the process $\mathbf{Y}$ is self-similar with exponent $H \ge 0$. If $\mathbf{Y}$ is a constant process, then once again, the claim of the proposition is trivial. Assume that this is not the case. Then the exponent of the self-similarity H is greater than zero.

For every $0 \le t_1 < t_2$ and n large enough, by stationarity,

$$\frac{1}{a_n}\big(M_{\lfloor nt_2 \rfloor} - M_{\lfloor nt_1 \rfloor}\big) \le \frac{1}{a_n} \max_{nt_1 < i \le nt_2} X_i \stackrel{st}{\le} \frac{1}{a_n} M_{\lfloor 2n(t_2 - t_1) \rfloor}\,.$$

Taking weak limits, we see that the difference $Y(t_2) - Y(t_1)$ is nonnegative and bounded stochastically by $Y\big(2(t_2 - t_1)\big)$. Therefore, it follows from the self-similarity of $(Y(t),\, t \ge 0)$ with $H > 0$ that it is continuous in probability.

It remains to check the stationarity of the max-increments. Let $r > 0$, and $t_i > 0$, $i = 1, \ldots, k$, for some $k \ge 1$. Write

$$\frac{1}{a_n} M_{\lfloor n(t_i + r) \rfloor} = \frac{1}{a_n} M_{\lfloor nr \rfloor} \bigvee \frac{1}{a_n} \max_{nr < j \le n(t_i + r)} X_j,\ i = 1, \ldots, k\,. \tag{8.13}$$

By the assumption of the theorem and stationarity of the process $(X_1, X_2, \ldots)$,

$$\begin{aligned} &\frac{1}{a_n} M_{\lfloor nr \rfloor} \Rightarrow Y(r) \ \text{ in } \mathbb{R}, \\ &\Big(\frac{1}{a_n} \max_{nr < j \le n(t_i + r)} X_j,\ i = 1, \ldots, k\Big) \Rightarrow \big(Y(t_1), \ldots, Y(t_k)\big) \ \text{ in } \mathbb{R}^k \end{aligned}$$

as $n \to \infty$. Since every weakly converging sequence is tight, and a sequence with tight marginals is itself tight, we conclude that

$$\Big(\frac{1}{a_n} M_{\lfloor nr \rfloor},\ \Big(\frac{1}{a_n} \max_{nr < j \le n(t_i + r)} X_j,\ i = 1, \ldots, k\Big)\Big)$$

is a tight sequence in $\mathbb{R} \times \mathbb{R}^k$. This tightness means that for every sequence $n_m \to \infty$, there exist a subsequence $n_{m(l)} \to \infty$ and a k-dimensional random vector $\big(Y^{(r)}(t_1), \ldots, Y^{(r)}(t_k)\big) \stackrel{d}{=} \big(Y(t_1), \ldots, Y(t_k)\big)$ such that as $l \to \infty$,

$$\begin{aligned}&\left(\frac{1}{a_{n_{m(l)}}} M_{\lfloor n_{m(l)} r \rfloor},\ \left(\frac{1}{a_{n_{m(l)}}} \max_{n_{m(l)} r < j \le n_{m(l)}(t_i + r)} X_j,\ i = 1, \ldots, k\right)\right)\\ &\Rightarrow \Big(Y(r),\ \big(Y^{(r)}(t_1), \ldots, Y^{(r)}(t_k)\big)\Big)\,.\end{aligned}$$

Let now τ_i, $i = 1, 2, \ldots$, be an enumeration of the rational numbers in $[0, \infty)$. A diagonalization argument shows that there exist a sequence $n_m \to \infty$ and a stochastic process $\big(Y^{(r)}(\tau_i),\ i = 1, 2, \ldots\big)$, with $\big(Y^{(r)}(\tau_i),\ i = 1, 2, \ldots\big) \stackrel{d}{=} \big(Y(\tau_i),\ i = 1, 2, \ldots\big)$, such that

$$\begin{aligned}&\left(\frac{1}{a_{n_m}} M_{\lfloor n_m r \rfloor},\ \left(\frac{1}{a_{n_m}} \max_{n_m r < j \le n_m(\tau_i + r)} X_j,\ i = 1, 2, \ldots\right)\right)\\ &\Rightarrow \Big(Y(r),\ \big(Y^{(r)}(\tau_i),\ i = 1, 2, \ldots\big)\Big)\end{aligned} \tag{8.14}$$

in finite-dimensional distributions, as $m \to \infty$. Obviously, the stochastic process $\big(Y^{(r)}(\tau_i),\ i = 1, 2, \ldots\big)$ has sample paths that are, a.s., nondecreasing on the rationals. We can therefore extend the process $\mathbf{Y}^{(r)}$ to the entire positive half-line by setting

$$Y^{(r)}(t) = \frac{1}{2}\Big(\lim_{\tau \uparrow t,\ \text{rational}} Y^{(r)}(\tau) + \lim_{\tau \downarrow t,\ \text{rational}} Y^{(r)}(\tau)\Big),\quad t \ge 0\,.$$

The continuity in probability implies that this process is a version of $(Y(t),\ t \ge 0)$. This continuity in probability, (8.14), and monotonicity imply that as $m \to \infty$,

$$\left(\frac{1}{a_{n_m}} M_{\lfloor n_m r \rfloor},\ \left(\frac{1}{a_{n_m}} \max_{n_m r < j \le n_m(t + r)} X_j,\ t \ge 0\right)\right) \Rightarrow \Big(Y(r),\ \big(Y^{(r)}(t),\ t \ge 0\big)\Big) \tag{8.15}$$

in finite-dimensional distributions. Now the stationarity of max-increments follows from (8.13), (8.15), and the continuous mapping theorem. □

8.2 General Properties of Self-Similar Processes

In this section, we explore the general properties of stochastic processes that are self-similar, or are both self-similar and have stationary increments. We will use certain common abbreviations. Thus, a self-similar process with be often called an

SS process, or an H-SS process if we want to emphasize the exponent H of self-similarity. Similarly, a self-similar process with stationary increments will be often called an SSSI process, or an H-SSSI process.

We begin with a simple statement about inclusion of time zero in the domain of the process.

Lemma 8.2.1. *Let* $\big(X(t),\, t \geq 0\big)$ *be self-similar with exponent* $H \neq 0$. *Then* $X(0) = 0$ *a.s.*

Proof. By self-similarity,

$$X(0) = X(c \cdot 0) \stackrel{d}{=} c^H X(0)$$

for every $c > 0$. Letting $c \to 0$ if $H > 0$ and $c \to \infty$ if $H < 0$ proves the claim. □

It is clear that there are many H-SS processes $\big(X(t),\, t > 0\big)$ with an arbitrary $H \in \mathbb{R}$; the deterministic process $X(t) = t^H$, $t > 0$, provides one example. Once we insist on the property of stationary increments as well, the domain of the Hurst exponent becomes, apart from a small number of mostly degenerate examples, much smaller.

Lemma 8.2.2. *Let* $\big(X(t),\, t \geq 0\big)$ *be an* H*-SSSI process with* $H < 0$. *Then* $\mathbf{X}$ *is the zero process, i.e.,* $P(X(t) = 0) = 1$ *for each* $t \geq 0$.

Proof. By self-similarity,

$$X(t) \stackrel{d}{=} t^H X(1) \to 0$$

in probability as $t \to \infty$. By Lemma 8.2.1 and the stationarity of the increments, for every $t \geq 0$,

$$X(t) = X(t) - X(0) \stackrel{d}{=} X(t+b) - X(b)$$

for every $b \geq 0$. Letting $b \to \infty$ shows that $X(t) = 0$ a.s. □

Lemma (8.2.2) rules out a negative Hurst exponent for a nontrivial H-SSSI process. The following example shows that there exist nontrivial H-SSSI processes with $H = 0$.

Example 8.2.3. Let $\big(X(t),\, t \geq 0\big)$ be an (uncountable) collection of i.i.d. random variables as in Example 1.3.1. This process is trivially self-similar with $H = 0$. It is also a stationary process, hence a process with stationary increments.

The uncountable collection of i.i.d. random variables in this example is an "ugly" process. Since this process is stationary but not continuous in probability (unless the i.i.d. random variables are equal to a constant), it follows from Theorem 1.3.3 that this process is not even measurable. It turns out that only trivial H-SSSI processes with $H = 0$ are measurable.

Proposition 8.2.4. *Let* $\big(X(t),\, t \geq 0\big)$ *be a measurable H-SSSI process with* $H = 0$. *Then* $\mathbf{X}$ *is a constant process, i.e.,* $P(X(t) = X(0)) = 1$ *for each* $t \geq 0$.

Proof. Let $t \geq 0$ and $c > 0$. Using first stationarity of the increments, then self-similarity, and then once again stationarity of the increments, we have

$$X(c+t) - X(t) \stackrel{d}{=} X(c+c) - X(c) \stackrel{d}{=} X(1+1) - X(1) \stackrel{d}{=} X(1) - X(0)\,.$$

Since the process $\mathbf{X}$ is measurable, Theorem 1.3.3 implies that it is continuous in probability. Letting $c \downarrow 0$, we immediately conclude that $P(X(1) = X(0)) = 1$. By self-similarity, $P(X(t) = X(0)) = 1$ for all $t > 0$. □

Lemma 8.2.2 and Proposition 8.2.4 show why in most applications we encounter only H-SSSI processes with $H > 0$. In the remainder of this section we will consider only this range of the exponent of self-similarity. The first property of such processes is elementary.

Lemma 8.2.5. *Every H-SSSI process with* $H > 0$ *is continuous in probability.*

Proof. Let $\big(X(t),\, t \geq 0\big)$ be an H-SSSI process with $H > 0$. For every $s, t \geq 0$ and $\varepsilon > 0$, using first stationarity of the increments together with Lemma 8.2.1 and then self-similarity, we see that

$$P\big(|X(t) - X(s)| > \varepsilon\big) = P\big(|X(|t-s|)| > \varepsilon\big) = P\big(|t-s|^H |X(1)| > \varepsilon\big) \to 0$$

as $|t - s| \to 0$, whence the continuity in probability. □

It turns out that finiteness of certain moments of H-SSSI processes constrains how large the exponent H can be.

Proposition 8.2.6. *Let* $\big(X(t),\, t \geq 0\big)$ *be an H-SSSI process such that* $P(X(1) \neq 0) > 0$.

(a) *Suppose that* $E|X(1)|^\gamma < \infty$ *for some* $0 < \gamma < 1$. *Then* $H < 1/\gamma$.
(b) *Suppose that* $E|X(1)| < \infty$. *Then* $H \leq 1$.

Proof. Since part (b) of the proposition is a trivial consequence of part (a), it is enough to prove the latter. Let

$$A_i = \big\{X(i) - X(i-1) \neq 0\big\},\ i = 1, 2, \ldots\,.$$

It follows from Lemma 8.2.1 and stationarity of the increments that for every $i = 1, 2, \ldots$, we have $P(A_i) = p > 0$, where $p = P(X(1) \neq 0)$. If $n > 1/p$, then the events $A_1, \ldots, A_n$ cannot be disjoint up to sets of probability zero, so that there are *different* numbers $1 \leq j_1, j_2 \leq n$ such that $P(A_{j_1} \cap A_{j_2}) > 0$. Consider the identity

$$E\big(|X(n)|^\gamma\big) = E\left(\left|\sum_{i=1}^{n} \big(X(i) - X(i-1)\big)\right|^\gamma\right).$$

A simple concavity argument shows that for all real numbers $a_1, \ldots, a_n$, we have

$$|a_1 + \ldots a_n|^\gamma \leq |a_1|^\gamma + \ldots + |a_n|^\gamma ,$$

and the inequality is strict unless at most one of the numbers $a_1, \ldots, a_n$ is different from zero. That is,

$$\left| \sum_{i=1}^{n} \big(X(i) - X(i-1)\big) \right|^\gamma \leq \sum_{i=1}^{n} \big|X(i) - X(i-1)\big|^\gamma ,$$

and the inequality is strict on a set of positive probability (at least on the set $A_{j_1} \cap A_{j_2}$ above). Therefore, by the stationarity of the increments,

$$E\big(|X(n)|^\gamma\big) < E\left(\sum_{i=1}^{n} \big|X(i) - X(i-1)\big|^\gamma \right) = nE\big(|X(1)|^\gamma\big) .$$

Since by self-similarity,

$$E\big(|X(n)|^\gamma\big) = n^{\gamma H} E\big(|X(1)|^\gamma\big),$$

and moreover,

$$0 < E\big(|X(1)|^\gamma\big) < \infty ,$$

we conclude that $H < 1/\gamma$, as claimed. □

Corollary 8.2.7. *Let $\big(X(t),\, t \geq 0\big)$ be an SαS H-SSSI process, $H > 0$, $0 < \alpha < 2$. Then the exponent H of self-similarity must be in the range*

$$H \in \begin{cases} (0, 1/\alpha] & \text{if } 0 < \alpha \leq 1, \\ (0, 1] & \text{if } 1 < \alpha < 2. \end{cases} \tag{8.16}$$

Are the bounds on the value of the exponent of self-similarity in Proposition 8.2.6 and Corollary 8.2.7 the best possible?

Example 8.2.8. SαS Lévy motion The law of a Lévy process $\big(X(t),\, t \geq 0\big)$ is determined by the one-dimensional distribution of $X(1)$. Let us choose the latter such that for some $0 < \alpha \leq 2$ and $\sigma > 0$,

$$Ee^{i\theta X(t)} = \big(Ee^{i\theta X(1)}\big)^t = e^{-t\sigma^\alpha |\theta|^\alpha}, \ \theta \in \mathbb{R}, \ t > 0 . \tag{8.17}$$

Then $X(1)$ is an SαS random variable if $0 < \alpha < 2$, and a zero-mean normal random variable if $\alpha = 2$. In the former case, the Lévy motion is an SαS Lévy motion, while in the latter case, it is a Brownian motion. By definition, these processes have stationary increments. They are also self-similar with $H = 1/\alpha$. To see this, note

that for every $c > 0$, the two processes in (3.71) are both Lévy motions, so we need to check only that they have the same one-dimensional marginal distribution at time $t = 1$. By (8.17),

$$Ee^{i\theta X(c\cdot 1)} = e^{-c\sigma^\alpha|\theta|^\alpha} = e^{-(c^H\sigma)^\alpha|\theta|^\alpha} = Ee^{i(c^H\theta)X(1)},$$

as required.

Let now $0 < \gamma < 1$. An SαS Lévy motion with $\gamma < \alpha < 2$ is a self-similar process with stationary increments and a finite absolute moment of order γ; see Example 4.2.8. The exponents of self-similarity $H = 1/\alpha$ in this family fill the interval $(1/2, 1/\gamma)$. The fractional Brownian motion of Example 3.5.1 is a Gaussian process, and hence all of its absolute moments are finite. It is also an H-SSSI process with exponent of self-similarity anywhere in the range $(0, 1)$. Since the union of the intervals $(1/2, 1/\gamma)$ and $(0, 1)$ is the interval $(0, 1/\gamma)$, we conclude that for every H in this interval, there is a nondegenerate H-SSSI process with a finite absolute moment of order γ. Therefore, the bounds on H in part (a) of Proposition 8.2.6 are the best possible.

The fractional Brownian motion of Example 3.5.1 also provides an example of a finite-mean H-SSSI process with any exponent of self-similarity in the range $(0, 1)$. Part (b) of Proposition 8.2.6 allows the value $H = 1$ as well. Such an H-SSSI process is the *straight-line process* in Example 8.2.9 below.

The bounds on H in Corollary 8.2.7 are also the best possible. The linear fractional symmetric stable motion of Example 3.5.2 and harmonizable fractional stable motion of Example 3.5.4 allow any $H \in (0, 1)$, regardless of the value of α, and the straight-line process of Example 8.2.9 below has $H = 1$, again regardless of the value of α. Finally, if $0 < \alpha < 1$, then the FBM-local time fractional stable motion of Example 3.5.5 allows any $H \in (1, 1/\alpha)$.

Example 8.2.9. Let $X(1)$ be an arbitrary random variable. Then the process $X(t) = tX(1)$, $t \geq 0$, is obviously an H-SSSI process with $H = 1$.

If $X(1)$ has a finite mean, then the straight-line process of Example 8.2.9, together with the fractional Brownian motion, shows that the bounds on H in part (b) of Proposition 8.2.6 are the best possible as well. The straight-line process is, however, a degenerate process, so one may wonder whether there exist nondegenerate examples of this type. It turns out that the answer is no, and the validity of this statement does not even require finiteness of the mean of the process $\mathbf{X}$. Only *existence* of the mean is required, as was shown in Vervaat (1985).

Recall that a random variable X has a mean if either $EX_+ < \infty$ or $EX_- < \infty$, but not necessarily both.

Proposition 8.2.10. *Let* $\big(X(t),\, t \geq 0\big)$ *be an SSSI process with* $H = 1$ *such that* $X(1)$ *has a mean. Then* $\mathbf{X}$ *is a straight-line process.*

Proof. By self-similarity, $X(n)/n \stackrel{d}{=} X(1)$ for all $n \geq 1$. Since the increment process $\mathbf{X}$ is stationary and has a mean, it follows from the pointwise ergodic theorem (2.7)

and Exercise 2.6.5 that

$$\frac{X(n)}{n} = \frac{1}{n}\sum_{j=1}^{n}\big(X(j) - X(j-1)\big) = \frac{1}{n}\sum_{j=1}^{n} Y_j \to E\Big(Y_1\Big|\mathcal{I}_{\mathbf{Y}}\Big) = E\Big(X(1)\Big|\mathcal{I}_{\mathbf{Y}}\Big)$$

with probability 1, where $\mathcal{I}_{\mathbf{Y}}$ is the invariant σ-field for the increment process $Y_j = X(j) - X(j-1)$, $j = 1, 2, \ldots$. We conclude that $X(1) \stackrel{d}{=} E\big(X(1)\big|\mathcal{I}_{\mathbf{Y}}\big)$. By Lemma 8.2.11 below, this implies that $X(1)$ is measurable with respect to the completion of $\mathcal{I}_{\mathbf{Y}}$. Since the observation $Y_1 = X(1)$ of the stationary process $\mathbf{Y}$ is measurable with respect to the completion of the invariant σ-field $\mathcal{I}_{\mathbf{Y}}$ for the process $\mathbf{Y}$, it follows that $Y_1 = Y_2 = \ldots$ with probability 1, and so $X(1) = X(n) - X(n-1)$ a.s. for all $n \geq 1$. That is, $X(n) = nX(1)$ a.s. for every $n = 1, 2, \ldots$. Rewriting this in the form $X(n)/n - X(m)/m = 0$ a.s. for all $n, m = 1, 2, \ldots$, we can use the self-similarity to see that $X(n/m) - (n/m)X(1) = 0$ a.s. for all $n, m = 1, 2, \ldots$, and so the relation $X(t) = tX(1)$ a.s. holds for every rational $t \geq 0$. Since the process $\mathbf{X}$ is, by Lemma 8.2.5, continuous in probability, we can extend this relation to an arbitrary $t > 0$ by selecting a sequence of rational numbers converging to t. □

Lemma 8.2.11. *Let X be a random variable with a mean on a probability space $\big(\Omega, \mathcal{F}, P\big)$. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. If $X \stackrel{d}{=} E\big(X|\mathcal{G}\big)$, then X is measurable with respect to the completion of $\mathcal{G}$.*

Proof. Assume that $EX_- < \infty$. It follows from the assumptions that $\big|E\big(X|\mathcal{G}\big)\big| < \infty$ a.s. For $a \in \mathbb{R}$, let

$$\varphi_a(x) = \begin{cases} 1 - \frac{1}{x+a+1} & \text{if } x \geq -a, \\ x + a & \text{if } x \leq -a. \end{cases}$$

Notice that φ_a is a concave function that is, moreover, strictly concave on $(-a, \infty)$. Moreover, $E\big|\varphi_a(X)\big| < \infty$, and the assumption of the lemma shows that

$$\varphi_a(X) \stackrel{d}{=} \varphi_a\Big(E\big(X|\mathcal{G}\big)\Big)\,. \tag{8.18}$$

By Jensen's inequality for conditional expectations, we obtain

$$\varphi_a\Big(E\big(X|\mathcal{G}\big)\Big) \geq E\big(\varphi_a(X)|\mathcal{G}\big) \text{ a.s.} \tag{8.19}$$

for every $a \in \mathbb{R}$.

Suppose that there are $s < t$ such that

$$P\Big(w:\, P\big(X \leq s|\mathcal{G}\big) > 0,\, P\big(X > t|\mathcal{G}\big) > 0\Big) > 0\,. \tag{8.20}$$

By the continuity from below for conditional and unconditional probabilities, there is $-\infty < b < s$ such that

$$P\Big(w:\ P\big(b < X \le s|\mathcal{G}\big) > 0,\ P\big(X > t|\mathcal{G}\big) > 0\Big) > 0\,. \tag{8.21}$$

Let $\Omega_{b,s,t}$ be the event of positive probability in (8.21), and choose $a > -b$. Let $\mu(\cdot,\omega)$ be the regular conditional distribution of X given $\mathcal{G}$. It follows from (8.21) that for almost every $\omega \in \Omega_{b,s,t}$, $\mu(\cdot,\omega)$ assigns positive values to both (b,s) and (t,∞). Since the function φ_a is strictly concave on $(-a,\infty) \supset (b,\infty)$ and $\mu(\cdot,\omega)$ assigns a positive value to that interval, and is not a point mass there, it follows that Jensen's inequality is strict: for almost every $\omega \in \Omega_{b,s,t}$,

$$\begin{aligned}\varphi_a\Big(E\big(X|\mathcal{G}\big)\Big) &= \varphi_a\left(\int_{-\infty}^{\infty} x\,\mu(dx,\omega)\right)\\ &> \int_{-\infty}^{\infty} \varphi_a(x)\,\mu(dx,\omega) = E\big(\varphi_a(X)|\mathcal{G}\big)\,.\end{aligned}$$

That is, the inequality (8.19) is strict on a set of positive probability, and hence

$$E\left[\varphi_a\Big(E\big(X|\mathcal{G}\big)\Big)\right] > E\left[E\big(\varphi_a(X)|\mathcal{G}\big)\right] = E\varphi_a(X).$$

This contradicts (8.18), and hence there are no numbers $s < t$ for which (8.20) holds. This means that there is an event $\Omega_0 \in \mathcal{G}$ with $P\big(\Omega_0\big) = 0$ such that for every $\omega \notin \Omega_0$ and rational numbers $\tau_1 < \tau_2$, either $P\big(X \le \tau_1|\mathcal{G}\big) = 0$ or $P\big(X > \tau_2|\mathcal{G}\big) = 0$.

The statement of the lemma will follow once we prove that for every rational τ, the event $\big\{X \le \tau\big\}$ is in the completion of the σ-field $\mathcal{G}$. This last claim is an immediate consequence of the following property, which we presently prove:

$$\text{if } A = \big\{w:\ X(\omega) \le \tau\big\} \text{ and } B = \big\{w:\ P\big(X \le \tau|\mathcal{G}\big)_{\omega} > 0\big\},$$

$$\text{then } P\big(A \triangle B\big) = 0. \tag{8.22}$$

Indeed, one the one hand,

$$P\big(A \cap B^c\big) = E\Big[\mathbf{1}_{P(X\le\tau|\mathcal{G})=0} P\big(X \le \tau|\mathcal{G}\big)\Big] = 0.$$

On the other hand,

$$\begin{aligned}P\big(A^c \cap B\big) &= P\Big(\big\{X > \tau,\ P\big(X \le \tau|\mathcal{G}\big) > 0\big\} \cap \Omega_0^c\Big)\\ &= \lim_{n\to\infty} P\Big(\big\{X > \tau + 1/n,\ P\big(X \le \tau|\mathcal{G}\big) > 0\big\} \cap \Omega_0^c\Big)\,.\end{aligned}$$

However, by the definition of the event Ω_0, for every $n \geq 1$,

$$P\Big(\big\{X > \tau + 1/n,\, P\big(X \leq \tau|\mathcal{G}\big) > 0\big\} \cap \Omega_0^c\Big)$$
$$= E\Big[\mathbf{1}_{P(X\leq\tau|\mathcal{G})>0}\, \mathbf{1}_{\Omega_0^c}\, P\big(X > \tau + 1/n|\mathcal{G}\big)\Big] = 0\,.$$

Therefore, $P\big(A^c \cap B\big) = 0$, and (8.22) follows. □

Nondegenerate H-SSSI processes with a finite mean must therefore satisfy $0 < H < 1$. For such processes, we have the following easy statement.

Lemma 8.2.12. *Let $\big(X(t),\, t \geq 0\big)$ be an H-SSSI process with $0 < H < 1$ and a finite mean. Then $EX(1) = 0$.*

Proof. The statement follows from the fact that $EX(2)$ is a finite number that is equal to both $2^H EX(1)$ by self-similarity and to $2EX(1)$ by the stationarity of increments. □

The correlation function of a finite-variance H-SSSI process turns out to be uniquely determined by the exponent of self-similarity H.

Proposition 8.2.13. *Let $\big(X(t),\, t \geq 0\big)$ be a finite-variance H-SSSI process, $0 < H < 1$. Let $\sigma^2 = EX(1)^2$. Then*

$$\mathrm{Cov}\big(X(s), X(t)\big) = \frac{\sigma^2}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right],\ s, t \geq 0\,. \tag{8.23}$$

Proof. By the stationarity of the increments and self-similarity, for $0 \leq s < t$,

$$E\big(X(t) - X(s)\big)^2 = EX(t-s)^2 = (t-s)^{2H}\sigma^2\,.$$

Now (8.23) follows from the obvious identity

$$\mathrm{Cov}\big(X(s), X(t)\big) = \frac{1}{2}\left[EX(t)^2 + EX(s)^2 - E\big(X(t) - X(s)\big)^2\right].$$

□

Since the covariance function uniquely determines the law of a centered Gaussian process, it follows immediately that up to a scale factor, the law of a Gaussian H-SSSI process is uniquely determined. We saw in Example 3.5.1 that a Gaussian H-SSSI process in fact exists, and it is a fractional Brownian motion. Its covariance function is, by necessity, given by (8.23). This shows, in particular, that the function on the right-hand side of (8.23) is nonnegative definite.

8.3 SSSI Processes with Finite Variance

We saw in Proposition 8.2.13 that all H-SSSI processes, $0 < H < 1$, with finite variance share the same covariance function, given by (8.23). This guarantees, of course, that there is a unique H-SSSI Gaussian process for each $0 < H < 1$, namely the fractional Brownian motion. Are there non-Gaussian H-SSSI processes with finite variance?

There are multiple ways by which such processes can be constructed; the simplest is to take a fractional Brownian motion $\big(B_H(t),\, t \geq 0\big)$ and a random variable A independent of it. Then the process

$$X(t) = AB_H(t),\ t \geq 0, \tag{8.24}$$

is clearly an H-SSSI process. It has a finite variance if A does. In this section, we will concentrate instead on a particular class of H-SSSI processes with finite variance, those given as multiple integrals with respect to a Brownian motion; see Section 10.6. Such processes are often constructed by starting with a representation of the fractional Brownian motion as described in Example 3.5.1 and extending it appropriately.

Let $k \geq 1$, and $0 < H < 1$. Suppose that a family of kernels $Q_t^{(k)} : \mathbb{R}^k \to \mathbb{R}$, $t \geq 0$, has the following properties. For every $t \geq 0$, $Q_t^{(k)} \in L^2(\mathbb{R}^k, \lambda_k)$. Furthermore, for all $0 \leq s < t$ and $c > 0$,

$$Q_t^{(k)}(x_1, \ldots, x_k) - Q_s^{(k)}(x_1, \ldots, x_k) = Q_{t-s}^{(k)}(x_1 - s, \ldots, x_k - s)\,, \tag{8.25}$$

and

$$Q_{ct}^{(k)}(cx_1, \ldots, cx_k) = c^{H-k/2} Q_t^{(k)}(x_1, \ldots, x_k) \tag{8.26}$$

for almost all $(x_1, \ldots, x_k)$. Let B be a standard Brownian motion, which we view as a Gaussian random measure on $\mathbb{R}$; see Examples 3.2.3 and 3.2.4. We define

$$X^{(k)}(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k)\, B(dx_1) \ldots B(dx_k),\ t \geq 0\,. \tag{8.27}$$

Theorem 8.3.1. *The process $\mathbf{X}^{(k)}$ in (8.27) is a well-defined H-SSSI process, and its increment process $X_i = X^{(k)}(i) - X^{(k)}(i-1)$, $i = 1, 2, \ldots$ is a stationary mixing process.*

Proof. The fact that the process $\mathbf{X}^{(k)}$ is well defined follows from Proposition 10.6.1. We will prove its self-similarity; the stationarity of the increments can be proven in a similar manner. Let $c > 0$. We need to check that for every $j = 1, 2 \ldots$ and $0 \leq t_1 < \ldots < t_j$,

$$\big(X^{(k)}(ct_1), \ldots, X^{(k)}(ct_j)\big) \stackrel{d}{=} \big(c^H X^{(k)}(t_1), \ldots, c^H X^{(k)}(t_j)\big)\,. \tag{8.28}$$

To this end, choose for each $i = 1, \dots, j$ a sequence of simple functions $\big(f_{i,n},\ n = 1, 2, \dots\big)$ vanishing on the diagonals such that $f_{i,n} \to Q_{t_i}^{(k)}$ in $L^2(\mathbb{R}^k, \lambda_k)$ as $n \to \infty$. Proposition 10.6.1 tells us that this is possible, and it also tells us that

$$\begin{aligned}&\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{i,n}(x_1, \dots, x_k)\, B(dx_1) \dots B(dx_k)\\ &\to \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_{t_i}^{(k)}(x_1, \dots, x_k)\, B(dx_1) \dots B(dx_k)\end{aligned}$$

in L^2 as $n \to \infty$ for $i = 1, \dots, j$. Furthermore, each scaled function $f_{i,n}\big(\cdot/c, \dots, \cdot/c\big)$ is also a simple function that vanishes on the diagonals. It is clear that we also have $f_{i,n}\big(\cdot/c, \dots, \cdot/c\big) \to Q_{t_i}^{(k)}\big(\cdot/c, \dots, \cdot/c\big)$ in $L^2(\mathbb{R}^k, \lambda_k)$, implying that

$$\begin{aligned}&\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{i,n}\big(x_1/c, \dots, x_k/c\big)\, B(dx_1) \dots B(dx_k)\\ &\to \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_{t_i}^{(k)}\big(x_1/c, \dots, x_k/c\big)\, B(dx_1) \dots B(dx_k)\end{aligned}$$

in L^2 as $n \to \infty$ for $i = 1, \dots, j$.

For every $i = 1, \dots, j$, we can use (8.26) to obtain

$$\begin{aligned}X^{(k)}(ct_i) &= c^{H-k/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_{t_i}^{(k)}\big(x_1/c, \dots, x_k/c\big)\, B(dx_1) \dots B(dx_k)\\ &= c^{H-k/2} \lim_{n\to\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{i,n}\big(x_1/c, \dots, x_k/c\big)\, B(dx_1) \dots B(dx_k).\end{aligned}$$

Suppose that the simple function $f_{i,n}$ is given by

$$f_{i,n}\big(x_1, \dots, x_k\big) = \sum_{l_1=1}^{r_{i,n}} \dots \sum_{l_k=1}^{r_{i,n}} a_{i,n}(l_1, \dots, l_k) \mathbf{1}\big(x_1 \in A_{l_1}^{i,n}\big) \dots \mathbf{1}\big(x_k \in A_{l_k}^{i,n}\big),$$

for some $r_{i,n} \geq 1$, some array $\big(a_{i,n}(l_1, \dots, l_k)\big)$, and some disjoint Borel sets $A_1^{i,n}, \dots, A_{r_{i,n}}^{i,n}$ of finite Lebesgue measure. Then by Proposition 10.6.1 and the $1/2$-self-similarity of the Brownian motion, we have, with convergence in L^2,

$$\begin{aligned}&\big(X^{(k)}(ct_1), \dots, X^{(k)}(ct_j)\big)\\ &= c^{H-k/2} \lim_{n\to\infty} \Bigg(\sum_{l_1=1}^{r_{1,n}} \dots \sum_{l_k=1}^{r_{1,n}} a_{1,n}(l_1, \dots, l_k) B\big(cA_{l_1}^{1,n}\big) \dots B\big(cA_{l_k}^{1,n}\big), \dots,\end{aligned}$$

$$\sum_{l_1=1}^{r_{j,n}} \dots \sum_{l_k=1}^{r_{j,n}} a_{j,n}(l_1,\dots,l_k) B\big(cA_{l_1}^{j,n}\big) \dots B\big(cA_{l_k}^{j,n}\big)\Bigg)$$

$$\stackrel{d}{=} c^{H-k/2} \lim_{n\to\infty} \Bigg(\sum_{l_1=1}^{r_{1,n}} \dots \sum_{l_k=1}^{r_{1,n}} a_{1,n}(l_1,\dots,l_k)\, c^{1/2} B\big(A_{l_1}^{1,n}\big) \dots c^{1/2} B\big(A_{l_k}^{1,n}\big), \dots,$$

$$\sum_{l_1=1}^{r_{j,n}} \dots \sum_{l_k=1}^{r_{j,n}} a_{j,n}(l_1,\dots,l_k)\, c^{1/2} B\big(A_{l_1}^{j,n}\big) \dots c^{1/2} B\big(A_{l_k}^{j,n}\big)\Bigg)$$

$$= c^H \big(X^{(k)}(t_1), \dots, X^{(k)}(t_j)\big)\,,$$

thus proving (8.28).

Since the stationarity of the increment process $\mathbf{X}$ follows from the stationarity of the increments of the process $\mathbf{X}^{(k)}$, it remains only to prove mixing of the process $\mathbf{X}$. For $n = 0, \pm 1, \pm 2, \dots$, consider the increments of the Brownian motion

$$\Delta_n(i,m) := B\big(n + i2^{-m}\big) - B\big(n + (i-1)2^{-m}\big)\,,$$

$m = 1, 2 \dots$, and $i = 1, \dots, 2^m$. Choose positive constants $\big(a(i,m),\ m = 1, 2 \dots,\ i = 1, \dots, 2^m\big)$ such that

$$\sum_{m=1}^{\infty} \sum_{i=1}^{2^m} a(i,m)^2 \Delta_n(i,m)^2 < \infty$$

a.s. for all n. By identifying $\big\{(m,i) :\ m = 1, 2, \dots,\ i = 1, \dots, 2^m\big\}$ with $\mathbb{N}$, we can view

$$\mathbf{a\Delta} = \Big(\big(a(i,m)\Delta_n(i,m),\ m = 1, 2, \dots,\ i = 1, \dots, 2^m\big),\ n \in \mathbb{Z}\Big)$$

as a two-sided sequence of i.i.d. random vectors in l^2.

Note that by (8.25),

$$X_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \big(Q_n^{(k)}(x_1,\dots,x_k) - Q_{n-1}^{(k)}(x_1,\dots,x_k)\big)\, B(dx_1) \dots B(dx_k) \tag{8.29}$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_1^{(k)}(x_1 - (n-1), \dots, x_k - (n-1))\, B(dx_1) \dots B(dx_k)\,,$$

$n = 1, 2, \dots$. Select a sequence of simple functions (f_m) vanishing on the diagonals such that $f_m \to Q_1^{(k)}$ in $L^2(\mathbb{R}^k, \lambda_k)$. Note that we can select the functions (f_m) in such a way that each function is of the form

$$f_m\big(x_1, \dots, x_k\big) = \sum_{l_1=1}^{r_m} \dots \sum_{l_k=1}^{r_m} a_m(l_1,\dots,l_k) \mathbf{1}\big(x_1 \in A_{l_1}^m\big) \dots \mathbf{1}\big(x_k \in A_{l_k}^m\big),$$

and each set A_l is a binary interval of the type $\left(i2^{-k}, j2^{-k}\right)$ for some $i < j$ in $\mathbb{Z}$ and $k \geq 1$. Select a subsequence of (f_m) if necessary to achieve the a.s. convergence in

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_m(x_1 - (n-1), \dots, x_k - (n-1))\, B(dx_1) \dots B(dx_k) \tag{8.30}$$
$$\to \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_1^{(k)}(x_1 - (n-1), \dots, x_k - (n-1))\, B(dx_1) \dots B(dx_k)$$

for each $n \in \mathbb{Z}$ as $m \to \infty$. Let θ be the left shift on $(l^2)^{\mathbb{Z}}$. For each m, because of the form of the function f_m, we have

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_m(x_1 - (n-1), \dots, x_k - (n-1))\, B(dx_1) \dots B(dx_k) = \varphi_m\left(\theta^n(\mathbf{a}\boldsymbol{\Delta})\right),$$

where $\varphi_m : l^2 \to \mathbb{R}$ is a measurable function. It follows from (8.29) and (8.30) that there is a measurable function $\varphi : l^2 \to \mathbb{R}$ such that $X_n = \varphi\left(\theta^n(\mathbf{a}\boldsymbol{\Delta})\right)$ for every $n = 1, 2 \dots$. The left shift on a sequence space with respect to a probability measure generated by an i.i.d. sequence is mixing. We can now apply Proposition 2.2.4 to conclude that the process $X_n = \varphi\left(\theta^n(\mathbf{a}\boldsymbol{\Delta})\right)$, $n \in \mathbb{Z}$ is mixing as well. □

Example 8.3.2. The Rosenblatt–Mori–Oodaira kernel. Let $0 < H < 1$. For $k = 1, 2, \dots$ and $1/2 < \gamma < 1/2 + 1/(2k)$, $H + k\gamma \neq 1 + k/2$, let

$$Q_t^{(k)}(x_1, \dots, x_k) = \int_{\max\{x_1, \dots, x_k\}}^{\infty} \prod_{j=1}^{k} (v - x_j)^{-\gamma} \left(|v|^{H+k\gamma-1-k/2} - |v-t|^{H+k\gamma-1-k/2}\right) dv\,,$$

$t \geq 0$. It is elementary (if a bit tedious) to check that this kernel is in $L^2(\mathbb{R}^k, \lambda_k)$. The fact that this kernel has the properties (8.25) and (8.26) is obvious.

More kernels satisfying the conditions of Theorem 8.3.1 can be created, starting with the explicit form of one-dimensional kernels in (3.74) and (3.75) and extending that to several dimensions. Our next examples are of this type.

Example 8.3.3. The Taqqu kernel. Let $1/2 < H < 1$. For $k \geq 1$, set

$$Q_t^{(k)}(x_1, \dots, x_k) = \int_0^t \prod_{j=1}^{k} \left((v - x_j)_+\right)^{-(1/2+(1-H)/k)} dv\,,$$

$t \geq 0$ (notice that the integral vanishes unless $\max(x_1, \dots, x_k) \leq t$.) Once again, $Q_t^{(k)} \in L^2(\mathbb{R}^k, \lambda_k)$ for all $t \geq 0$, and (8.25) and (8.26) hold.

One can define a version of the process $\mathbf{X}^{(k)}$ corresponding to the kernel in Example 8.3.3 in the case $0 < H < 1/2$ as well.

Example 8.3.4. Let $0 < H < 1/2$. For $k \geq 1$, set for $t \geq 0$,

$$Q_t^{(k)}(x_1, \ldots, x_k) = \mathbf{1}\big(\max(x_1, \ldots, x_k) < t\big) \int_t^\infty \prod_{j=1}^k (v - x_j)^{-(1/2+(1-H)/k)} dv$$

$$- \mathbf{1}\big(\max(x_1, \ldots, x_k) < 0\big) \int_0^\infty \prod_{j=1}^k (v - x_j)^{-(1/2+(1-H)/k)} dv \,.$$

Then $Q_t^{(k)} \in L^2(\mathbb{R}^k, \lambda_k)$ for all $t \geq 0$, and (8.25) and (8.26) hold. If $k \geq 2$, then this kernel works in the entire range $0 < H < 1$.

Like all other multiple Wiener–Itô integrals with respect to the Brownian motion, the H-SSSI processes defined by (8.27) with kernels in $L^2(\mathbb{R}^k, \lambda_k)$ satisfying the conditions (8.25) and (8.26) (such as given, for instance, in Examples 8.3.2, 8.3.3, and 8.3.4) have finite moments of all orders. They must share with the fractional Brownian motion its correlation function, but they are not Gaussian processes if $k \geq 2$. In fact, they can be viewed as polynomials of order k in the Brownian motion, and are said to be in the *kth Gaussian chaos*, using the terminology of Wiener (1938).

Suppose now that we have, for each $k \geq 1$, a kernel $(Q_t^{(k)})$ as above. Let (a_k) be a sequence of real numbers such that

$$\sum_{k=1}^\infty a_k^2 \, k! \int_{-\infty}^\infty \ldots \int_{-\infty}^\infty \widetilde{Q_t^{(k)}}(x_1, \ldots, x_k)^2 \, dx_1 \ldots dx_k < \infty \,,$$

where $\widetilde{Q_t^{(k)}}$ is the symmetrization of $Q_t^{(k)}$; see Section 10.6. Then the series defining the process

$$X(t) = \sum_{k=1}^\infty a_k \int_{-\infty}^\infty \ldots \int_{-\infty}^\infty Q_t^{(k)}(x_1, \ldots, x_k) \, B(dx_1) \ldots B(dx_k) \,, \tag{8.31}$$

$t \geq 0$, converges in L^2; see Proposition 10.6.1.

Proposition 8.3.5. *The process* $\mathbf{X}$ *in* (8.31) *is a well-defined second-order stochastic process. It is H-SSSI, and its increment process is a stationary mixing process.*

Proof. The fact that $\mathbf{X}$ has finite variance follows from the L^2 convergence of the series defining it. The rest of the statement can be established in a manner similar to the argument in the proof of Theorem 8.3.1. □

Of course, the process in (8.31) can no longer be viewed as a polynomial functional of the Brownian motion unless only finitely many of the numbers (a_k) are different from zero.

Even more versions of H-SSSI processes with finite variance can be introduced by replacing some of the repeated Brownian motions in (8.27) with independent Brownian motions or, more generally, with correlated Brownian motions. We will not pursue this point here.

8.4 SSSI Processes Without a Finite Variance

The most common SSSI processes without a finite variance are SαS SSSI processes. Examples in Section 3.6 provide three different classes of such processes: linear fractional symmetric stable motions, harmonizable fractional stable motion, and FBM-local time fractional stable motions. Recall that all these processes reduce to the fractional Brownian motion in the Gaussian case, but are different processes in the proper α-stable case, $0 < \alpha < 2$. More classes of SαS SSSI processes exist. One more such class is described in the next example.

Example 8.4.1. A *subordinator* is a Lévy motion of Example 3.1.2 whose sample paths are nondecreasing. The only stable processes among subordinators are those with stability index strictly smaller than 1, the Lévy measure concentrated on the positive half-line, and nonnegative shift. Strict stability requires the shift to vanish; see Bertoin (1996). For $0 < \beta < 1$, let $\big(S_\beta(t), t \geq 0\big)$ be a strictly β-stable subordinator. It has a Laplace transform of the form

$$Ee^{-\theta S_\beta(t)} = \exp\{-a^\beta t\theta^\beta\},\ \theta \geq 0,\ t \geq 0\,, \tag{8.32}$$

for a positive constant a. We define the inverse process of the subordinator by

$$M_\beta(t) = S_\beta^{\leftarrow}(t) = \inf\big\{u \geq 0 :\ S_\beta(u) \geq t\big\},\ t \geq 0\,, \tag{8.33}$$

and call it the Mittag-Leffler process. Since the stable subordinator is obviously self-similar with exponent $1/\beta$, it follows from Exercise 8.7.8 that the Mittag-Leffler process is self-similar with exponent β.

The marginal distributions of the Mittag-Leffler process are the Mittag-Leffler distributions. They have all moments finite. Moreover, if we standardize the stable subordinator by choosing the scaling constant in (8.32) to be $a = 1$, then

$$E\exp\{\theta M_\beta(t)\} = \sum_{n=0}^{\infty} \frac{(\theta t^\beta)^n}{\Gamma(1+n\beta)}, \quad \theta \in \mathbb{R}; \tag{8.34}$$

see Proposition 1(a) in Bingham (1971). The Mittag-Leffler process has a continuous and nondecreasing version; we will always choose such a version.

We now introduce the process we call the *β-Mittag-Leffler fractional SαS motion*. Once again, we start with a probability space $(\Omega', \mathcal{F}', P')$ supporting a Mittag-Leffler process $\big(M_\beta(t),\ t \geq 0\big)$. Let ν be a σ-finite measure on $[0, \infty)$ with

$$\nu(dx) = (1-\beta)x^{-\beta}\,dx\,.$$

Letting $M_{\alpha,\beta}$ be an SαS random measure on $\Omega' \times [0,\infty)$ with modified control measure $P' \times \nu$, we define

$$Y_{\alpha,\beta}(t) = \int_{\Omega' \times [0,\infty)} M_\beta\big((t-x)_+, \omega'\big)\, M_{\alpha,\beta}(d\omega',\, dx), \quad t \geq 0\,. \tag{8.35}$$

This process is well defined, since by the monotonicity of the Mittag-Leffler process and the fact that $M_\beta(0) = 0$, we have

$$\int_{[0,\infty)} \int_{\Omega'} M_\beta((t-x)_+, \omega')^\alpha P'(d\omega)\nu(dx) \leq t^{1-\beta} E' M_\beta(t)^\alpha < \infty\,.$$

Further, by the β-self-similarity of the process M_β, we have for every $k \geq 1$, $t_1 \ldots t_k \geq 0$, and $c > 0$, for all real $\theta_1, \ldots, \theta_k$,

$$\begin{aligned}
&E \exp\left\{ i \sum_{j=1}^k \theta_j Y_{\alpha,\beta}(ct_j) \right\} \\
&= \exp\left\{ -\int_0^\infty E' \Big| \sum_{j=1}^k \theta_j M_\beta((ct_j - x)_+) \Big|^\alpha (1-\beta) x^{-\beta} dx \right\} \\
&= \exp\left\{ -c^{1-\beta+\alpha\beta} \int_0^\infty E' \Big| \sum_{j=1}^k \theta_j M_\beta((t_j - y)_+) \Big|^\alpha (1-\beta) y^{-\beta} dy \right\} \\
&= E \exp\left\{ i \sum_{j=1}^k \theta_j c^{\beta + (1-\beta)/\alpha} Y_{\alpha,\beta}(t_j) \right\},
\end{aligned}$$

which shows that the β-Mittag-Leffler fractional SαS motion is self-similar with $H = \beta + (1-\beta)/\alpha$. This process also has stationary increments; see Exercise 8.7.9.

We will check next that the β-Mittag-Leffler fractional SαS noise, the increment process of the β-Mittag-Leffler fractional SαS motion, like the increment process the FBM-local time fractional stable motion, is generated by a conservative null flow. The increment process of the β-Mittag-Leffler fractional SαS motion has the representation

$$X_n = \int_{\Omega' \times [0,\infty)} \Big(M_\beta\big((n+1-x)_+, \omega'\big) - M_\beta\big((n-x)_+, \omega'\big) \Big)\, M_{\alpha,\beta}(d\omega',\, dx)\,,$$

$n = 0, 1, \ldots$. We begin by showing that for every $x > 0$, on a set of P' probability 1,

$$\limsup_{n\to\infty} M_\beta\big((n+1-x)_+\big) - M_\beta\big((n-x)_+\big) > 0\,. \tag{8.36}$$

To see this, note that a positive stable random variable has positive density on $(0, \infty)$, so the probability

$$P'\big(S_\beta(1) \in (a, a+1/2),\, S_\beta(2) - S_\beta(1) < 1/2\big)$$

is, as a function of $0 \le a \le 1$, bounded from below by some strictly positive number, say γ. Translated to the language of the Mittag-Leffler process, this implies that

$$P'\big(M_\beta(a+1) - M_\beta(a) \ge 1\big) \ge P'\big(M_\beta(a) \le 1,\, M_\beta(a+1) \ge 2\big) \ge \gamma \tag{8.37}$$

for each $0 \le a \le 1$. Fix now $x > 0$, and define recursively a sequence of random positive integers, a sequence of random numbers in $[0, 1]$, and a sequence of events (A_k) as follows. Let $n_0 = \lceil x \rceil$, $a_0 = \lceil x \rceil - x$, and

$$A_0 = \big\{M_\beta(a_0 + 1) - M_\beta(a_0) \ge 1\big\}\,.$$

For $k \ge 1$, we let

$$n_k = \lceil S_\beta\big(M_\beta(a_{k-1} + 1)\big) + x \rceil,\; a_k = n_k - x\,,$$
$$A_k = \big\{M_\beta(a_k + 1) - M_\beta(a_k) \ge 1\big\}\,,$$

with the stable subordinator and the Mittag-Leffler process related by (8.33). One can interpret a_k as follows. Take the value of the subordinator the first time the latter exceeds $a_{k-1}+1$. Then a_k is the distance from that value to the nearest from the right point of the type $m - x$, m an integer. The integer m is n_k. Clearly, $n_k \ge n_{k-1} + 1$, and by (8.37) and the strong Markov property of the stable subordinator,

$$P\Big(\cap_{k=n+1}^{n+m} A_n^c\Big) \le (1-\gamma)^m$$

for every $n, m \ge 1$. This implies (8.36). Now we can use Theorem 3.95 to conclude that the β-Mittag-Leffler fractional SαS noise is generated by a conservative flow. In order to prove that this process is generated by a null flow, choose $\beta < \tau < 1$, and let $w_n = n^{-\tau}$, $n = 0, 1, \ldots$. We will check that

$$E' \int_0^\infty \sum_{n=0}^{\infty} w_n \left[M_\beta\big((n+1-x)_+\big) - M_\beta\big((n-x)_+\big)\right] e^{-x}\, dx < \infty\,. \tag{8.38}$$

In that case, it will follow by Theorem 3.95 that the β-Mittag-Leffler fractional SαS noise is generated by a null flow if $\alpha = 1$, and hence by Proposition 3.6.8, the same is true for every $0 < \alpha < 2$. However, by the β-self-similarity of the Mittag-Leffler process, for some $c_\beta \in (0, \infty)$,

$$E' \sum_{n=0}^{\infty} w_n \left[M_\beta\big((n+1-x)_+\big) - M_\beta\big((n-x)_+\big) \right]$$
$$\le c_\beta \left(1 + \sum_{n=1}^{\infty} \big((n+1)^\beta - n^\beta\big)\big(n + \lceil x \rceil\big)^{-\tau} \right)$$
$$\le c_\beta \left(1 + \sum_{n=1}^{\infty} \big((n+1)^\beta - n^\beta\big) n^{-\tau} \right),$$

which proves (8.38) because of the choice of τ.

Example 8.4.1 as well as Examples 3.5.2, 3.5.4, and 3.5.5 demonstrate the large variety of different SαS SSSI processes. One can also construct SSSI processes without a finite variance that are not SαS. One can, for example, follow the recipe in (8.24) but choose the random factor A to have an infinite second moment. In that case, we could also replace the fractional Brownian motion in (8.24) by any of the H-SSSI processes defined previously, or any other H-SSSI process. In the remainder of this section, we will consider instead SSSI processes without a finite variance that are multiple integrals with respect to SαS random measures. We will follow a path similar to that used in Section 8.3 to construct processes with a finite variance.

Let $k \ge 1$ and $H > 0$. Suppose we are given a family of kernels $Q_t^{(k)} : \mathbb{R}^k \to \mathbb{R}$, $t \ge 0$, satisfying the increment property (8.25). The scaling property (8.26) is now replaced by the property

$$Q_{ct}^{(k)}(cx_1, \ldots, cx_k) = c^{H-k/\alpha} Q_t^{(k)}(x_1, \ldots, x_k) \tag{8.39}$$

for all $t \ge 0$, $c > 0$, and almost all $(x_1, \ldots, x_k)$. The integrability requirement $Q_t^{(k)} \in L^2(\mathbb{R}^k, \lambda_k)$ for every $t \ge 0$ has now to be replaced by the following requirement: there is a measurable function $\psi : S \to (0, \infty)$ with

$$\int_S \psi(s)^\alpha \, ds < \infty \tag{8.40}$$

such that

$$\int_S \cdots \int_S |Q_t(s_1, \ldots, s_k)|^\alpha \left(1 + \log_+ \frac{|Q_t(s_1, \ldots, s_k)|}{\psi(s_1) \ldots \psi(s_k)} \right)^{k-1} ds_1 \ldots ds_k < \infty \tag{8.41}$$

for each $t \ge 0$. See Section 10.6.

Let M be an SαS Lévy motion, which we view as an SαS random measure on $\mathbb{R}$. We define

$$X^{(k)}(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k) \, M(dx_1) \ldots M(dx_k), \ t \ge 0 \,. \tag{8.42}$$

Analogously to Theorem 8.3.1, we have the following result.

Theorem 8.4.2. *The process* $\mathbf{X}^{(k)}$ *in* (8.42) *is a well-defined H-SSSI process, and its increment process* $X_i = X^{(k)}(i) - X^{(k)}(i-1)$, $i = 1, 2, \ldots$, *is a stationary mixing process.*

Proof. Repeat the steps in the proof of Theorem 8.3.1, using Proposition 10.6.2 instead of Proposition 10.6.1. □

Example 8.4.3. The Surgailis kernel. Assume that $1 < \alpha < 2$ and $H \in (1/\alpha, 1)$. For $k \geq 1$, we let

$$Q_t^{(k)}(x_1, \ldots, x_k) = \int_0^t \prod_{j=1}^{k} \big((v - x_j)_+\big)^{-(1/\alpha + (1-H)/k)} dv \,, \tag{8.43}$$

$t \geq 0$. It is not difficult (but tedious) to check that this kernel satisfies (8.41), with ψ proportional to the Cauchy density, e.g., $\psi(x) = 1/(1 + x^2)$, $x \in \mathbb{R}$. The fact that (8.25) and (8.39) hold is elementary.

The resulting SSSI process, is, of course, a direct generalization of the SSSI process with finite variance obtained using the Taqqu kernel of Example 8.3.3.

Example 8.4.4. The finite-variance SSSI process obtained using the kernel in Example 8.3.4 can be generalized to the case of multiple integrals with respect to an SαS motion using

$$\begin{aligned} Q_t^{(k)}(x_1, \ldots, x_k) = {} & \mathbf{1}\big(\max(x_1, \ldots, x_k) < t\big) \int_t^\infty \prod_{j=1}^{k} (v - x_j)^{-(1/\alpha + (1-H)/k)} dv \\ & - \mathbf{1}\big(\max(x_1, \ldots, x_k) < 0\big) \int_0^\infty \prod_{j=1}^{k} (v - x_j)^{-(1/\alpha + (1-H)/k)} dv \,. \end{aligned} \tag{8.44}$$

As in the finite-variance case, for $k = 1$, the resulting process is well defined only if $H < 1/\alpha$, whereas for $k \geq 2$, the resulting process is well defined for all $0 < H < 1$. Once again, the condition (8.41) holds with ψ proportional to the Cauchy density, while (8.25) and (8.39) are obvious.

It is clear that one can construct other SSSI processes without a finite variance. For example, one can mimic the recipe (8.31), and construct SSSI processes in the "infinite SαS chaos." One chooses a sequence (a_k) of real numbers and a family $(Q_t^{(k)})$, $k = 1, 2, \ldots$, that satisfies, for each k, conditions (8.41), (8.25), and (8.39). With these ingredients, we can define

$$X(t) = \sum_{k=1}^{\infty} a_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k)\, M(dx_1) \ldots M(dx_k) \,, \tag{8.45}$$

assuming that the series converges in probability for each $t \geq 0$. There are no known necessary and sufficient conditions on the sequence (a_k) for this to happen, but it is clear that the convergence will hold if the sequence (a_k) converges to zero fast enough. Under these conditions, a statement analogous to Proposition 8.3.5 will hold, and the resulting process will also be H-SSSI, and its increment process will be a mixing process.

8.5 What Is in the Hurst Exponent? Ergodicity and Mixing

In the previous two sections, we have seen a large number of self-similar processes with stationary increments, with or without a finite variance. In particular, many of these processes share the same Hurst exponent H. We would like to understand the effect the Hurst exponent has on the memory of the increment processes. We will see in Chapter 9 that this effect is important, but limited. In this section, we take a preliminary step toward understanding this effect. Our goal here is very specific: given an H-SSSI process $\big(Y(t),\ t \geq 0\big)$, we would like to understand the effect of the Hurst exponent H on the presence or absence of ergodicity and/or mixing properties in the stationary increment process $\mathbf{X} = \big(Y(n+1) - Y(n),\ n = 0, 1, \ldots\big)$. This is an important question, because we can view absence of ergodicity as infinite memory, and, correspondingly, presence of ergodicity as finite memory; see Section 5.3. The property of mixing is, of course, a related and stronger property, which can be viewed as a link between ergodicity and the strong mixing properties. We will see that the link between the Hurst exponent H on the one hand and ergodicity or mixing of the increment process of an H-SSSI process on the other is not strong.

We begin with the example of the families of the H-SSSI processes that are defined as multiple integrals of the kernels satisfying conditions (8.25) and (8.26) in the finite-variance case, and conditions (8.25) and (8.39) in the infinite-variance case. Recall that in the finite-variance case, the integrals are with respect to the Brownian motion, as in (8.27), whereas in the infinite-variance case the integrals are with respect to the SαS Lévy motion, as in (8.42). In the finite-variance case, it follows from Theorem 8.3.1 that the increment process $\mathbf{X}$ is mixing, regardless of the value of $0 < H < 1$. The special case of a single integral, $d = 1$, reduces to the fractional Brownian motion, whose increment process is the fractional Gaussian noise, and its mixing follows simply from the fact that its covariance function asymptotically vanishes; see (1.16). In the infinite-variance case, the increment process is mixing as well; see Theorem 8.4.2. The special case of $d = 1$ reduces now to the linear fractional SαS motion of Example 3.5.2. Therefore, the increment process, linear fractional SαS noise, is always mixing, regardless of the value of $0 < H < 1$.

Not all H-SSSI processes have an increment process that is mixing. Since most of the examples we have of SSSI processes with infinite variance are symmetric α-stable processes, we need to understand when an SαS process is ergodic and/or mixing. We begin with a more general problem of characterizing ergodicity and

mixing of a stationary infinitely divisible process. Since we already know when a stationary Gaussian process is ergodic or mixing (Examples 2.2.18 and 2.2.8), we will concentrate on stationary infinitely divisible processes without a Gaussian component.

Let $\mathbf{X} = \big(X_n, n \in \mathbb{Z}\big)$ be a stationary infinitely divisible process with Lévy measure μ (on $\mathbb{R}^{\mathbb{Z}}$) and no Gaussian component.

Theorem 8.5.1. *(i) The process* $\mathbf{X}$ *is ergodic if and only if it is weakly mixing, and this happens if and only if the Lévy measure of the component* $\mathbf{X}^{(p)}$ *of* $\mathbf{X}$*, generated by a positive flow, vanishes.*

(ii) Assume that the one-dimensional Lévy measure of X_0 *does not have an atom of the form* $2\pi k$*,* $k \in \mathbb{Z}$*. Then the process* $\mathbf{X}$ *is mixing if and only if*

$$\lim_{n\to\infty} Ee^{i(X_n - X_0)} = Ee^{iX_0}Ee^{-iX_0} \left(= \left|Ee^{iX_0}\right|^2\right). \tag{8.46}$$

(iii) If the process $\mathbf{X}$ *is generated by a dissipative flow, then it is mixing.*

Proof. We begin with part (ii) of the theorem. Observe that if a process is mixing, then Theorem 2.2.7 says that (8.46) must hold regardless of infinite divisibility of the process and of any of the features of its Lévy measure. We need to prove that under our assumptions, (8.46), in turn, implies mixing. We claim that (8.46) implies that for every complex-valued random variable $Y \in L^2$ (of the probability space on which the process is defined),

$$\lim_{n\to\infty} E\big(e^{iX_n}\bar{Y}\big) = Ee^{iX_0}E\bar{Y}\,. \tag{8.47}$$

To see this, note that by (8.46) and stationarity, (8.47) holds for every $Y \in H$, the closure in L^2 of the linear space spanned by the constant 1 and e^{iX_k}, $k \in \mathbb{Z}$. Therefore, it is enough to prove that (8.47) also holds for every $Y \in H^{\perp}$, the orthogonal complement to H. However, such a Y is orthogonal to constants and hence has mean zero. Since it is also orthogonal to each e^{iX_n}, the claim (8.47) is a trivial statement that the limit of a sequence of zeros is zero.

We now use (8.47) with $Y = e^{-iX_0}$ to conclude that we also have

$$\lim_{n\to\infty} Ee^{i(X_n + X_0)} = \big(Ee^{iX_0}\big)^2\,. \tag{8.48}$$

Armed with (8.46) and (8.48), we are ready to check that the conditions for mixing in Theorem 2.2.7 hold. Fix $k = 1, 2, \ldots$. By stationarity, the sequence of the laws of the vectors on the left-hand side of (2.14) is clearly tight in $\mathbb{R}^{2k}$, so we need to prove only that every subsequential weak limit of that sequence coincides with the law of the random vector on the right-hand side of (2.14). Let, therefore, $\{n_m\}$ be an increasing sequence of positive integers, and suppose that

$$\big(X_1, \ldots, X_k, X_{n_m+1}, \ldots, X_{n_m+k}\big) \Rightarrow \big(X_1, \ldots, X_k, Y_1, \ldots, Y_k\big) \tag{8.49}$$

as $m \to \infty$, and we need to prove that the vectors $(X_1, \dots, X_k)$ and $(Y_1, \dots, Y_k)$ on the right-hand side of (8.49) are independent. Since the family of infinitely divisible distributions is closed under weak convergence, the $2k$-dimensional random vector on the right-hand side of (8.49) is infinitely divisible. Furthermore, it follows by stationarity, (8.46), and (8.48) that for every $j_1, j_2 = 1, \dots, k$, the random variables on the right-hand side of (8.49) satisfy

$$Ee^{i(X_{j_1} - Y_{j_2})} = Ee^{iX_{j_1}} Ee^{-iY_{j_2}}, \quad Ee^{i(X_{j_1} + Y_{j_2})} = Ee^{iX_{j_1}} Ee^{iY_{j_2}} .$$

We conclude that X_{j_1} and Y_{j_2} are independent; see Exercise 3.8.4. Since this is true for any two components of the vectors $(X_1, \dots, X_k)$ and $(Y_1, \dots, Y_k)$, we conclude that the vectors themselves are independent; see Exercise 3.8.3. By Theorem 2.2.7, we conclude that the process $\mathbf{X}$ is mixing.

We proceed to prove part (iii) of the theorem. We will prove that if $\mathbf{X}$ is generated by a dissipative flow, then condition (8.46) is satisfied. Observe that by part (ii) of the theorem, this will suffice for the statement if the one-dimensional Lévy measure of X_0 does not have an atom of the form $2\pi k$, $k \in \mathbb{Z}$. In the general case, since that Lévy measure has at most countably many atoms, there exists $c > 0$ such that there are no atoms of the form $2c\pi k$, $k \in \mathbb{Z}$. Then the one-dimensional Lévy measure of the stationary process $\mathbf{Y} = \mathbf{X}/c$ does not charge the set $\{2\pi k, k \in \mathbb{Z}\}$. Obviously, the process $\mathbf{Y}$ is generated by a dissipative flow if the process $\mathbf{X}$ is. Therefore, we conclude that the process $\mathbf{Y}$ is mixing, which implies that so is the process $\mathbf{X}$.

It remains to prove that (8.46) holds if $\mathbf{X}$ is generated by a dissipative flow, which is the same as proving that

$$\begin{aligned} &\int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{i(x_n - x_0)} - 1 - i\big([\![x_n]\!] - [\![x_0]\!]\big)\right) \mu(d\mathbf{x}) \\ &- \int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{ix_n} - 1 - i[\![x_n]\!]\right) \mu(d\mathbf{x}) - \int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{-ix_0} - 1 + i[\![x_0]\!]\right) \mu(d\mathbf{x}) \\ &= \int_{\mathbb{R}^{\mathbb{Z}}} \big(e^{ix_n} - 1\big)\big(e^{-ix_0} - 1\big)\, \mu(d\mathbf{x}) \to 0 \end{aligned} \tag{8.50}$$

as $n \to \infty$. To this end, let $\varepsilon > 0$ and set $A = \big\{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : |x_0| > \varepsilon\big\}$. Note that $\mu(A) < \infty$, so by Theorem 2.40, $\mathbf{1}_A \circ \theta^n \to 0$ as $n \to \infty$ m-a.e. Therefore,

$$\mu\big(\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : |x_0| > \varepsilon,\ |x_n| > \varepsilon\big) \to 0 \ \text{ as } n \to \infty. \tag{8.51}$$

We can now use the obvious bound $|e^{ix} - 1| \le 2|[\![x]\!]|$ for a real x and the Cauchy–Schwarz inequality to bound

$$\left| \int_{\mathbb{R}^{\mathbb{Z}}} \big(e^{ix_n} - 1\big)\big(e^{-ix_0} - 1\big)\, \mu(d\mathbf{x}) \right| \tag{8.52}$$

$$\leq 4 \int_{\mathbb{R}^{\mathbb{Z}}} |[\![x_0]\!]||[\![x_n]\!]| \, \mu(d\mathbf{x})$$

$$\leq 8 \left(\int_{\mathbb{R}^{\mathbb{Z}}} [\![x_0]\!]|^2 \, \mu(d\mathbf{x}) \right)^{1/2} \left(\int_{\mathbb{R}^{\mathbb{Z}}} [\![x_0]\!]|^2 \mathbf{1}_{|x_0| \leq \varepsilon} \, \mu(d\mathbf{x}) \right)^{1/2}$$

$$+4\mu\big(\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} : \, |x_0| > \varepsilon, \, |x_n| > \varepsilon\big) \, .$$

It follows from (8.51) that for every $\varepsilon > 0$,

$$\limsup_{n \to \infty} \left| \int_{\mathbb{R}^{\mathbb{Z}}} \big(e^{ix_n} - 1\big)\big(e^{-ix_0} - 1\big) \, \mu(d\mathbf{x}) \right|$$

$$\leq 8 \left(\int_{\mathbb{R}^{\mathbb{Z}}} [\![x_0]\!]|^2 \, \mu(d\mathbf{x}) \right)^{1/2} \left(\int_{\mathbb{R}^{\mathbb{Z}}} [\![x_0]\!]|^2 \mathbf{1}_{|x_0| \leq \varepsilon} \, \mu(d\mathbf{x}) \right)^{1/2} ,$$

and we obtain (8.50) by letting $\varepsilon \to 0$.

We now turn to proving part (i) of the theorem. We first check the claimed equivalence of ergodicity and weak mixing. We assume initially that the one-dimensional Lévy measure of X_0 does not have an atom of the form $2\pi k$, $k \in \mathbb{Z}$.

Suppose that the process $\mathbf{X}$ is ergodic. It follows from (2.8) that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} e^{iX_j} \to E e^{iX_0} \ \text{ a.s.},$$

and hence by the bounded convergence theorem,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} E e^{i(X_j - X_0)} \to \left| E e^{iX_0} \right|^2 .$$

The computation in (8.50) shows that we can rewrite this statement as

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \exp \left\{ \int_{\mathbb{R}^{\mathbb{Z}}} \big(e^{ix_n} - 1\big)\big(e^{-ix_0} - 1\big) \, \mu(d\mathbf{x}) \right\} = 1 \, . \tag{8.53}$$

Denote

$$h_n = \int_{\mathbb{R}^{\mathbb{Z}}} \big(e^{ix_n} - 1\big)\big(e^{-ix_0} - 1\big) \, \mu(d\mathbf{x}), \ n \in \mathbb{Z} \, .$$

We claim that the complex-valued sequence (h_n) is nonnegative definite. Indeed, by stationarity, for every $n = 1, 2, \ldots$ and complex numbers $a_1, \ldots, a_n$,

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j\bar{a}_k h_{j-k} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_j\bar{a}_k \int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{ix_j}-1\right)\left(e^{-ix_k}-1\right)\mu(d\mathbf{x})$$

$$= \int_{\mathbb{R}^{\mathbb{Z}}} \left|\sum_{j=1}^{n} a_j\left(e^{ix_j}-1\right)\right|^2 \mu(d\mathbf{x}) \geq 0\,.$$

By Theorem 10.1.2, there is a finite measure μ on $(-\pi,\pi]$ such that

$$h_n = \int_{(-\pi,\pi]} e^{inx}\,\mu(dx),\ n\in\mathbb{Z}\,.$$

Define another finite measure on $(-\pi,\pi]$ by

$$\exp(\mu) = \sum_{k=0}^{\infty}\frac{1}{k!}\mu^{*k,f}\,,$$

where $\mu^{*k,f}$ is the folded kth convolution power of μ, $k = 1,2,\ldots$, defined in Theorem 6.3.4. Since

$$\int_{(-\pi,\pi]} e^{inx}\,\exp(\mu)(dx) = \exp\left\{\int_{(-\pi,\pi]} e^{inx}\,\mu(dx)\right\}$$

for each $n\in\mathbb{Z}$, we can rewrite (8.53) as

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\int_{(-\pi,\pi]} e^{ijx}\,\exp(\mu)(dx) = 1\,.$$

By Exercise 6.5.2, this is the same as saying that the mass at zero of the measure $\exp(\mu)$ is equal to 1. Since

$$\exp(\mu)(\{0\}) \geq \exp\{\mu(\{0\})\}\,,$$

we conclude that $\mu(\{0\}) = 0$. By Exercise 6.5.2, this means that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\int_{\mathbb{R}^{\mathbb{Z}}}\left(e^{ix_j}-1\right)\left(e^{-ix_0}-1\right)\mu(d\mathbf{x}) = 0\,. \tag{8.54}$$

Concentrating on convergence of the real parts, we obtain

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\left(\int_{\mathbb{R}^{\mathbb{Z}}}(1-\cos x_j)(1-\cos x_0)\,\mu(d\mathbf{x}) + \int_{\mathbb{R}^{\mathbb{Z}}}\sin x_j\sin x_0\,\mu(d\mathbf{x})\right) = 0\,.$$

The stationarity of the process once again tells us that each of the two integrals above defines a nonnegative definite sequence, and appealing twice to Exercise 6.5.2, we conclude that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \int_{\mathbb{R}^{\mathbb{Z}}} (1-\cos x_j)(1-\cos x_0)\, \mu(d\mathbf{x}) = 0\,.$$

Since this is a statement about averages of a nonnegative sequence, it follows from Lemma 2.2.12 that

$$\int_{\mathbb{R}^{\mathbb{Z}}} (1-\cos x_n)(1-\cos x_0)\, \mu(d\mathbf{x}) \to 0 \ \text{ as } n \to \infty \text{ in density}, \tag{8.55}$$

as in Proposition 2.2.13. We will prove that (8.55) implies that

$$(X_1, X_{n+1}) \Rightarrow (X_1, Y_1) \ \text{ in density}, \tag{8.56}$$

where Y_1 is an independent copy of X_1. Before doing so, we show that (8.56) implies weak mixing of the process. To this end, we use Theorem 2.2.17. Let $k = 1, 2, \ldots$. It follows from (8.56) that there is a set K_k of positive integers of density zero such that

$$(X_i, X_{n+j}) \Rightarrow (X_1, Y_1) \ \text{ in density} \tag{8.57}$$

for each $i, j = 1, \ldots, k$. We claim that convergence in (2.24) holds outside of the set K_k. Indeed, suppose that this convergence statement is false. Then the tightness of the sequence of the laws of the random vectors on the left-hand side of (2.24) implies that there is an increasing sequence $\{n_m\}$ of positive integers *outside of* K_k such that

$$\big(X_1, \ldots, X_k, X_{n_m+1}, \ldots, X_{n_m+k}\big) \Rightarrow \big(X_1, \ldots, X_k, Y_1, \ldots, Y_k\big)$$

as $m \to \infty$, and the random vectors $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_k)$ on the right-hand side are not independent. This is, however, impossible, as we noticed earlier: these two random vectors are jointly infinitely divisible, and (8.57) shows that X_i and Y_j are independent for each $i, j = 1, \ldots, k$. By Exercise 3.8.3, the random vectors $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_k)$ on the right-hand side must be independent. The obtained contradiction proves the claimed weak mixing, and so it remains only to prove that (8.55) implies (8.56).

Let K be a set of density zero such that (8.55) holds outside of that set. Then for every $\varepsilon > 0$, we have

$$\mu\big\{\mathbf{x} :\ 1-\cos x_0 > \varepsilon \text{ and } 1-\cos x_n > \varepsilon\big\} \to 0 \ \text{ as } n \to \infty \text{ outside of } K$$

and hence also

$$\mu\big\{\mathbf{x} :\ |\sin x_0| > \varepsilon \text{ and } |\sin x_n| > \varepsilon\big\} \to 0 \ \text{ as } n \to \infty \text{ outside of } K.$$

Since these two statements hold for every $\varepsilon > 0$, we conclude that

$$\int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{ix_n} - 1\right)\left(e^{-ix_0} - 1\right) \mu(d\mathbf{x}) \to 0 \text{ as } n \to \infty \text{ outside of } K.$$

By (8.50), this proves that (8.46) holds as $n \to \infty$ outside of K. While proving part (ii) of the theorem, we saw that this (together with the assumption that the one-dimensional Lévy measure of X_0 does not charge the set $\{2\pi k, k \in \mathbb{Z}\}$) implies (8.56).

We have therefore established equivalence of ergodicity and weak mixing under the assumption that the one-dimensional Lévy measure of X_0 has no atoms of the form $2\pi k$, $k \in \mathbb{Z}$. In the general case, we proceed in the same manner as in the proof of part (iii): Choose $c > 0$ such that the one-dimensional Lévy measure of X_0 has no atoms of the form $2c\pi k$, $k \in \mathbb{Z}$. If the process $\mathbf{X}$ is ergodic, then so is the process $\mathbf{Y} = \mathbf{X}/c$, which is then weakly mixing. This, in turn, implies weak mixing of the process $\mathbf{X}$.

We still have to prove that ergodicity (and weak mixing) are equivalent to absence of the component $\mathbf{X}^{(p)}$ of $\mathbf{X}$, generated by a positive flow. Suppose first that the Lévy measure μ_p of $\mathbf{X}^{(p)}$ does not vanish. Then there exists a probability measure ν on $\mathbb{R}^{\mathbb{Z}}$ equivalent to μ_p and invariant under the left shift. The Radon–Nikodym derivative $d\mu_p/d\nu$ is then shift-invariant, and there is $0 < a < \infty$ such that $\nu(A_a) > 0$, where

$$A_a = \left\{\mathbf{x} : \frac{d\mu_p}{d\nu}(\mathbf{x}) \le a\right\} .$$

Let $\mu_{p,a}$ be the restriction of μ_p to the invariant set A_a. Then $\mathbf{X}^{(p)} \stackrel{d}{=} \mathbf{X}^{(p,a)} + \tilde{\mathbf{X}}^{(p,a)}$, where $\mathbf{X}^{(p,a)}$ and $\tilde{\mathbf{X}}^{(p,a)}$ are independent stationary infinitely divisible processes, with $\mathbf{X}^{(p,a)}$ having Lévy measure $\mu_{p,a}$, and

$$E \exp\left\{ i \sum_{n\in\mathbb{Z}} \theta(n) X_n^{(p,a)} \right\} = \exp\left\{ \int_{\mathbb{R}^{\mathbb{Z}}} \left(e^{i(\boldsymbol{\theta},\mathbf{x})} - 1 - i\big(\boldsymbol{\theta}, [\![\mathbf{x}]\!]\big) \right) \mu_{p,a}(d\mathbf{x}) \right\}$$

for every $\boldsymbol{\theta} \in \mathbb{R}^{(\mathbb{Z})}$. Since $\mu_p(A_a) \le a < \infty$, $\mu_{p,a}$ is a finite measure; hence the process $\mathbf{X}^{(p,a)}$ can be represented, in law, as a shifted compound Poisson process,

$$X_n^{(p,a)} = \sum_{i=1}^{N} Y_{i,n} + b, \ n \in \mathbb{Z}, \tag{8.58}$$

where N is a Poisson random variable with mean $\lambda = \mu_{p,a}(\mathbb{R}^{\mathbb{Z}})$, independent of the sequence of i.i.d. stationary processes $\big(Y_{i,n}, n \in \mathbb{Z}\big)$, $i = 1, 2, \ldots$, with the common law $\lambda^{-1}\mu_{p,a}$. Here $b = -\lambda E[\![Y_{1,0}]\!]$; see Exercise 3.8.9. It follows from (8.58) that the process $\mathbf{X}^{(p,a)}$ can be represented as a nontrivial mixture of stationary processes. Therefore, so can be the process $\mathbf{X}^{(p)}$ (it is the sum of the process $\mathbf{X}^{(p,a)}$

and an independent stationary process), and hence so can be the process $\mathbf{X}$. By Proposition 2.1.6, the process $\mathbf{X}$ is not ergodic.

In the final stage of the proof, suppose that the left shift is null with respect to the Lévy measure μ of the process. We will prove that the process $\mathbf{X}$ is ergodic. We may assume that the one-dimensional Lévy measure of X_0 does not charge the set $\{2\pi k,\ k \in \mathbb{Z}\}$. In this case, to prove ergodicity it is enough to show that (8.54) holds, since we have already established that (8.54) implies weak mixing of the process. By (8.52), it is enough to prove that for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mu\big(\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\ |x_0| > \varepsilon,\ |x_j| > \varepsilon\big) = 0\,. \tag{8.59}$$

Let $A = \{\mathbf{x} \in \mathbb{R}^{\mathbb{Z}} :\ |x_0| > \varepsilon\}$. Then $\mu(A) < \infty$. For $\theta > 0$, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{j=1}^{n} \mu\left(\mathbf{x} \in A :\ |x_j| > \varepsilon\right) = \int_A \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\left(\mathbf{x} \in A :\ |x_j| > \varepsilon\right)\, \mu(d\mathbf{x}) \\
&= \int_A \left(\frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\left(\mathbf{x} \in A :\ |x_j| > \varepsilon\right)\right) \mathbf{1}\left(\frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) > \theta\right)\, \mu(d\mathbf{x}) \\
&+ \int_A \left(\frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\left(\mathbf{x} \in A :\ |x_j| > \varepsilon\right)\right) \mathbf{1}\left(\frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) \le \theta\right)\, \mu(d\mathbf{x}) \\
&\le \mu\left(\mathbf{x} \in A :\ \frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) > \theta\right) \\
&+ (\varepsilon^2 \wedge 1)^{-1} \int_A \frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1)\, \mathbf{1}\left(\frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) \le \theta\right)\, \mu(d\mathbf{x}) \\
&\le \mu\left(\mathbf{x} \in A :\ \frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) > \theta\right) + \frac{\theta}{\varepsilon^2 \wedge 1} \mu(A).
\end{aligned} \tag{8.60}$$

Since by the assumption, the left shift is null with respect to the Lévy measure μ, and the function $f(\mathbf{x}) = x_0^2 \wedge 1$ is in $L^1(\mu)$ by (3.5), Exercise 2.6.13 tells us that

$$\frac{1}{n} \sum_{j=1}^{n} (x_j^2 \wedge 1) \to 0 \ \ \mu\text{-a.e.}$$

as $n \to \infty$. Since the set A has finite measure, we conclude that the first term in (8.60) vanishes as $n \to \infty$ for every $\theta > 0$. Since we can take θ as small as

we wish, we can make the second term in (8.60) as small as we like, and so (8.59) follows. This completes the proof. □

Applying the results of Theorem 8.5.1 to different classes of SαS self-similar processes with stationary increments, we obtain some immediate conclusions:

- The linear fractional SαS noise is generated by a dissipative flow (Example 3.6.9) and hence is a mixing process, regardless of the value of the Hurst exponent.
- The harmonizable fractional SαS noise is generated by a conservative flow (Example 3.6.10) and hence is not an ergodic process, regardless of the value of the Hurst exponent.
- The FBM-H-local time fractional SαS noise and the β-Mittag-Leffler fractional SαS noise are generated by conservative null flows (Examples 3.6.11 and 8.4.1) and hence are ergodic and weakly mixing processes, regardless of the value of the Hurst exponent (which belongs, in these examples, to restricted ranges).

From this discussion, we see that at least as far as ergodicity and mixing of the increment process of a stable self-similar process with stationary increments are concerned, very little is determined by the value of the Hurst exponent, and the type of the flow that generates the increment process is of a major importance.

8.6 Comments on Chapter 8

Comments on Section 8.1

The original Lamperti's theorem with continuous scaling is in Lamperti (1962); it allows an additional centering. An even more general situation is tackled in Laha and Rohatgi (1982), where the original process $\mathbf{U}$ is vector-valued, and the scaling is by a continuous parameter family of linear operators. The discrete scaling is in Weissman (1975), with the conclusion stated in terms of the one-dimensional marginal distributions.

Comments on Section 8.2

The foundations of the systematic study of self-similar processes, particularly those with stationary increments, were laid by W. Vervaat and his collaborators; two of the important papers are O'Brien and Vervaat (1983) and Vervaat (1985).

The statement of Lemma 8.2.11 had been formulated as a problem by D. Heath and W. Vervaat, and a solution was published in Smit (1983).

Comments on Section 8.3

The process (8.27) corresponding to the kernel in Example 8.3.2 is mentioned in Mori and Oodaira (1986), following a similar process introduced in Rosenblatt (1979). The process corresponding to the kernel in Example 8.3.3 was introduced in Taqqu (1978); it appeared as a limit in a "noncentral limit theorem" in Taqqu (1979) and, in a more general situation, in Surgailis (1981a); see also Avram and Taqqu (1987).

Comments on Section 8.4

The β-Mittag-Leffler fractional SαS motion was introduced in Owada and Samorodnitsky (2015a), where it is also shown that in the range $0 < \beta \leq 1/2$, this process has the same law as the $\hat{\beta}$-stable local time fractional SαS motion introduced and studied in Dombry and Guillotin-Plantard (2009), with $\hat{\beta} = (1-\beta)^{-1}$.

The multiple integral process of Example 8.4.3 was introduced in Surgailis (1981b). In the case $k = 2$, this process appears as a limit in the "noncentral limit theorem" setting, as shown in Astrauskas (1983).

Comments on Section 8.5

The equivalence of ergodicity and weak mixing for general stationary infinitely divisible processes was proved in Rosiński and Żak (1997). Many ideas in Theorem 8.5.1 are taken from Rosiński and Żak (1996), Rosiński and Żak (1997), and Samorodnitsky (2005). A very general discussion of the relations between the properties of stationary infinitely divisible processes and their Lévy measures is in Roy (2007).

8.7 Exercises to Chapter 8

Exercise 8.7.1. *Prove Theorem 8.1.6.*

Exercise 8.7.2. (a) *Prove that if X and Y are defined on the same probability space, $X \stackrel{d}{=} Y$ and $X \geq Y$ a.s., then $X = Y$ a.s.*

(b) *Let $\mathbf{X}$ be a stochastic process with stationary max-increments; see Definition 8.1.9. Prove that for every $s \geq 0$, $X(s) \vee X^{(s)}(0) = X(s)$ a.s.*

(c) *Prove that a stochastic process $\mathbf{X}$ with stationary max-increments satisfies $X(t_2) \geq X(t_1)$ a.s. for every $0 \leq t_1 \leq t_2$.*

Exercise 8.7.3. *Complete the missing step at the beginning of the proof of Proposition 8.1.10 and show that with $A(\lambda) = a_n$ if $n - 1 \leq \lambda < n$, $n = 1, 2, \ldots$, the process $\mathbf{U}$ satisfies the continuous scaling assumption (8.11).*

Exercise 8.7.4. *Use Exercise 8.7.2 to extend Lemma 8.2.2, Proposition 8.2.4, and Lemma 8.2.5 from the case of self-similar processes with stationary increments to the case of self-similar processes with stationary max-increments. Show that in the latter case, an a priori assumption of measurability in Proposition 8.2.4 is not required.*

Exercise 8.7.5. *Let $\big(X(t),\, t \geq 0\big)$ be H-self-similar with $H > 0$ and have stationary increments. Let*

$$p_0 = P\big(X(\tau) = 0 \ \textit{for all rational } \tau\big)\,.$$

Assume that $p_0 < 1$. Show that there is an H-self-similar process with stationary increments $\big(Y(t),\, t \geq 0\big)$ such that $P(Y(t) = 0) = 0$ for all $t > 0$ such that

$$\mathbf{X} \stackrel{d}{=} \begin{cases} \mathbf{N} \text{ with probability } p_0, \\ \mathbf{Y} \text{ with probability } 1 - p_0, \end{cases}$$

where $\mathbf{N}$ is the null (identically zero) process.

Exercise 8.7.6. *Use Exercise 8.7.5 to prove the following version of Proposition 8.2.6. Let $\big(X(t),\, t \geq 0\big)$ be an H-self-similar process with stationary max-increments such that $P(X(1) \neq 0) > 0$. Suppose that $E|X(1)|^\gamma < \infty$ for some $\gamma > 0$. Then $H < 1/\gamma$.*

Exercise 8.7.7. *Show that the covariance function of the fractional Brownian motion given in (8.23) remains a valid covariance function if one formally sets $H = 1$ in (8.23), and that in this case, it is the covariance function of the straight-line process $X(t) = tX$, $t \geq 0$, with $EX^2 < \infty$. This is, of course, expected due to Proposition 8.2.10.*

Exercise 8.7.8. *Let $\big(X(t),\, t \geq 0\big)$ be a right-continuous H-self-similar process with $H > 0$ such that $X(1) > 0$ a.s. Show that*

$$Y(t) = \inf\big\{u \geq 0 :\, X(u) \geq t\big\},\ t \geq 0,$$

is a well-defined stochastic process that is self-similar with exponent $1/H$.

Exercise 8.7.9. *Prove that the β-Mittag-Leffler fractional SαS motion of Example 8.4.1 has stationary increments. The strong Markov property of the stable subordinator is helpful for that purpose as well as the known distribution of the overshoot $\delta_r = S_\beta\big(M_\beta(r)\big) - r$ of the level $r > 0$ by the β-stable subordinator $\big(S_\beta(t), t \geq 0\big)$ related to $\big(M_\beta(t), t \geq 0\big)$ by (8.33). The law of δ_r is given by*

$$P(\delta_r \in dx) = \frac{\sin \beta\pi}{\pi} r^\beta (r + x)^{-1} x^{-\beta}\, dx,\ x > 0\,;$$

see Exercise 5.6 in Kyprianou (2006).

Chapter 9
Long-Range Dependence as a Phase Transition

9.1 Why Phase Transitions?

Long-range dependence in a stationary process has been understood as corresponding to a particular second-order behavior, to a particular range of the Hurst parameter, or of fractional integration. All these points of view on long-range dependence really describe the situation in which the stationary process under consideration is very different from "the norm," and "normal behavior" is understood as the behavior of the i.i.d. sequence with the same marginal distribution as the stationary process being considered. A unified point of view would therefore directly regard the phenomenon of long-range dependence as a phase transition, and we proceed to describe such a direct approach.

Suppose that we are considering a family of laws of stationary stochastic processes on $\mathbb{Z}$; this is a family of shift-invariant probability measures $\left(P_\theta,\ \theta \in \Theta\right)$ on $\mathbb{R}^{\mathbb{Z}}$. That is, for each value of the parameter $\theta \in \Theta$, the probability measure P_θ describes the finite-dimensional distributions of a stationary stochastic process $\mathbf{X} = \left(X_n,\ n \in \mathbb{Z}\right)$. In general, as the parameter θ varies, both the one-dimensional marginal distribution of the process $\mathbf{X}$ changes, and so does the memory in the process $\mathbf{X}$. There are two basic assumptions.

1. We measure the memory in the process $\mathbf{X}$ through the behavior of a measurable functional $\phi : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\infty}$ under the probability measure P_θ.
2. The one-dimensional marginal distributions of the process $\mathbf{X}$ do not change significantly as θ varies.

Some examples of the functional ϕ in assumption 1 are the sequence of the partial sums $\phi_n(\mathbf{x}) = \sum_{j=1}^{n} x_j$ for $n \geq 1$ and the sequence of the partial maxima, for which $\phi_n(\mathbf{x}) = \max_{j=1}^{n} x_j$ for $n \geq 1$. Among other possibilities, we may be interested in weak limits associated with the functional ϕ, or in large deviations associated with this functional.

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_9

As far as assumption 2 is concerned, as θ varies, we may allow, for example, a change in scale or another parameter not relevant to the behavior of the functional ϕ we are considering. Typically, we do not want to allow a serious change in the marginal tails of the process $\mathbf{X}$. For example, we do not want the variance to be finite for θ in one part of the set Θ and infinite in another part. For example, $\left(P_\theta,\ \theta \in \Theta\right)$ might be the family of laws of zero-mean unit-variance stationary Gaussian processes, in which case a change in the parameter θ will change only the correlation function of the process. Alternatively, $\left(P_\theta,\ \theta \in \Theta\right)$ might be the family of laws of infinite moving-average processes of Section 1.4. In this case, a change in the parameter θ corresponds to a change in the sequence of coefficients (φ_n).

We denote by Θ_0 the subset of the parameter space corresponding to the choices of the parameter θ such that under P_θ, the process $\mathbf{X}$ is a sequence of i.i.d. random variables; the set Θ_0 may be a singleton, as is the case in the example of the family of laws of zero-mean unit-variance stationary Gaussian processes parametrized by their correlation function. Suppose that there is a partition $\Theta = \Theta_1 \cup \Theta_2$ of the parameter space into disjoint parts, with $\Theta_0 \subset \Theta_1$, such that the behavior of the functional ϕ undergoes a major change as the parameter θ crosses the boundary between Θ_1 and Θ_2. This often means that as long as the parameter θ stays within Θ_1, the behavior of the functional ϕ does not change much, and it remains similar to the behavior of ϕ in the i.i.d. case, $\theta \in \Theta_0$, perhaps "up to a multiplicative constant." Once the parameter θ crosses the boundary into Θ_2, there is a change "in the order of magnitude" in the behavior of the functional ϕ. Moreover, there continue to be significant changes as the parameter θ moves within Θ_2.

In such a situation, we view the part Θ_1 of the parameter space as corresponding to short-memory models, and the part Θ_2 of the parameter space as corresponding to long-memory models. That is, if the law of a stationary process $\mathbf{X}$ is P_θ with $\theta \in \Theta_2$, we say that the process $\mathbf{X}$ has long-range dependence, whereas if the law of $\mathbf{X}$ is P_θ with $\theta \in \Theta_1$, then the process does not have long-range dependence. From this point of view, the boundary between Θ_1 and Θ_2 is the boundary between short and long memory, and so the appearance of long-range dependence is a phase transition.

It is easy to criticize the point of view on long-range dependence we are offering. We have clearly avoided rigorous definitions. Furthermore, we have tied the notion of long-range dependence to a particular functional of the sample paths of the process, and perhaps also to a particular aspect of the behavior of that functional. Finally, it might be possible to identify additional significant boundaries within the set Θ_2 and hence additional significant phase transitions.

One defense of viewing long-range dependence as a phase transition is that some ambiguity may be inevitable. Long-range dependence is really an entire group of phenomena, and it may well be, in some cases, a hierarchical system. The reason it has proven to be so difficult to define long-range dependence is that one has tried to give a single definition to what is not a single phenomenon.

In the remainder of this chapter, we point out some of the critical boundaries indicating long-range dependence in a number of interesting situations.

9.2 Phase Transitions in Partial Sums

In this section, we investigate long-range dependence from the point of view of possible phase transitions related to the behavior of the partial sums of a stationary process. Let $\mathbf{X} = (X_n,\ n \in \mathbb{Z})$ be a stationary stochastic process. As in Chapter 5, we use the notation

$$S_n = X_1 + \ldots + X_n,\ n = 0, 1, 2, \ldots, \tag{9.1}$$

for the sequence of partial sums of the process. We will introduce certain assumptions on the one-dimensional marginal distribution of X (a generic representative of the stationary process) and investigate when the behavior of the partial sums is *qualitatively different* from the behavior of the corresponding partial sums of an i.i.d. sequence with the same (or similar) marginal distribution. We will consider two types of assumptions on the marginal distribution of the process.

Assumption 9.2.1. *Either*

- *X has a finite variance, or*
- *X has balanced regularly varying tails with exponent $\alpha \in (0, 2)$.*

We will study not only the behavior of the partial sums S_n in (9.1) as $n \to \infty$, but even more informatively, the partial sum processes

$$\mathbf{S}_n = \left(S_{[nt]},\ t \geq 0\right). \tag{9.2}$$

Under each of the situations in Assumption 9.2.1, we will investigate the order of magnitude of the partial sums in (9.1). We will also investigate possible limiting processes $\mathbf{Y}$ in weak limit results of the type

$$\frac{1}{a_n}\left(\mathbf{S}_n - \mathbf{b}_n\right) \Rightarrow \mathbf{Y} \tag{9.3}$$

in some "reasonable topology" and with a nondegenerate limit $\mathbf{Y}$, for some sequence of real-valued functions $(\mathbf{b}_n)$ and a sequence $a_n \to \infty$. Recall from Section 8.1 that we expect the limiting process $\mathbf{Y}$ to be self-similar and have stationary increments. We think of a_n in (9.3) as the "order of magnitude" of the partial sums of the process $\mathbf{X}$.

If a statement as strong as a functional central limit theorem of the type (9.3) is not available, we can still try to measure the "size" of the partial sums, perhaps by looking just at the sequence of partial sums S_n in (9.1). If, however, a functional central limit theorem can be proved, then the limiting process $\mathbf{Y}$ in that theorem can be used to detect more phase transitions: we will see that there are boundaries such that different limiting processes are obtained on different sides of a boundary, perhaps even with the same order of magnitude of the partial sums.

We begin by considering the partial sums of finite-variance processes. Since the common mean of the process is of no interest in our discussion of dependence, we will simply assume that $\mathbf{X} = (X_n,\ n \in \mathbb{Z})$ is a stationary finite-variance zero-mean stochastic process. Let us recall that in this case, the behavior of the partial sums associated with the i.i.d. sequence is the invariance principle:

$$\big(n^{-1/2}S_n(t),\ t \geq 0\big) \Rightarrow \big(\sigma B(t),\ t \geq 0\big) \ \text{ as } n \to \infty \tag{9.4}$$

weakly in the Skorokhod J_1 topology on $D[0,\infty)$. Here $\sigma^2 = EX^2$ is the variance of the process $\mathbf{X}$, and $(B(t),\ t \geq 0)$ is a standard Brownian motion; see, for example, Billingsley (1999). Therefore, in the context of the behavior of the partial sums of finite-variance stationary processes, we associate short memory with the order of magnitude of the partial sums equal to the square root of the sample size, with the limiting process in a functional limit theorem being the Brownian motion.

It is natural to measure the order of magnitude of the partial sums of a zero-mean finite-variance process with the standard deviations (s_n) of the partial sums. That is,

$$s_n^2 = \mathrm{Var}S_n = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right),\ \ n = 1, 2, \ldots\ . \tag{9.5}$$

Note, however, that the standard deviation is an imperfect tool for measuring the distributional order of magnitude of S_n. The imperfection comes from the fact that while the family of the laws of $\big(S_n/s_n,\ n = 1, 2, \ldots\big)$ (assuming that s_n does not vanish) is clearly tight in $\mathbb{R}$, it is possible that $S_n/s_n \to 0$ in probability as $n \to \infty$; see Example 9.10.1. The appropriate behavior of the sequence of higher-order moments can sometimes be used to eliminate this imperfection, as the following proposition shows.

Proposition 9.2.2. *Let $\mathbf{X} = (X_n,\ n \in \mathbb{Z})$ be a stationary finite-variance zero-mean stochastic process. Assume that $E|X_0|^p < \infty$ for some $p > 2$. If*

$$\liminf_{n\to\infty} \frac{s_n^p}{E|S_n|^p} > 0\,, \tag{9.6}$$

then

$$\liminf_{n\to\infty} P\big(|S_n|/s_n > \varepsilon\big) > 0 \tag{9.7}$$

for all $0 < \varepsilon < 1$.

Proof. We use Problem 9.10.2 with $X = S_n^2$ and $p/2 > 1$ instead of p. For every $0 < \varepsilon < 1$,

$$\big(P\big(|S_n|/s_n > \varepsilon\big)\big)^{p/2-1} \geq (1-\varepsilon^2)^{p/2}\frac{(ES_n^2)^{p/2}}{E|S_n|^p}\,,$$

and hence (9.7) follows immediately from (9.6). □

A phase transition is transparent in the behavior of the partial sums of a process if we consider the family of zero-mean stationary *Gaussian* processes. In that case, we have the following easy result.

Proposition 9.2.3. *Let* $\mathbf{X} = (X_n,\, n \in \mathbb{Z})$ *be a zero-mean stationary Gaussian process. Then*

$$\big(s_n^{-1} S_n(t),\, t \geq 0\big) \Rightarrow \big(B(t),\ t \geq 0\big) \ \text{as } n \to \infty \tag{9.8}$$

weakly in the Skorokhod J_1 *topology on* $D[0,\infty)$ *if and only if the sequence* (s_n) *is regularly varying with exponent* $1/2$*. In particular,* (9.4) *holds for some* $\sigma > 0$ *if and only if* $s_n^2/n \to \sigma^2$ *as* $n \to \infty$*.*

Proof. Since the process $s_n^{-1}\mathbf{S}_n$ is a zero-mean Gaussian process for every n, the finite-dimensional distributions convergence part of (9.8) is equivalent to convergence of their covariance functions. Since all the processes start at zero, this is, in turn, equivalent to convergence of the corresponding incremental variances. However, for $0 \leq t_1 < t_2$,

$$\mathrm{Var}\big(s_n^{-1} S_{[nt_2]} - s_n^{-1} S_{[nt_1]}\big) = \frac{s^2_{[nt_2]-[nt_1]}}{s_n^2}\,,$$

and it is clear that the ratio converges to the incremental variance of the Brownian motion, $t_2 - t_1$, if and only if (s_n^2) is regularly varying with exponent 1. To show that in the latter case, tightness in the Skorokhod J_1 topology holds as well, it is enough to prove that there is $c > 0$ such that for all n large enough,

$$P\big(|S_n(t_2) - S_n(t_1)| > \lambda s_n,\ |S_n(t_3) - S_n(t_2)| > \lambda s_n\big) \leq c\lambda^{-4}(t_3 - t_1)^2 \tag{9.9}$$

for all $0 \leq t_1 < t_2 < t_3 < 1$ and all $\lambda > 0$; see Theorem 15.6 in Billingsley (1968). Note, however, that at least one of the differences, $S_n(t_2) - S_n(t_1)$ and $S_n(t_3) - S_n(t_2)$, is a sum, of length smaller than $n(t_3 - t_1)$, of the consecutive values of the process $\mathbf{X}$. Letting G denote a standard normal random variable, we use (with $\epsilon = 1/2$) the Potter bounds of Corollary 10.5.8 to show that the probability in (9.9) does not exceed

$$P\left(\frac{\max_{k<n(t_3-t_1)} s_k}{s_n}|G| > \lambda\right) \leq P\big(|G| > 2(t_3 - t_1)^{-1/2}\lambda\big)$$

for n large enough. Now (9.9) follows by applying Markov's inequality of order 4. □

Since the difference between a slowly varying sequence with exponent $1/2$ and the sequence $a_n = n^{1/2}$ is not significant enough to qualify as a phase transition, it is logical to interpret the result of Proposition 9.2.3 as saying that a stationary Gaussian process has short memory from the point of view of the behavior of its partial sums

if and only if the sequence of the variances of the partial sums is regularly varying with exponent 1. This occurs, in particular, if the covariances of the process are absolutely summable and add up to a positive number; see Proposition 6.1.1 and the discussion following that proposition. From this point of view, we must regard as long-range dependent any stationary Gaussian process for which the sequence of the variances of the partial sums is not regularly varying with exponent 1. Note that in this long-memory regime, the functional limit theorem of the type (9.3) exists only in a narrow situation described in the following extension of Proposition 9.2.3. The proof is identical to that of the former result.

Proposition 9.2.4. *Let* $\mathbf{X} = (X_n, n \in \mathbb{Z})$ *be a zero-mean stationary Gaussian process. Let* $0 < H < 1$. *Then*

$$\left(s_n^{-1} S_n(t),\ t \geq 0\right) \Rightarrow \left(B_H(t),\ t \geq 0\right) \text{ as } n \to \infty \tag{9.10}$$

weakly in the Skorokhod J_1 *topology on* $D[0, \infty)$ *if and only if the sequence* (s_n) *is regularly varying with exponent* H.

Note that in Proposition 9.2.4, as in Proposition 9.2.3, regular variation of the sequence of the standard deviations of the partial sums is required even for convergence of the finite-dimensional distributions.

Remark 9.2.5. An interesting consequence of the above discussion is that from the point of view of behavior of the partial sums, we view stationary Gaussian processes for which the variance of the partial sums is regularly varying with exponent less than $1/2$ as being long-range dependent. Note, in particular, that according to Proposition 9.2.4, a functional central limit theorem of the type (9.3) holds, and the limit is different from a Brownian motion. In particular, fractional Gaussian noise with Hurst exponent either $H < 1/2$ or $H > 1/2$ has, from this point of view, long memory.

It is, however, common in the literature to state that only when the partial sums of the process have a magnitude *larger than the magnitude of the partial sums of an i.i.d. sequence* is the memory long. In the case of fractional Gaussian noise, this claims long-range dependence only if $H > 1/2$.

Let us consider now phase transitions for partial sums in another situation, that of *fractional noises*. The name simply extends the terminology we have used before: if $\mathbf{Y} = \left(Y(t),\ t \geq 0\right)$ is a self-similar process with stationary increments, we call the stationary process $\mathbf{X}$ obtained by taking the increments of the process $\mathbf{Y}$ (at lag 1) a fractional noise. That is, $X_n = Y(n) - Y(n-1)$, $n = 1, 2, \ldots$. The following proposition is elementary.

We remind the reader that the abbreviation "H-SSSI process" means "self-similar process with stationary increments and Hurst exponent H".

Proposition 9.2.6. *Let* $\mathbf{X}$ *be a fractional noise corresponding to an H-SSSI process* $\mathbf{Y}$ *with* $H > 0$. *Then*

$$\left(n^{-H} S_n(t),\ t \geq 0\right) \Rightarrow \left(Y(t),\ t \geq 0\right) \text{ as } n \to \infty \tag{9.11}$$

in finite-dimensional distributions. If $\mathbf{Y}$ *has sample functions in* $D[0,\infty)$*, then the convergence also holds as weak convergence in the Skorokhod* J_1 *topology on that space.*

Proof. For all $t_1,\dots,t_k \geq 0$, by the self-similarity of the process $\mathbf{Y}$,

$$\begin{aligned}\left(n^{-H}S_n(t_j),\, j=1,\dots,k\right) &= \left(n^{-H}Y\left([nt_j]\right),\, j=1,\dots,k\right)\\ &\stackrel{d}{=} \left(Y\left(n^{-1}[nt_j]\right),\, j=1,\dots,k\right).\end{aligned}$$

Since the process $\mathbf{Y}$ is continuous in probability by Lemma 8.2.5, we conclude that

$$\left(n^{-H}S_n(t_j),\, j=1,\dots,k\right) \Rightarrow \left(Y(t_1),\dots,Y(t_k)\right) \text{ as } n\to\infty$$

as $n\to\infty$.

Suppose now that $\mathbf{Y}$ has sample functions in $D[0,\infty)$, and write

$$\left(n^{-H}S_n(t),\, t\geq 0\right) \stackrel{d}{=} \left(Y\left(n^{-1}[nt]\right),\, t\geq 0\right)$$

in the sense of equality of two probability measures on $D[0,\infty)$. Now the statement of the proposition follows from the fact that

$$\left(Y\left(n^{-1}[nt]\right),\, t\geq 0\right) \to \left(Y(t),\ t\geq 0\right)$$

a.s. in $D[0,\infty)$ as $n\to\infty$; see Exercise 9.10.3. □

Remark 9.2.7. A conclusion from Proposition 9.2.6 is that the only fractional noise process with a finite variance for which a Brownian motion appears as the limit in a functional central limit theorem of the type (9.3) is the fractional noise corresponding to the Brownian motion itself, that is, an i.i.d. sequence of Gaussian random variables. Therefore, from the point of view of behavior of the partial sums, all fractional noises with finite variance, apart from the i.i.d. sequence, are long-range dependent. This is true in the entire range of the Hurst exponent, $0 < H < 1$. Recall from Section 8.3 that there exist multiple SSSI processes with finite variance and $H = 1/2$. The fractional noise corresponding to such a process should also be viewed as long-range dependent as long as it is not an i.i.d. sequence.

If, however, one insists that under long memory, the magnitude must be larger than the magnitude of the partial sums of an i.i.d. sequence, then the only finite-variance fractional noises with long-range dependence are those with $1/2 < H < 1$.

9.3 Partial Sums of Finite-Variance Linear Processes

We will now consider phase transition in the behavior of the partial sums of the finite-variance linear processes of Section 1.4. Let $(\varepsilon_n,\ n \in \mathbb{Z})$ be a sequence of i.i.d. random variables with zero mean and finite variance, and let (φ_n) be deterministic coefficients satisfying

$$\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty\,. \tag{9.12}$$

By Theorem 1.4.1, the process

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\,\varepsilon_j = \sum_{j=-\infty}^{\infty} \varphi_j\,\varepsilon_{n-j}\,, \quad n \in \mathbb{Z}\,, \tag{9.13}$$

is a well-defined stationary process with zero mean and finite variance. For this family of processes, we will consider phase transitions for the partial sums. In search of such a phase transition, let us assume first that the coefficients (φ_n) are absolutely summable:

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty\,. \tag{9.14}$$

Under this assumption, we have the following result.

Theorem 9.3.1. *Let* $\mathbf{X}$ *be a finite-variance infinite-moving-average process with coefficients* (φ_n) *satisfying* (9.14)*. Then*

$$\left(n^{-1/2} S_n(t),\ t \geq 0\right) \Rightarrow \left(a_\varphi \sigma_\varepsilon B(t),\ \ t \geq 0\right) \ \text{as } n \to \infty \tag{9.15}$$

weakly in the Skorokhod J_1 *topology on* $D[0,\infty)$*, where* $\mathbf{B}$ *is a standard Brownian motion,* $\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_0)$*, and*

$$a_\varphi = \sum_{j=-\infty}^{\infty} \varphi_j\,. \tag{9.16}$$

Proof. It is convenient to consider the symmetric sums

$$\hat{S}_n = \sum_{k=-n}^{n} X_k,\ \ n = 0, 1, 2, \ldots\,,$$

and prove the following statement, equivalent to (9.15):

$$\left((2n)^{-1/2} \hat{S}_n(t),\ t \geq 0\right) \Rightarrow \left(a_\varphi \sigma_\varepsilon B(t),\ \ t \geq 0\right) \ \text{as } n \to \infty\,. \tag{9.17}$$

Note that

$$\begin{aligned}\hat{S}_n &= \sum_{j=-\infty}^{\infty}\left(\sum_{k=-n}^{n}\varphi_{k-j}\right)\varepsilon_j \qquad (9.18)\\ &= \sum_{j=-n}^{n}\left(\sum_{k=-n-j}^{n-j}\varphi_k\right)\varepsilon_j + \sum_{|j|>n}\left(\sum_{k=-n-j}^{n-j}\varphi_k\right)\varepsilon_j\\ &= a_\varphi\sum_{j=-n}^{n}\varepsilon_j - \sum_{j=-n}^{n}\left(\sum_{k:\,|k+j|>n}\varphi_k\right)\varepsilon_j + \sum_{|j|>n}\left(\sum_{k=-n-j}^{n-j}\varphi_k\right)\varepsilon_j\\ &:= A_n - B_n + C_n\,.\end{aligned}$$

By the invariance principle (9.4), we have, in the obvious notation,

$$\big((2n)^{-1/2}A_n(t),\ t\ge 0\big) \Rightarrow \big(a_\varphi\sigma_\varepsilon B(t),\ \ t\ge 0\big) \ \text{ as } n\to\infty\,.$$

Therefore, the statement (9.17) will follow once we prove that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\Big(n^{-1/2}\sup_{0\le t\le 1}\big|B_{[nt]}\big| > \epsilon\Big) = \lim_{n\to\infty} P\Big(n^{-1/2}\sup_{0\le t\le 1}\big|C_{[nt]}\big| > \epsilon\Big) = 0\,. \qquad (9.19)$$

We begin with the first statement in (9.19). Since we are free to switch from the sequence (φ_j) to the sequence (φ_{-j}), it is enough to prove that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\Big(n^{-1/2}\sup_{0\le t\le 1}\big|\hat{B}_{[nt]}\big| > \epsilon\Big) = 0\,, \qquad (9.20)$$

where

$$\hat{B}_n = \sum_{j=0}^{n}\left(\sum_{k=n-j+1}^{\infty}\varphi_k\right)\varepsilon_j,\ \ n = 0,1,2,\ldots\,.$$

Let (d_n) be a sequence of positive integers growing to infinity, $d_n = o(n)$. The probability on the left-hand side of (9.20) can be rewritten in the form

$$\begin{aligned}&P\left(n^{-1/2}\max_{m=0,1,\ldots,n}\left|\sum_{j=0}^{m}\left(\sum_{k=m-j+1}^{\infty}\varphi_k\right)\varepsilon_j\right| > \epsilon\right) \qquad (9.21)\\ &= P\left(n^{-1/2}\max_{m=0,1,\ldots,n}\left|\sum_{0\le j\le m-d_n}\left(\sum_{k=m-j+1}^{\infty}\varphi_k\right)\varepsilon_j\right| > \epsilon/2\right)\end{aligned}$$

$$+P\left(n^{-1/2}\max_{m=0,1,\dots,n}\left|\sum_{j=\max(0,m-d_n)+1}^{m}\left(\sum_{k=m-j+1}^{\infty}\varphi_k\right)\varepsilon_j\right|>\epsilon/2\right).$$

The first term in (9.21) is

$$\begin{aligned}
&P\left(n^{-1/2}\max_{m=d_n+1,\dots,n}\left|\sum_{k=d_n+1}^{\infty}\left(\sum_{j=\max(0,m-k)+1}^{m-d_n}\varepsilon_j\right)\varphi_k\right|>\epsilon/2\right)\\
&\le P\left(n^{-1/2}\sum_{k=d_n+1}^{\infty}\max_{m=d_n+1,\dots,n}\left|\sum_{j=\max(0,m-k)+1}^{m-d_n}\varepsilon_j\right||\varphi_k|>\epsilon/2\right)\\
&\le\frac{2}{\epsilon}n^{-1/2}\sum_{k=d_n+1}^{\infty}|\varphi_k|E\max_{m=d_n+1,\dots,n}\left|\sum_{j=\max(0,m-k)+1}^{m-d_n}\varepsilon_j\right|\\
&\le\frac{2}{\epsilon}n^{-1/2}\sum_{k=d_n+1}^{\infty}|\varphi_k|\left[E\left(\max_{m=d_n+1,\dots,n}\left|\sum_{j=\max(0,m-k)+1}^{m-d_n}\varepsilon_j\right|\right)^2\right]^{1/2}\\
&\le\frac{2}{\epsilon}n^{-1/2}\sum_{k=d_n+1}^{\infty}|\varphi_k|2\sigma_\varepsilon(n-d_n)^{1/2}\le\frac{4\sigma_\varepsilon}{\epsilon}\sum_{k=d_n+1}^{\infty}|\varphi_k|\to 0,
\end{aligned}$$

since $d_n\to\infty$ (in the penultimate step, we have used Doob's maximal inequality for martingales). On the other hand, the second term in (9.21) does not exceed

$$\begin{aligned}
&P\left(n^{-1/2}\max_{m=0,1,\dots,n}\sum_{j=\max(0,m-d_n)+1}^{m}|\varepsilon_j|>\frac{\epsilon}{2\sum_{k=-\infty}^{\infty}|\varphi_k|}\right)\\
&\le P\left(\max_{j=0,1,\dots,n}|\varepsilon_j|>c\frac{n^{1/2}}{d_n}\right)=1-\left(1-P\big(|\varepsilon_0|>cn^{1/2}/d_n\big)\right)^n\to 0
\end{aligned}$$

if we choose (d_n) such that

$$P\big(|\varepsilon_0|>cn^{1/2}/d_n\big)=o(1/n)\,,$$

which can be done, since ε_0 has a finite second moment. Here $c=\epsilon/(2\sum|\varphi_k|)$. This proves (9.20) and hence the first statement in (9.19).

The second statement in (9.19) can be proved in a similar way. We leave the details to Exercise 9.10.4. □

Remark 9.3.2. We see from Theorem 9.3.1 that a finite-variance linear process with coefficients (φ_n) satisfying (9.14) and $a_\varphi\neq 0$ in (9.16) satisfies a functional central

limit theorem of the type (9.3) with a Brownian limit. Such a process should be viewed as a short-memory process from the point of view of behavior of the partial sums.

We consider now a finite-variance infinite-moving-average process (9.13) whose coefficients do not satisfy the summability condition (9.14). Instead, we will impose a certain balanced regular variation assumption. Specifically, assume that there is a regularly varying sequence (b_n) with exponent $\beta \in (-1, -1/2)$ such that

$$\lim_{n\to\infty} \frac{\varphi_n}{b_n} = c_+, \quad \lim_{n\to\infty} \frac{\varphi_{-n}}{b_n} = c_-, \tag{9.22}$$

for some $c_+, c_- \in \mathbb{R}$, at least one of which is different from zero. Clearly, such a sequence of coefficients (φ_n) satisfies the square summability assumption (9.12) but not the absolute summability assumption (9.14). Now the limiting behavior of the partial sums is different.

Theorem 9.3.3. *Let* $\mathbf{X}$ *be a finite-variance infinite-moving-average process with coefficients* (φ_n) *satisfying* (9.22)*,* $-1 < \beta < -1/2$*. Then*

$$\left(\frac{1}{n^{3/2}b_n} S_n(t),\ t \geq 0\right) \Rightarrow \left(c_\varphi \sigma_\varepsilon B_H(t),\ t \geq 0\right) \text{ as } n \to \infty \tag{9.23}$$

weakly in the Skorokhod J_1 *topology on* $D[0,\infty)$*, where* $\mathbf{B}_H$ *is the standard fractional Brownian motion with* $H = 3/2 + \beta$*, satisfying* $E\big(B_H(t)^2\big) = t^{2H}/2$*,* $t \geq 0$*. Furthermore,* $\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_0)$*, and*

$$c_\varphi^2 = 2(1+\beta)^{-2} \int_{-\infty}^{\infty} \left\{c_+\left[(1+y)_+^{1+\beta} - y_+^{1+\beta}\right] - c_-\left[(1+y)_-^{1+\beta} - y_-^{1+\beta}\right]\right\}^2 dy\,. \tag{9.24}$$

Proof. We begin by writing

$$S_n = X_1 + \ldots + X_n = \sum_{j=-\infty}^{\infty} \left(\sum_{i=1}^{n} \varphi_{i-j}\right) \varepsilon_j,\ n = 1, 2, \ldots\,, \tag{9.25}$$

so that

$$s_n^2 = \mathrm{Var} S_n = \sigma_\varepsilon^2 \sum_{j=-\infty}^{\infty} \left(\sum_{i=1}^{n} \varphi_{i-j}\right)^2.$$

We will show that

$$\lim_{n\to\infty} \frac{s_n^2}{n^3 b_n^2} = c_\varphi^2/2\,, \tag{9.26}$$

with c_φ^2 given in (9.24). We postpone the proof of (9.26) for a moment to see how this statement implies the claim of the theorem.

In order to prove convergence of the finite-dimensional distributions, we will use the Lindeberg central limit theorem; see, e.g., Theorem 27.2 in Billingsley (1995), the multivariate version of which follows immediately from the usual univariate version by the Cramér–Wold device. Let $0 < t_1 < t_2 < \ldots < t_d$. In order to prove that

$$\left(\frac{1}{n^{3/2}b_n}S_{[nt_1]}, \ldots, \frac{1}{n^{3/2}b_n}S_{[nt_d]}\right) \Rightarrow \left(c_\varphi \sigma_\varepsilon B_H(t_1), \ldots, c_\varphi \sigma_\varepsilon B_H(t_d)\right), \tag{9.27}$$

we use the representation (9.25) of the vector on the left-hand side of (9.27) as a sum of independent random variables. Define

$$\theta_{j,n} = \sum_{i=1}^{n} \varphi_{i-j},\ j \in \mathbb{Z},\ n = 1, 2, \ldots .$$

Then (9.27) will follow once we check the incremental variance convergence

$$\lim_{n\to\infty} \frac{\mathrm{Var}\left(S_{[nt]} - S_{[ns]}\right)}{n^3 b_n^2} = c_\varphi^2 \sigma_\varepsilon^2 (t-s)^{2H}/2 \tag{9.28}$$

for every $0 \le s < t < \infty$, as well as the Lindeberg condition

$$\lim_{n\to\infty} \frac{1}{n^3 b_n^2} \sum_{j=-\infty}^{\infty} \theta_{j,n}^2 E\left[\varepsilon_0^2 \mathbf{1}\left(|\varepsilon_0| > \frac{n^{3/2}|b_n|}{|\theta_{j,n}|}\epsilon\right)\right] = 0 \tag{9.29}$$

for every $\epsilon > 0$. However, once (9.26) has been proved, the incremental variance property (9.28) is immediate, since

$$\lim_{n\to\infty} \frac{\mathrm{Var}\left(S_{[nt]} - S_{[ns]}\right)}{n^3 b_n^2} = \lim_{n\to\infty} \frac{\mathrm{Var}\left(S_{[nt]-[ns]}\right)}{n^3 b_n^2}$$

$$= \frac{c_\varphi^2 \sigma_\varepsilon^2}{2} \lim_{n\to\infty} \frac{\left([nt] - [ns]\right)^3 b_{[nt]-[ns]}^2}{n^3 b_n^2} = c_\varphi^2 \sigma_\varepsilon^2 (t-s)^{2H}/2,$$

by the regular variation of the sequence (b_n). Furthermore, it follows easily by the assumption (9.22) and Exercise 10.9.9 that there is $C \in (0, \infty)$ such that for each $j \in \mathbb{Z}$,

$$|\theta_{j,n}| \le C \sum_{i=1}^{n} |b_i| \sim C(1+\beta)^{-1} n |b_n|$$

as $n \to \infty$. Therefore, as $n \to \infty$,

$$\sum_{j=-\infty}^{\infty} \theta_{j,n}^2 E\left[\varepsilon_0^2 \mathbf{1}\left(|\varepsilon_0| > \frac{n^{3/2}|b_n|}{|\theta_{j,n}|}\epsilon\right)\right] = o(1) \sum_{j=-\infty}^{\infty} \theta_{j,n}^2 = o(1) s_n^2 ,$$

and (9.29) will follow once (9.26) has been proved.

In order to prove tightness, we use, once again, Theorem 15.6 in Billingsley (1968). Since $H > 1/2$, according to this theorem it is enough to prove that there is $c > 0$ such that for all n large enough,

$$P\big(|S_n(t_2) - S_n(t_1)| > \lambda n^{3/2} b_n,\ |S_n(t_3) - S_n(t_2)| > \lambda n^{3/2} b_n\big) \le c\lambda^{-2}(t_3 - t_1)^{2H}$$

for all $0 \le t_1 < t_2 < t_3 < 1$ and all $\lambda > 0$. As in the proof of (9.9), this probability does not exceed

$$\max_{m \le n(t_3 - t_1)} P\big(|S_m| > \lambda n^{3/2} b_n\big) \le \lambda^{-2} \max_{m \le n(t_3 - t_1)} \frac{s_m^2}{n^3 b_n^2} .$$

Since by (9.26) and Exercise 10.9.9, $\max_{m \le n} s_m^2 \sim s_n^2$ as $n \to \infty$, the required bound follows from (9.26).

We finish the proof of the theorem with proving (9.26). We may assume that the sequence (b_n) is eventually positive. Define

$$\theta_{+,n} = \sum_{i=1}^{n} \varphi_i = \theta_{0,n}, \quad \theta_{-,n} = \sum_{i=0}^{n} \varphi_{-i} = \theta_{n+1,n+1} ,$$

and note that by (9.22) and Theorem 10.5.6,

$$\lim_{n\to\infty} \frac{\theta_{\pm,n}}{n b_n} = \frac{c_\pm}{1+\beta} . \tag{9.30}$$

We can write

$$s_n^2 = \sigma_\varepsilon^2 \left[\sum_{j=-\infty}^{0} \theta_{j,n}^2 + \sum_{j=1}^{n-1} \theta_{j,n}^2 + \sum_{j=n}^{\infty} \theta_{j,n}^2\right] =: \sigma_\varepsilon^2 \big(V_1(n) + V_2(n) + V_3(n)\big) .$$

The following three claims will imply (9.26):

$$\lim_{n\to\infty} \frac{V_1(n)}{n^3 b_n^2} = (1+\beta)^{-2} c_+^2 \int_0^\infty \big((1+y)_+^{1+\beta} - y_+^{1+\beta}\big)^2 \, dy , \tag{9.31}$$

$$\lim_{n\to\infty} \frac{V_2(n)}{n^3 b_n^2} = (1+\beta)^{-2} \int_0^1 \big(c_+(1-y)_+^{1+\beta} + c_- y_+^{1+\beta}\big)^2 \, dy , \tag{9.32}$$

$$\lim_{n\to\infty} \frac{V_3(n)}{n^3 b_n^2} = (1+\beta)^{-2} c_-^2 \int_0^\infty \left((1+y)_+^{1+\beta} - y_+^{1+\beta}\right)^2 dy\,. \tag{9.33}$$

We will prove (9.31). The statements (9.32) and (9.33) can be proved in a similar way, and are left to the reader in Exercise 9.10.5.

For the purpose of proving (9.31), we will assume that $c_+ > 0$. If $c_+ < 0$, one can multiply the entire process by -1. If $c_+ = 0$, one can use an obvious monotonicity argument and prove that the limit in (9.31) is zero once the truth of that statement for $c_+ > 0$ has been established. Let $0 < \epsilon < 1$ be a small number. Write

$$V_1(n) = \sum_{k=0}^{\infty} \sum_{kn\epsilon \le j < (k+1)n\epsilon} \left(\theta_{+,n+j} - \theta_{+,j}\right)^2.$$

By the assumption, the moving-average coefficients (φ_n) are eventually (as $n \to \infty$) positive, so that the sequence $(\theta_{+,n})$ is eventually increasing. Therefore, for all n large enough,

$$V_1(n) \le (1+n\epsilon) \sum_{k=0}^{\infty} \left(\theta_{+,[n(1+(k+1)\epsilon)]} - \theta_{+,[nk\epsilon]}\right)^2 + O(1)\,.$$

We claim that

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{\left(\theta_{+,[n(1+(k+1)\epsilon)]} - \theta_{+,[nk\epsilon]}\right)^2}{\theta_{+,n}^2} = \sum_{k=0}^{\infty} \left((1+(k+1)\epsilon)^{1+\beta} - (k\epsilon)^{1+\beta}\right)^2. \tag{9.34}$$

Once (9.34) has been verified, it will follow from (9.30) that for $0 < \epsilon < 1$,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{V_1(n)}{n^3 b_n^2} &\le (1+\beta)^{-2} c_+^2 \epsilon \sum_{k=0}^{\infty} \left((1+(k+1)\epsilon)^{1+\beta} - (k\epsilon)^{1+\beta}\right)^2 \\
&\le (1+\beta)^{-2} c_+^2 \epsilon \int_0^\infty \left((1+2\epsilon+x\epsilon)^{1+\beta} - (x\epsilon)^{1+\beta}\right)^2 dx \\
&\quad (1+\beta)^{-2} c_+^2 \int_0^\infty \left((1+2\epsilon+x)^{1+\beta} - x^{1+\beta}\right)^2 dx\,.
\end{aligned}$$

Letting $\epsilon \to 0$ establishes then the upper bound part in (9.31).

In order to prove (9.34), we note that the sequence $(\theta_{+,n})$ is regularly varying with exponent $1+\beta$, so that each term in the sum on the left-hand side of (9.34) converges, as $n \to \infty$, to the corresponding term in the sum on the right-hand side. Therefore, we need only to exhibit a dominating function. We use the Potter bounds of Corollary 10.5.8. Let $0 < \delta < -(\beta + 1/2)$. Denoting by C a finite positive constant that is allowed to change from appearance to appearance, we have, for n large enough, for all $k \ge 1/\epsilon$,

$$\begin{aligned}\frac{\theta_{+,[n(1+(k+1)\epsilon)]} - \theta_{+,[nk\epsilon]}}{\theta_{+,n}} &\le Cn^{-1} \sum_{i=[nk\epsilon]+1}^{[n(1+(k+1)\epsilon)]} \frac{a_i}{a_n} \\ &\le Cn^{-1} \sum_{i=[nk\epsilon]+1}^{[n(1+(k+1)\epsilon)]} (i/n)^{\beta+\delta} \\ &\le Ck^{\beta+\delta}\,.\end{aligned}$$

By the choice of δ, the square of this function is summable. Hence, (9.34) holds.

Similarly, for all n large enough,

$$V_1(n) \ge (n\epsilon - 1) \sum_{k=1}^{\infty} \left(\theta_{+,[n(1+k\epsilon)]} - \theta_{+,[n(k+1)\epsilon]}\right)^2,$$

and proceeding in the same manner as above, we obtain the lower bound part of (9.31). □

Since a functional central limit theorem of the type (9.3) holds for a finite-variance linear process with coefficients (φ_n) satisfying the regular variation assumption (9.22), but the limit is a fractional Brownian motion with exponent $H > 1/2$ (and a function of the exponent of the regular variation of the coefficients), such an infinite-moving-average process should be viewed as a long-memory process from the point of view of behavior of the partial sums.

As far as the behavior of the partial sums is concerned, long-range dependence in finite-variance linear processes is also possible when the coefficients (φ_n) are summable but their sum a_φ is equal to zero. This possibility is similar to the case of the Hurst exponent being in the range $0 < H < 1/2$ for fractional noises in Proposition 9.2.6, because now the "size" of the partial sums may be smaller than that under the short-memory conditions of Theorem 9.3.1. Once again, we will impose a balanced regular variation assumption. Specifically, we will assume that (9.22) holds, but this time, the exponent of regular variation of the sequence (b_n) is assumed to be in the range $\beta \in (-3/2, -1)$. This clearly implies that the coefficients (φ_n) satisfy the absolute summability assumption (9.14).

Theorem 9.3.4. *Let* $\mathbf{X}$ *be a finite-variance infinite-moving-average process with coefficients* (φ_n) *satisfying* (9.22)*,* $-3/2 < \beta < -1$*. Assume that*

$$a_\varphi := \sum_{j=-\infty}^{\infty} \varphi_j = 0\,. \tag{9.35}$$

Then, in the notation of Theorem 9.3.3,

$$\left(\frac{1}{n^{3/2} b_n} S_n(t),\ t \ge 0\right) \Rightarrow \left(c_\varphi \sigma_\varepsilon B_H(t),\ t \ge 0\right) \ \text{as } n \to \infty \tag{9.36}$$

in finite-dimensional distributions. Furthermore, if we also have $E|\varepsilon_0|^p < \infty$ *for some* $p > 2/(3+2\beta)$*, then* (9.36) *also holds in the sense of weak convergence in the Skorokhod* J_1 *topology on* $D[0,\infty)$.

The proof of the theorem is similar to that of Theorem 9.3.3 and is left to Exercise 9.10.6.

Remark 9.3.5. One can summarize the results of Theorems 9.3.1, 9.3.4, and 9.3.3 and say that under the assumption (9.14) of the summability of coefficients and nonvanishing sum a_φ of the coefficients, the finite-variance infinite moving average has short memory from the point of view of the behavior of its partial sums, while from the same point of view, the memory is long if the coefficients satisfy the regular variation assumption (9.22) with either $\beta \in (-1,-1/2)$ or $-3/2 < \beta < -1$, but $a_\varphi = 0$. The presence of a phase transition is clear, even though we have not made an effort to draw the exact boundary between the short- and long-memory cases. In particular, it is clear that the arguments in Theorems 9.3.4 and 9.3.3 require little beyond regular variation of certain partial sums of the coefficients, instead of the regular variation of the coefficients themselves.

9.4 Partial Sums of Finite-Variance Infinitely Divisible Processes

Our task in this section is to find phase transitions from the point of view of partial sums for stationary infinitely divisible processes without a Gaussian component given in the form (3.91). That is,

$$X_n = \int_E f \circ \phi^n(s)\, M(ds),\ n \in \mathbb{Z}\,, \tag{9.37}$$

where $(E, \mathcal{E}, m)$ is a σ-finite measure space and M a homogeneous infinitely divisible random measure on E without a Gaussian component with control measure m and a constant local Lévy measure ρ. We assume that the constant local shift of M vanishes: $b = 0$. Furthermore, $f \in L_0(M)$, and $\phi : E \to E$ is a nonsingular measure-preserving map. For simplicity, we will assume that the local Lévy measure ρ is symmetric. Recall that by Theorem 3.3.2, the assumption $f \in L_0(M)$ means that

$$\int_E \left(\int_{-\infty}^{\infty} \min\bigl(1, x^2 f(s)^2\bigr)\, \rho(dx) \right) m(ds) < \infty\,. \tag{9.38}$$

In this section, we will assume that the process $\mathbf{X}$ has finite variance. This is equivalent to saying that

$$\int_{-\infty}^{\infty} x^2\, \rho(dx) < \infty \ \text{ and } \ \int_E f(s)^2\, m(ds) < \infty\,; \tag{9.39}$$

see, e.g., Sato (1999). This condition already implies (9.38), hence the integrability of f.

The laws of stationary stochastic processes $\mathbf{X}$ we have defined by (9.37) are thus determined by several parameters: the local Lévy measure ρ, the kernel f, and ergodic-theoretical properties of the map ϕ with respect to the control measure m. We will find in this section that for finite-variance processes, the behavior of the partial sums is largely determined by the ergodic-theoretical properties of the map ϕ and by the "size" of the kernel f.

We begin with the case in which the map ϕ is dissipative with respect to the control measure m. In order to obtain behavior of the partial sums consistent with short memory, we will impose additional "size" constraints on the function f. Specifically, we assume that

$$\int_E |f(s)| \sum_{k=-\infty}^{\infty} |f| \circ \phi^k(s)\, m(ds) < \infty\,. \tag{9.40}$$

Notice that (9.40) implies that

$$\sum_{k=-\infty}^{\infty} |f| \circ \phi^k(s) < \infty \;\; m\text{-a.e.}, \tag{9.41}$$

which would follow from the dissipativity of ϕ by Theorem 2.4.5 if we had $f \in L_1(m)$. The latter condition is not, however, guaranteed just by the assumption (9.39). Moreover, by Theorem 2.4.5, (9.41) itself implies dissipativity of ϕ, at least on (the only relevant) ϕ-invariant set

$$E_f = \left\{ s \in E : f \circ \phi^k(s) \neq 0 \;\text{ for some } k \in \mathbb{Z} \right\}. \tag{9.42}$$

Theorem 9.4.1. *Let* $\mathbf{X}$ *be a finite-variance stationary symmetric infinitely divisible process without a Gaussian component given by* (9.37)*. Suppose that the map* ϕ *is dissipative with respect to the measure m, and that* (9.40) *holds. Then*

$$\left(n^{-1/2} S_n(t),\, t \geq 0\right) \Rightarrow \left(\sigma_{\mathbf{X}} B(t),\; t \geq 0\right) \;\; as\; n \to \infty \tag{9.43}$$

in finite-dimensional distributions, where

$$\sigma_{\mathbf{X}}^2 = \int_{-\infty}^{\infty} x^2\, \rho(dx) \left(\sum_{k=-\infty}^{\infty} \int_E f(s) f \circ \phi^k(s)\, m(ds) \right). \tag{9.44}$$

Proof. By the linearity of the integral (Theorem 3.3.2), we know that for $n = 1, 2, \ldots,$

$$S_n = \int_E \sum_{k=1}^{n} f \circ \phi^k(s)\, m(ds)\,, \tag{9.45}$$

which is a symmetric infinitely divisible random variable without a Gaussian component. A standard argument shows that it is enough to prove weak convergence of the type

$$n^{-1/2}\big(S_{j_1 n},\dots,S_{j_m n}\big) \Rightarrow \sigma_{\mathbf{X}}\big(B(j_1),\dots,B(j_m)\big) \tag{9.46}$$

for every $m = 1,2,\dots$ and positive integers $j_1,\dots,j_m$; see Exercise 9.10.7. Conditions for convergence of a sequence of infinitely divisible random vectors to a Gaussian limit are given, for example, in Theorem 13.14 in Kallenberg (1997). With the representation (9.45), we need to verify two statements. First of all, we have to check that for all nonnegative integers $j_1 < j_2$ and $\kappa > 0$,

$$\int_E \int_{-\infty}^{\infty} x^2 \frac{1}{n} \left(\sum_{k=j_1 n+1}^{j_2 n} f\circ\phi^k(s) \right)^2 \mathbf{1}\left(x^2 \frac{1}{n} \left(\sum_{k=j_1 n+1}^{j_2 n} f\circ\phi^k(s) \right)^2 \le \kappa \right) \rho(dx)\, m(ds)$$

$$\to \sigma_{\mathbf{X}}^2 (j_2 - j_1) \tag{9.47}$$

as $n \to \infty$, and we also have to check that for every positive integer j,

$$\int_E \int_{-\infty}^{\infty} \mathbf{1}\left(x^2 \frac{1}{n} \left(\sum_{k=1}^{jn} f\circ\phi^k(s) \right)^2 > \mathbf{1} \right) \rho(dx)\, m(ds) \to 0 \tag{9.48}$$

as $n \to \infty$. Since the map ϕ is measure-preserving, it is clearly enough to consider only the case $j_1 = 0, j_2 = 1$ in (9.47), and $j = 1$ in (9.48).

We begin by proving (9.48) (with $j = 1$). Let M be a large positive number. We write the integral in (9.48) as $I_1^{(M)}(n) + I_2^{(M)}(n)$, with the two terms given, respectively, by

$$\int_E \mathbf{1}\left(\left(\sum_{k=1}^{n} f\circ\phi^k(s) \right)^2 \le \frac{n}{M} \right) \left(\int_{-\infty}^{\infty} \mathbf{1}\left(x^2 > \frac{n}{\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2} \right) \rho(dx) \right) m(ds)$$

and

$$\int_E \mathbf{1}\left(\left(\sum_{k=1}^{n} f\circ\phi^k(s) \right)^2 > \frac{n}{M} \right) \left(\int_{-\infty}^{\infty} \mathbf{1}\left(x^2 > \frac{n}{\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2} \right) \rho(dx) \right) m(ds)\,.$$

Then (9.48) will follow from the following two claims:

$$\lim_{M\to\infty} \limsup_{n\to\infty} I_1^{(M)}(n) = 0 \tag{9.49}$$

and

$$\lim_{n\to\infty} I_2^{(M)}(n) = 0 \text{ for every } M > 0. \tag{9.50}$$

We prove (9.49) first. Note that

$$\begin{aligned} I_1^{(M)}(n) &\le 2\int_{M^{1/2}}^{\infty} x^2\,\rho(dx)\int_E\int_{-\infty}^{\infty}\frac{1}{n}\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 m(ds) \\ &= 2\int_{M^{1/2}}^{\infty} x^2\,\rho(dx)\left(\frac{1}{n}\sum_{k=-(n-1)}^{n-1}(n-|k|)\int_E f(s)f\circ\phi^k(s)\,m(ds)\right) \\ &\le 2\int_{M^{1/2}}^{\infty} x^2\,\rho(dx)\sum_{k=-(n-1)}^{n-1}\left|\int_E f(s)f\circ\phi^k(s)\,m(ds)\right|. \end{aligned}$$

Now the claim (9.49) follows from (9.40) and (9.39).

In order to prove (9.50) for a fixed M, we write

$$\begin{aligned} I_2^{(M)}(n) &\le 2\rho\big((1,\infty)\big)\int_E \mathbf{1}\left(\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 > \frac{n}{M}\right) m(ds) \\ &\quad + 2\int_E \mathbf{1}\left(\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 > n\right)\rho\left(\left(\frac{n^{1/2}}{\left|\sum_{k=1}^{n} f\circ\phi^k(s)\right|},1\right)\right) m(ds) \\ &:= I_{21}^{(M)}(n) + I_{22}(n)\,. \end{aligned}$$

For $K > 0$, let

$$A_K = \left\{ s\in E : \sum_{k=-\infty}^{\infty} |f|\circ\phi^k(s) > K\right\} \downarrow \emptyset \text{ up to an } m\text{-null set as } K\to\infty, \tag{9.51}$$

where the convergence follows from (9.41). Letting c be a finite (M-dependent) constant that is allowed to change from appearance to appearance, we have for a fixed K, for all n large enough,

$$\begin{aligned} I_{21}^{(M)}(n) &= c\int_{A_K} \mathbf{1}\left(\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 > \frac{n}{M}\right) m(ds) \\ &\le c\int_E \mathbf{1}_{A_K}(s)\,n^{-1}\left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 m(ds) \end{aligned} \tag{9.52}$$

$$\leq c \int_E \mathbf{1}_{A_K}(s)|f(s)| \sum_{k=-\infty}^{\infty} |f| \circ \phi^k(s)\, m(ds)\,,$$

and the last expression vanishes as $K \to \infty$ by (9.40), (9.51), and the monotone convergence theorem. Therefore, $I_{21}^{(M)}(n) \to 0$ as $n \to \infty$.

For a large $K > 0$, we have

$$I_{22}(n) \leq 2 \int_E \mathbf{1}\left(\left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 > Kn\right) \rho\left(\left(\frac{n^{1/2}}{\left|\sum_{k=1}^n f \circ \phi^k(s)\right|}, 1\right)\right) m(ds)$$

$$+ 2\rho\big((K^{-1/2}, 1)\big) \int_E \mathbf{1}\left(\left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 > n\right) m(ds)\,.$$

We already know that the second term vanishes as $n \to \infty$ for every $K > 0$. Using the fact that $x^2 \rho\big((x, 1)\big) \to 0$ as $x \downarrow 0$, we estimate the first term as

$$o_K(1) \int_E \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 m(ds)\,,$$

where $o_K(1) \to 0$ as $K \to \infty$. Since the integral is bounded by (9.40), we conclude that $I_{22}^{(M)}(n) \to 0$ as $n \to \infty$, which proves (9.50). Hence (9.48) holds.

We now prove (9.47) (with $j_1 = 0$, $j_2 = 1$). Write the integral in (9.47) as

$$\int_{-\infty}^{\infty} x^2\, \rho(dx) \int_E \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 m(ds)$$

$$- \int_E \int_{-\infty}^{\infty} x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 \mathbf{1}\left(x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 > \kappa\right) \rho(dx)\, m(ds)\,.$$

Since by (9.40), the first term converges, as $n \to \infty$, to $\sigma_{\mathbf{X}}^2$, it remains to prove that the second term vanishes as $n \to \infty$. For $M > 0$, we write this term as

$$2 \int_E \int_M^{\infty} x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 \mathbf{1}\left(x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 > \kappa\right) \rho(dx)\, m(ds)$$

$$+ 2 \int_E \int_0^M x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 \mathbf{1}\left(x^2 \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s)\right)^2 > \kappa\right) \rho(dx)\, m(ds)$$

$$\leq 2 \int_M^\infty x^2 \, \rho(dx) \int_E \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s) \right)^2 m(ds)$$

$$+ 2 \int_0^M x^2 \, \rho(dx) \int_E \frac{1}{n} \left(\sum_{k=1}^n f \circ \phi^k(s) \right)^2 \mathbf{1} \left(\left(\sum_{k=1}^n f \circ \phi^k(s) \right)^2 > \frac{n\kappa}{M^2} \right) m(ds)$$

$$:= J_1^{(M)}(n) + J_2^{(M)}(n) \, .$$

It follows from (9.40) that

$$\lim_{M \to \infty} \limsup_{n \to \infty} J_1^{(M)}(n) = 0 \, ,$$

and the argument used in (9.52) shows that

$$\lim_{n \to \infty} J_2^{(M)}(n) = 0 \ \text{ for any } M > 0.$$

Therefore, (9.47) follows, and the proof of the theorem is complete. □

Remark 9.4.2. We have stated and proved Theorem 9.4.1 in terms of convergence of finite-dimensional distributions. It is clear that under appropriate conditions on the function f, the result will hold in the sense of weak convergence in the Skorokhod J_1 topology on $D[0, \infty)$. In order to keep the conditions of the theorem as simple as possible, we have avoided additional conditions on f.

Remark 9.4.3. As long as the limiting variance $\sigma_{\mathbf{X}}^2$ in (9.44) does not vanish, we should view the stationary symmetric infinitely divisible process of the type (9.37) as having short memory from the point of view of the behavior of its partial sums if it satisfies (9.40).

The assumption (9.40) implies that the map ϕ in (9.44) is dissipative, at least on the set E_f in (9.42). In the remainder of this section, we will try to shed some additional light on the role the dissipativity (or its opposite, conservativity) of ϕ play. Note that the condition (9.40) is essentially a condition on the rate of decay of the function $|f|$ over the trajectories of the flow (ϕ^n). This condition may or may not hold, even when the map ϕ is dissipative.

Example 9.4.4. The linear process (9.13) is a special case of the stationary infinitely divisible process (9.37) if the noise variables $(\varepsilon_n, \, n \in \mathbb{Z})$ are themselves infinitely divisible random variables. Assume that

$$E e^{i\theta\varepsilon_0} = \exp \left\{ \int_{-\infty}^\infty \left(e^{i\theta x} - 1 - i\theta [\![x]\!] \right) \rho(dx) \right\} \, , \ \theta \in \mathbb{R} \, , \tag{9.53}$$

for a one-dimensional symmetric Lévy measure ρ satisfying the first condition in (9.39). Then the linear process (9.13) is of the form (9.37) with $E = \mathbb{Z}$, the

control measure m of the symmetric infinitely divisible random measure M being the counting measure on $\mathbb{Z}$, and the function f given by $f(j) = \varphi_j$, $j \in \mathbb{Z}$. The map ϕ is, of course, the left shift on $\mathbb{Z}$. It is easy to verify that in this special case, the assumption (9.40) coincides with the assumption (9.14) of Theorem 9.3.1, and the conclusion (9.43) of Theorem 9.4.1 coincides with the conclusion (9.15) of Theorem 9.3.1 (apart from a different mode of convergence).

We can learn a lesson from Example 9.4.4. Namely, even when the map ϕ in (9.44) is dissipative, if the growth condition (9.40) on the kernel f fails, then the partial sums of the stationary infinitely divisible process can converge to a limit different from a Brownian motion. Theorem 9.3.3 shows, in particular, that the partial sums of an infinitely divisible infinite-moving-average process as in Example 9.4.4 can converge, with proper normalization, to a fractional Brownian motion with $H > 1/2$. It is clear that in the more general situation described by (9.37) and a dissipative map ϕ, appropriate regular variation assumptions on the kernel f will also lead to convergence to a fractional Brownian motion with $H > 1/2$. We will not present such results here. Similarly, if the growth condition (9.40) on the kernel f holds but the limiting variance $\sigma_{\mathbf{X}}^2$ in (9.44) vanishes, then under proper regular variation assumptions on the kernel f, the partial sums of the stationary infinitely divisible process will converge to a fractional Brownian motion with $H < 1/2$. Theorem 9.3.4 is an example of such a situation for an infinitely divisible infinite-moving-average process as in Example 9.4.4. We will not present more general results of this type here.

The above discussion shows that from the point of view of the behavior of the partial sums, stationary symmetric infinitely divisible processes of the type (9.37) can have either short memory or long memory if the map ϕ is dissipative. The key, in this case, is the "size" of the kernel f.

When the map ϕ in (9.37) is conservative, the "size" of the kernel f alone can no longer guarantee that the infinitely divisible process has short memory as far as the behavior of the partial sums is concerned. The following proposition shows that if the kernel f is nonnegative, then short memory according to the behavior of the partial sums is impossible.

Proposition 9.4.5. *Assume that the map ϕ in (9.37) is conservative, and that $f \geq 0$ m-a.e. on the set E_f in (9.42). If $m(E_f) > 0$, then there is no $\sigma \geq 0$ such that (9.4) holds in terms of convergence of finite-dimensional distributions.*

Proof. We may assume that $E = E_f$, and that $\rho\big((1,\infty)\big) > 0$. Suppose, to the contrary, that there is $\sigma \geq 0$ such that (9.4) holds in terms of convergence of finite-dimensional distributions. Then both (9.47) and (9.48) must hold (the former statement has to hold with $\sigma_{\mathbf{X}}$ replaced by σ). For $n \geq 1$, let

$$A_n = \left\{ s \in E : \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right| > n^{1/2} \right\} .$$

Note that the expression in (9.48) with $j = 1$ is at least $m(A_n)\rho\big((1,\infty)\big)$, so if $m(A_n)$ does not converge to zero as $n \to \infty$, then (9.48) does not hold. Assume, therefore, that $m(A_n) \to 0$ as $n \to \infty$. Let $\kappa > 0$ be so large that $\rho\big((1,\kappa)\big) > 0$, and let k be a large positive integer. Then the expression in (9.47) with $j_1 = 0$, $j_2 = 1$ cannot be smaller than

$$
\begin{aligned}
&2\rho\big((1,\kappa)\big)\frac{1}{n}\int_E \left(\sum_{k=1}^{n} f\circ\phi^k(s)\right)^2 \mathbf{1}\big(s\in A_n^c\big)\, m(ds)\\
\geq &2\rho\big((1,\kappa)\big)\frac{1}{n}\sum_{j_1=k+1}^{n}\sum_{j_2=j_1-k}^{j_1-1}\int_E f\circ\phi^{j_1}(s) f\circ\phi^{j_2}(s)\,\mathbf{1}\big(s\in A_n^c\big)\, m(ds)\\
= &2\rho\big((1,\kappa)\big)\left[\frac{n-k}{n}\sum_{i=1}^{k}\int_E f(s) f\circ\phi^{i}(s)\, m(ds)\right.\\
&\left.-\frac{1}{n}\sum_{j_1=k+1}^{n}\sum_{j_2=j_1-k}^{j_1-1}\int_E f\circ\phi^{j_1}(s) f\circ\phi^{j_2}(s)\,\mathbf{1}\big(s\in\big)\, m(ds)\right].
\end{aligned}
$$

Since ϕ preserves the measure m,

$$
m\big(\phi^j(A_n)\big) = m(A_n) \to 0 \ \text{ as } n\to\infty
$$

for every j, and hence

$$
\begin{aligned}
&\int_E f\circ\phi^{j_1}(s) f\circ\phi^{j_2}(s)\,\mathbf{1}\big(s\in A_n\big)\, m(ds)\\
&= \int_E f(s) f\circ\phi^{j_2-j_1}(s)\,\mathbf{1}\big(s\in\phi^{j_1}(A_n)\big)\, m(ds)\\
&\leq \left(\int_E f(s)^2\mathbf{1}\big(s\in\phi^{j_1}(A_n)\big)\, m(ds)\right)^{1/2}\left(\int_E f^2\circ\phi^{j_2-j_1}(s)\, m(ds)\right)^{1/2}
\end{aligned}
$$

uniformly in j_1, j_2. Therefore, the lower limit of the expression in (9.47) with $j_1 = 0$, $j_2 = 1$ is at least

$$
2\rho\big((1,\kappa)\big)\sum_{i=1}^{k}\int_E f(s) f\circ\phi^i(s)\, m(ds) = 2\rho\big((1,\kappa)\big)\int_E f(s)\sum_{i=1}^{k} f\circ\phi^i(s)\, m(ds)\,.
$$

Since this is true for an arbitrarily large k, the lower limit cannot be smaller than

$$
2\rho\big((1,\kappa)\big)\int_E f(s)\sum_{i=1}^{\infty} f\circ\phi^i(s)\, m(ds)\,.
$$

However,

$$\sum_{i=1}^{\infty} f \circ \phi^i(s) = \infty \ \ m\text{-a.e.}, \tag{9.54}$$

since ϕ is conservative and $E = E_f$; just use Exercise 2.6.12 with

$$B = \left\{ s \in E : f \circ \phi^k(s) \geq \varepsilon \ \text{ for some } k \in \mathbb{Z} \right\}$$

with $\varepsilon > 0$ becoming arbitrarily small. Therefore, (9.47) does not hold.

The resulting contradiction shows that there is no $\sigma \geq 0$ such that (9.4) holds in terms of convergence of finite-dimensional distributions. □

Remark 9.4.6. One can see from Proposition 9.4.5 that conservative maps tend to produce stationary infinitely divisible processes that do not have short memory from the point of view of the behavior of the partial sums. Of course, other factors are present. With the assumption of a nonnegative kernel f made in Proposition 9.4.5, for example, one can expect that if a functional central limit theorem of the type (9.3) holds, then the limiting process $\mathbf{Y}$ may be a fractional Brownian motion with $H > 1/2$, and hence the memory in the stationary infinitely divisible process is long. Theorem 9.4.7 demonstrates this for a special type of conservative map. Intuitively, this happens because the assumption that the kernel f is nonnegative prevents certain cancellation effects in the partial sums of an infinitely divisible process. Such cancellation effects may reduce the Hurst exponent of the limiting process $\mathbf{Y}$ if a functional central limit theorem of the type (9.3) holds. In particular, even with a conservative map ϕ, cancellation effects can result in a Brownian motion as a limit in (9.3) and hence in short memory from the point of view of the behavior of the partial sums.

In the remainder of this section, we keep the assumption of a nonnegative kernel f. In order to better understand how the properties of a conservative map ϕ affect a functional central limit theorem for the partial sums of the process, we present a result about an infinitely divisible process in (9.37) corresponding to a particular conservative map. We will see that in this special situation, even with a "very small" but nonnegative kernel, a properly normalized partial sum process converges to a fractional Brownian motion with $H > 1/2$, and hence the infinitely divisible process has long-range dependence as far as the partial sums of the process are concerned. We will consider a situation similar to that described in Example 2.4.12.

Let $(p_{ij},\ i,j \in \mathbb{Z})$ be a set of transition probabilities of an irreducible null recurrent Markov chain on $\mathbb{Z}$. Let $(\pi_i,\ i \in \mathbb{Z})$ be a σ-finite invariant measure for the Markov chain; it is necessarily unique up multiplication by a positive constant; see, e.g., Resnick (1992). Note that

$$m(A) = \sum_{i \in \mathbb{Z}} \pi_i P_i\big(\text{the trajectory of the Markov chain is in } A\big) \tag{9.55}$$

for a measurable subset A of $E = \mathbb{Z}^{\mathbb{Z}}$ defines a σ-finite measure on the cylindrical σ-field on E that is invariant under the left shift ϕ on E. We choose a very simple, and "small," function f: let

$$f(\mathbf{x}) = 1(x_0 = 0) \text{ for } \mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{Z}^{\mathbb{Z}}. \tag{9.56}$$

Note that

$$\int_E f(s)^2 \, m(ds) = \pi_0 < \infty \,,$$

so that the conditions (9.39) are satisfied, and (9.37) specifies a well-defined symmetric infinitely divisible stochastic process with finite variance. In this case, the left shift on $E = \mathbb{Z}^{\mathbb{Z}}$ is conservative; see Exercise 9.10.8.

It turns out that the behavior of the partial sums of this stationary infinitely divisible process is determined by how quickly the Markov chain returns to a given state. For $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathbb{Z}^{\mathbb{Z}}$, define

$$\tau_1 = \tau_1(\mathbf{x}) = \inf\{n \geq 1 : x_n = 0\} \,. \tag{9.57}$$

For $i \in \mathbb{Z}$, we will use the notation $P_i(\tau_1 \in \cdot)$ to describe the law of the first hitting time of the origin by the Markov chain after $n = 0$. For example,

$$P_i(\tau_1 = n) = \pi_i \sum_{j_1 \neq 0} \cdots \sum_{j_{n-1} \neq 0} p_{ij_1} \cdots p_{j_{n-2} j_{n-1}} p_{j_{n-1} 0}, \; n = 1, 2, \ldots \,.$$

A useful property of the first return times used in the proof of the next theorem is in Exercise 9.10.9.

Theorem 9.4.7. *Assume that the sequence $P_0(\tau_1 > n)$, $n = 1, 2, \ldots$, is regularly varying with exponent $-\beta \in (-1, 0)$. Then*

$$\left(\left(\frac{P_0(\tau_1 > n)}{n} \right)^{1/2} S_n(t), \; t \geq 0 \right) \Rightarrow \left(c_\beta B_H(t), \; t \geq 0 \right) \text{ as } n \to \infty \tag{9.58}$$

in the Skorokhod J_1 topology on $D[0, \infty)$, where $\mathbf{B}_H$ is the standard fractional Brownian motion with $H = (1 + \beta)/2$, satisfying $E\left(B_H(t)^2\right) = t^{2H}/2$, $t \geq 0$. Here

$$c_\beta^2 = \pi_0 \int_{-\infty}^{\infty} x^2 \, \rho(dx) \, \frac{\Gamma(1 - \beta)\Gamma(1 + 2\beta)}{\Gamma(2 + \beta)} E\left(Z_\beta^{-2\beta}\right) \,,$$

where Z_β is a positive β-stable random variable with characteristic triple $(0, \mu, 0)$, and the Lévy measure is given by $\mu(dx) = \beta x^{-(1+\beta)} \, dx$ for $x > 0$.

Proof. As in the proof of Theorem 9.4.1, for the convergence of the finite-dimensional distributions we need to prove two statements. First of all, for every $\kappa > 0$,

$$\int_E \int_{-\infty}^{\infty} x^2 \frac{P_0(\tau_1 > n)}{n} \left(\sum_{k=1}^{n} f \circ \phi^k(s) \right)^2 \mathbf{1}\left(x^2 \frac{P_0(\tau_1 > n)}{n} \left(\sum_{k=1}^{n} f \circ \phi^k(s) \right)^2 \le \kappa \right) \rho(dx)\, m(ds) \to c_\beta^2 \tag{9.59}$$

as $n \to \infty$, and also

$$\int_E \int_{-\infty}^{\infty} \mathbf{1}\left(x^2 \frac{P_0(\tau_1 > n)}{n} \left(\sum_{k=1}^{n} f \circ \phi^k(s) \right)^2 > \mathbf{1} \right) \rho(dx)\, m(ds) \to 0 \tag{9.60}$$

as $n \to \infty$.

Let N_n be the number of times the Markov chain visits state 0 in the time interval $\{1, \dots, n\}$. Then the expression in (9.59) can be written in the form

$$\int_{-\infty}^{\infty} x^2\, \rho(dx)\, \frac{P_0(\tau_1 > n)}{n} \sum_{i=-\infty}^{\infty} \pi_i E_i(N_n^2)$$

$$- \frac{P_0(\tau_1 > n)}{n} \int_{-\infty}^{\infty} x^2 \left[\sum_{i=-\infty}^{\infty} \pi_i E_i \left[N_n^2 \mathbf{1}\left(N_n^2 > \kappa x^{-2} \frac{n}{P_0(\tau_1 > n)} \right) \right] \right] \rho(dx)$$

$$:= I_1(n) - I_2(n)\,.$$

By the strong Markov property of the Markov chain and Exercise 9.10.9,

$$\sum_{i=-\infty}^{\infty} \pi_i E_i(N_n^2) = \sum_{k=1}^{n} \sum_{i=-\infty}^{\infty} \pi_i P_i(\tau_1 = k) E_0((1 + N_{n-k})^2)$$

$$= \pi_0 \sum_{k=1}^{n} P_0(\tau_1 \ge k) E_0((1 + N_{n-k})^2)\,.$$

By Exercise 9.10.10, the sequence $E_0(N_n^2)$, $n = 1, 2, \dots$, is regularly varying with exponent 2β. We can use now part (i) of Theorem 10.5.10 and, once again, Exercise 9.10.10 to conclude that

$$I_1(n) \to \frac{\Gamma(1-\beta)\Gamma(1+2\beta)}{\Gamma(2+\beta)} \pi_0 \int_{-\infty}^{\infty} x^2\, \rho(dx) E\left(Z_\beta^{-2\beta}\right) = c_\beta^2 \tag{9.61}$$

as $n \to \infty$. We will prove that $I_2(n) \to 0$ for every $\kappa > 0$. This will establish both (9.59) and (9.60). Using, once again, Exercise 9.10.9, we can write, for every $a > 0$,

$$\begin{aligned} I_2(n) &= \pi_o \frac{P_0(\tau_1 > n)}{n} \sum_{k=1}^{n} P_0(\tau_1 \geq k) \\ &\qquad \int_{-\infty}^{\infty} x^2 E_0 \left[(1 + N_{n-k})^2 \mathbf{1} \left((1 + N_{n-k})^2 > \kappa x^{-2} \frac{n}{P_0(\tau_1 > n)} \right) \right] \rho(dx) \\ &= \pi_o \frac{P_0(\tau_1 > n)}{n} \sum_{k=1}^{n} P_0(\tau_1 \geq k) \left[\int_{|x| \leq a} \cdot + \int_{|x| > a} \cdot \right] \\ &:= I_{21}^{(a)}(n) + I_{22}^{(a)}(n) \,. \end{aligned}$$

Using Exercise 9.10.10 and Theorem 10.5.6, we see, for large n and for some positive finite constant c, that

$$I_{21}^{(a)}(n) \leq c E_0 \left[\left(P_0(\tau_1 > n) N_n \right)^2 \mathbf{1} \left(\left(P_0(\tau_1 > n) N_n \right)^2 > \frac{\kappa}{4a^2} n P_0(\tau_1 > n) \right) \right] \to 0$$

as $n \to \infty$ for every $a > 0$, since $nP_0(\tau_1 > n) \to \infty$. On the other hand, the argument leading to (9.61) immediately gives us

$$\limsup_{n \to \infty} I_{21}^{(a)}(n) \leq \frac{\Gamma(1-\beta)\Gamma(1+2\beta)}{\Gamma(2+\beta)} \pi_0 \int_{|x|>a} x^2 \, \rho(dx) E\left(Z_\beta^{-2\beta} \right) ,$$

and the expression on the right-hand sides vanishes as $a \to \infty$. Therefore, $I_2(n) \to 0$, and the proof of convergence of the finite-dimensional distributions is complete. Since the limiting fractional Brownian motion satisfies $H > 1/2$, the required tightness in the Skorokhod J_1 topology on $D[0, \infty)$ follows as in the proof of Theorem 9.3.3. We leave the details to Exercise 9.10.11. □

Remark 9.4.8. The normalizing sequence $\left(n^{1/2} P_0(\tau_1 > n) \right)^{-1/2}$ in Theorem 9.4.7 is regularly varying with exponent $(1 + \beta)/2$, where $0 < \beta < 1$ describes how heavy the tail of the first return time τ_1 is. It is interesting to observe that as $\beta \to 0$, the Markov chain becomes closer to being transient, while the left shift ϕ on $\mathbb{Z}^{\mathbb{Z}}$ becomes closer to being dissipative (see Example 2.4.12). At the same time, the exponent of regular variation of the normalizing sequence becomes closer to $1/2$, and the fractional Brownian limit in Theorem 9.4.7 becomes closer to the Brownian motion. That is, as $\beta \to 0$, the stationary infinitely divisible process in the theorem becomes closer to the boundary between long memory and short memory as far as the behavior of the partial sums is concerned.

9.5 Partial Sums of Infinite-Variance Linear Processes

In this section, we begin the detailed search for phase transitions in the behavior of the partial sums of stationary stochastic processes with infinite variance. Recall Assumption 9.2.1: in this context, we will consider stationary processes $\mathbf{X} = (X_n,\, n \in \mathbb{Z})$ for which the marginal distributions have balanced regularly varying tails with exponent $\alpha \in (0, 2)$. We will assume, for simplicity, that the process $\mathbf{X}$ is symmetric, i.e., that $-\mathbf{X} \stackrel{d}{=} \mathbf{X}$; without this assumption, partial sums of the process $\mathbf{X}$ may require centering, in addition to scaling, in order to obtain convergence. The necessity of such centering will complicate the discussion without helping us to understand the phase transitions between short and long memory. In order to have a benchmark for short memory in this case, let us recall the behavior of the partial sums of an i.i.d. sequence with such balanced regularly varying tails. In this case, the analogue of the invariance principle (9.4) takes the following form. As in (4.40), the appropriate normalizing sequence is

$$a_n = \inf\bigl\{x > 0 :\, P(X_0 > x) \le 1/n\bigr\},\ n = 1, 2, \ldots\,. \tag{9.62}$$

Recall that the sequence (a_n) is regularly varying with exponent $1/\alpha$. Then, under the assumption of symmetry,

$$\bigl(a_n^{-1} S_n(t),\, t \ge 0\bigr) \Rightarrow \bigl(Y_\alpha(t),\ t \ge 0\bigr) \ \text{ as } n \to \infty \tag{9.63}$$

weakly in the Skorokhod J_1 topology on $D[0, \infty)$. Here $\bigl(Y_\alpha(t),\ t \ge 0\bigr)$ is a symmetric α-stable Lévy motion with Lévy measure μ satisfying

$$\mu(dx) = \alpha |x|^{-(\alpha+1)}\, dx\,;$$

see Gikhman and Skorohod (1996). Therefore, a different normalization from that given in (9.62) or a limit different from a stable Lévy motion is an indication of long-range dependence from the point of view of the behavior of the partial sums of symmetric stationary processes whose marginal tails are balanced regularly varying with exponent $\alpha \in (0, 2)$.

Before we proceed to study the partial sums of infinite-variance linear processes, which is the main topic of this section, we would like to mention a new general phenomenon that appears when the marginal tails of a stationary process are sufficiently heavy: the partial sums of the process can no longer grow faster than the partial sums of an i.i.d. sequence with the same marginal distributions. This is described in the following proposition.

Proposition 9.5.1. *Let X be a random variable such that $E|X|^\beta = \infty$ for some $0 < \beta < 1$. Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a stochastic process with each $X_i \stackrel{d}{=} X$, and*

let $\mathbf{Y} = (Y_1, Y_2, \ldots)$ *be a sequence of independent copies of* X*. Let* $a_n \uparrow \infty$ *be a sequence of positive numbers such that*

$$\limsup_{n\to\infty} \frac{a_{n+1}}{a_n} < \infty\,.$$

If

$$\liminf_{n\to\infty} P(|X_1 + X_2 + \ldots + X_n| > a_n) > 0, \tag{9.64}$$

then also

$$\limsup_{n\to\infty} P\big(|Y_1 + Y_2 + \ldots + Y_n| > a_n/5\big) > 0\,. \tag{9.65}$$

If X *is symmetric, then*

$$\limsup_{n\to\infty} P(|Y_1 + Y_2 + \ldots + Y_n| > a_n) > 0\,. \tag{9.66}$$

Proof. Note that

$$\begin{aligned}
&P(|X_1 + X_2 + \ldots + X_n| > a_n)\\
&\le P\Big(\max_{i=1,\ldots,n} |X_i| > a_n\Big) + P\Big(\Big|\sum_{i=1}^{n} X_i \mathbf{1}(|X_i| \le a_n)\Big| > a_n\Big)\\
&\le nP(|X| > a_n) + \frac{nE\big(|X|\mathbf{1}(|X| \le a_n)\big)}{a_n}\,,
\end{aligned}$$

using Markov's inequality at the last step. Using Lemma 9.5.3 below and the growth assumption on the sequence (a_n) (in order to switch from the limit over a continuous variable to a limit over a discrete variable), we see that

$$\limsup_{n\to\infty} \frac{a_n P(|X| > a_n)}{E\big(|X|\mathbf{1}(|X| \le a_n)\big)} > 0\,,$$

and therefore,

$$\limsup_{n\to\infty} nP(|X| > a_n) > 0\,.$$

Choose $\delta > 0$ and a sequence $n_k \uparrow \infty$ such that

$$n_k P\big(P|X| > a_{n_k}\big) \ge \delta \ \text{ for all } k \ge 1. \tag{9.67}$$

Suppose first that X is a symmetric random variable. By symmetry and independence, we have

$$\begin{aligned} &P(|Y_1 + Y_2 + \ldots + Y_{n_k}| > a_{n_k}) \geq \frac{1}{2} P\Big(\max_{i=1,\ldots,n_k} |Y_i| > a_{n_k}\Big) \\ =&\frac{1}{2}\left(1 - \Big(1 - P\big(|X| > a_{n_k}\big)\Big)^{n_k}\right) \\ \geq&\frac{1}{2}\left(1 - \left(1 - \frac{\delta}{n_k}\right)^{n_k}\right) \to \frac{1}{2}\Big(1 - e^{-\delta}\Big) > 0 \end{aligned} \tag{9.68}$$

as $k \to \infty$. This proves (9.66).

Without assuming symmetry, suppose that to the contrary, (9.65) fails. By the Lévy–Ottaviani inequality of Theorem 10.7.3 with $s = t = a_n/5$, we have

$$\lim_{n\to\infty} P\Big(\max_{i=1,\ldots,n} |Y_1 + Y_2 + \ldots + Y_n| > 2a_n/5\Big) = 0\,. \tag{9.69}$$

For $n = 1, 2 \ldots$, define $N_n = \inf\{k \geq 1 : |Y_k| > a_n\}$. With the sequence (n_k) as in (9.67), we have

$$\begin{aligned} &P\big(|Y_1 + Y_2 + \ldots + Y_{n_k}| > a_{n_k}/5\big) \\ \geq\, &P\big(N_{n_k} \leq n_k,\ \max_{i\leq N_{n_k}} |Y_1 + \ldots + Y_i| \leq 2a_{n_k}/5,\ |Y_{N_{n_k}+1} + \ldots + Y_{n_k}| \leq 2a_{n_k}/5\big) \\ \geq\, &\sum_{j=1}^{n_k} P(N_{n_k} = j)\, P(|Y_{j+1} + \ldots + Y_{n_k}| \leq 2a_{n_k}/5) \\ &- P\Big(\max_{i=1,\ldots,n_k} |Y_1 + Y_2 + \ldots + Y_{n_k}| > 2a_{n_k}/5\Big) \\ \geq\, &P(N_{n_k} \leq n_k)\Big(1 - P\Big(\max_{i=1,\ldots,n_k} |Y_1 + Y_2 + \ldots + Y_{n_k}| > 2a_{n_k}/5\Big)\Big) \\ &- P\Big(\max_{i=1,\ldots,n_k} |Y_1 + Y_2 + \ldots + Y_{n_k}| > 2a_{n_k}/5\Big)\,. \end{aligned}$$

However, the computation in (9.68) and the assumption (9.69) imply that the above expression cannot converge to zero. This contradicts our assumption that (9.65) fails. □

The proof of the proposition uses Lemma 9.5.3 below. First we need a preliminary estimate.

Lemma 9.5.2. *Let (a_n) be a sequence of positive numbers such that*

$$\lim_{n\to\infty} \frac{a_n}{1 + \sum_{k=1}^{n-1} a_k} = 0\,. \tag{9.70}$$

Then for every $\theta > 0$,

$$\lim_{k\to\infty} 2^{-\theta k} a_k = 0\,. \tag{9.71}$$

Proof. If the statement of the lemma fails, then for some $\theta > 0$, we have

$$\limsup_{k\to\infty} 2^{-\theta k} a_k = \infty\,.$$

Fixing such a θ, define, for $B > 0$,

$$n_B = \min\bigl\{k \ge 1 :\ 2^{-\theta k} a_k > B\bigr\}\,,$$

and note that $n_B \uparrow \infty$ as $B \uparrow \infty$. Then

$$\frac{a_{n_B}}{1 + \sum_{k=1}^{n_B-1} a_k} > \frac{B 2^{\theta n_B}}{1 + \sum_{k=1}^{n_B-1} B 2^{\theta k}} \ge \frac{2^\theta - 1}{2}$$

for every $B \ge 2^\theta - 1$, and this contradicts (9.70). □

Lemma 9.5.3. *Let X be a positive random variable such that $EX^\beta = \infty$ for some $0 < \beta < 1$. Then*

$$\limsup_{x\to\infty} \frac{xP(X > x)}{E\bigl(X\mathbf{1}(X \le x)\bigr)} > 0\,.$$

Proof. For $m \ge 1$,

$$\begin{aligned} E\bigl(X\mathbf{1}(X \le 2^m)\bigr) &\le \int_0^{2^m} P(X > y)\,dy \\ &\le 1 + \sum_{n=0}^{m-1} 2^n P\bigl(X > 2^n\bigr)\,. \end{aligned}$$

If the statement of the lemma fails, then we can use Lemma 9.5.2 with $a_n = 2^n P\bigl(X > 2^n\bigr)$ for $n \ge 1$ to conclude that

$$P\bigl(X > 2^n\bigr) = o\Bigl(2^{-\beta n}\Bigr) \text{ as } n \to \infty$$

for every $0 < \beta < 1$, which contradicts the assumption $EX^\beta = \infty$ for some $0 < \beta < 1$. □

Remark 9.5.4. An immediate conclusion from Proposition 9.5.1 is as follows. Let $\mathbf{X}$ be a symmetric stationary process whose marginal tails are balanced regularly varying with exponent $\alpha \in (0, 1)$. Then it is impossible that there is a sequence (b_n)

satisfying $b_n/a_n \to \infty$ (with a_n defined by (9.62)) such that a functional central limit theorem of the type (9.3) holds with a nonzero limit process $\mathbf{Y}$. That is, any deviations of the rate of the growth of the partial sums of the process from the rate of growth of the partial sums of the corresponding i.i.d. sequence (specified by a_n) must be in the direction of a slower rate of growth. This is a point worth keeping in mind when looking for phase transitions between short memory and long memory. It is sometimes claimed that only a faster rate of growth of the partial sums qualifies as an indication of long memory. We see now that this is impossible with sufficiently heavy tails. From the point of view of a phase transition, a deviation in the direction of a rate of growth of the partial sums slower than a_n should be viewed as an indication of long memory as well.

Let us consider now the partial sums of the infinite-variance linear processes of Section 1.4. We will assume that the noise variables $(\varepsilon_n,\ n \in \mathbb{Z})$ are i.i.d. symmetric random variables, with balanced regularly varying tails with exponent $0 < \alpha < 2$. Further, we assume that the coefficients $(\varphi_n,\ n \in \mathbb{Z})$ satisfy

$$\sum_{n=-\infty}^{\infty} |\varphi_n|^{\alpha-\varepsilon} < \infty \ \text{ for some } 0 < \varepsilon < \alpha\,. \tag{9.72}$$

It follows by Corollary 4.2.12 that the process (9.13) is a well-defined symmetric stationary process such that

$$\lim_{x\to\infty} \frac{P(X_0 > x)}{P(|\varepsilon_0| > x)} = \frac{1}{2} \sum_{n=-\infty}^{\infty} |\varphi_n|^{\alpha}\,. \tag{9.73}$$

We begin a search for phase transitions in the behavior of the partial sums of such processes by proving an analogue of Theorem 9.3.1.

Theorem 9.5.5. *Let* $\mathbf{X}$ *be a symmetric infinite-variance infinite-moving-average process as just defined. If* $1 < \alpha < 2$*, assume that the coefficients* (φ_n) *satisfy the absolute summability assumption* (9.14)*. Let* a_n *be defined by* (9.62)*. Then*

$$\left(a_n^{-1} S_n(t),\ t \geq 0\right) \Rightarrow \left(\frac{a_\varphi}{\|\varphi\|_\alpha} Y_\alpha(t),\ t \geq 0\right) \ \text{ as } n \to \infty \tag{9.74}$$

in finite-dimensional distributions, where $\mathbf{Y}_\alpha$ *is the symmetric* α*-stable Lévy motion in* (9.63)*,* a_φ *is given by* (9.16)*, and*

$$\|\varphi\|_\alpha = \left(\sum_{n=-\infty}^{\infty} |\varphi_n|^{\alpha}\right)^{1/\alpha}\,.$$

Proof. Let

$$\hat{a}_n = \inf\left\{x > 0 :\ P(\varepsilon_0 > x) \leq 1/n\right\},\ n = 1, 2, \ldots\,.$$

It follows from (9.73) that $a_n/\hat{a}_n \to \|\varphi\|_\alpha$ as $n \to \infty$. Therefore, the statement (9.74) is equivalent to the statement

$$\left(\hat{a}_n^{-1} S_n(t),\ t \geq 0\right) \Rightarrow \left(a_\varphi Y_\alpha(t),\ t \geq 0\right) \quad \text{as } n \to \infty \tag{9.75}$$

in finite-dimensional distributions, which we now proceed to prove. We will use the decomposition (9.18). By the central limit theorem (9.63), the claim (9.75) will follow once we show that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\hat{a}_n^{-1}\left|B_n\right| > \epsilon\right) = \lim_{n\to\infty} P\left(\hat{a}_n^{-1}\left|C_n\right| > \epsilon\right) = 0\,. \tag{9.76}$$

For the first statement in (9.76), we note that for every $M = 1, 2, \ldots,$

$$\limsup_{n\to\infty} P\left(\hat{a}_n^{-1}\left|B_n\right| > \epsilon\right) \leq \limsup_{n\to\infty} P\left(\hat{a}_n^{-1}\left|\sum_{j=-n+M}^{n-M} \theta_{j,n}\varepsilon_j\right| > \frac{\epsilon}{2}\right),$$

where

$$\theta_{j,n} = \sum_{k:\,|k+j|>n} \varphi_k\,.$$

The assumptions of the theorem imply that the absolute summability assumption (9.14) holds regardless of the value of α. Therefore, for every $\gamma > 0$, we have $|\theta_{j,n}| \leq \gamma$ for all n and all $-n + M \leq j \leq n - M$ if M is large enough. By the contraction principle in Theorem 10.7.5, we conclude that for such M,

$$\begin{aligned}\limsup_{n\to\infty} P\left(\hat{a}_n^{-1}\left|B_n\right| > \epsilon\right) &\leq 2\limsup_{n\to\infty} P\left(\hat{a}_n^{-1}\left|\sum_{j=-n+M}^{n-M} \varepsilon_j\right| > \frac{\epsilon}{2\gamma}\right)\\ &= 2P\left(\left|Y_\alpha(2)\right| > \frac{\epsilon}{2\gamma}\right),\end{aligned}$$

where at the last step we used (9.63). Letting $\gamma \to 0$, we obtain the first statement in (9.76).

For the second statement in (9.76), we write

$$C_n = \sum_{j=n+1}^{\infty}\left(\sum_{k=-n-j}^{n-j} \varphi_k\right)\varepsilon_j + \sum_{j=-\infty}^{-n-1}\left(\sum_{k=-n-j}^{n-j} \varphi_k\right)\varepsilon_j := C_{n,1} + C_{n,2}$$

and prove that for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\hat{a}_n^{-1}\left|C_{n,j}\right| > \epsilon\right) = 0,\ j = 1, 2\,. \tag{9.77}$$

We will consider only the case $j = 1$; the case $j = 2$ is similar. Note that

$$P\left(\hat{a}_n^{-1}|C_{n,1}| > \epsilon\right) = P\left(\left|\sum_{j=1}^{\infty}\left(\sum_{k=j}^{2n+j}\varphi_{-k}\right)\varepsilon_j\right| > \epsilon\hat{a}_n\right)$$

$$\leq P\left(\left|\varepsilon_j\sum_{k=j}^{2n+j}\varphi_{-k}\right| > \hat{a}_n \text{ for some } j = 1, 2, \ldots\right)$$

$$+ P\left(\left|\sum_{j=1}^{\infty}\left(\sum_{k=j}^{2n+j}\varphi_{-k}\right)\varepsilon_j \mathbf{1}\left(\left|\varepsilon_j\sum_{k=j}^{2n+j}\varphi_{-k}\right| \leq \hat{a}_n\right)\right| > \epsilon\hat{a}_n\right)$$

$$:= p_{n,1} + p_{n,2}\,.$$

We now use the Potter bounds of Corollary 10.5.8 to check that for every $\theta > 0$, for all n large enough,

$$p_{n,1} \leq \sum_{j=1}^{\infty} P\left(|\varepsilon_0| > \frac{\hat{a}_n}{\left|\sum_{k=j}^{2n+j}\varphi_{-k}\right|}\right)$$

$$\leq (1+\theta)\sum_{j=1}^{\infty} P(|\varepsilon_0| > \hat{a}_n)\left|\sum_{k=j}^{2n+j}\varphi_{-k}\right|^{\alpha-\theta}$$

$$\leq 3(1+\theta)\, n^{-1}\sum_{j=1}^{\infty}\left|\sum_{k=j}^{2n+j}\varphi_{-k}\right|^{\alpha-\theta}, \tag{9.78}$$

with the definition of $\hat{a}_n$ being used in the last step. We will prove that if $\theta > 0$ is small enough, the expression in (9.78) converges to zero as $n \to \infty$.

Suppose first that $1 < \alpha < 2$, and choose $0 < \theta < \alpha - 1$ in (9.78). For every $M = 1, 2, \ldots$, we have

$$\limsup_{n\to\infty} n^{-1}\sum_{j=1}^{\infty}\left|\sum_{k=j}^{2n+j}\varphi_{-k}\right|^{\alpha-\theta} = \limsup_{n\to\infty} n^{-1}\sum_{j=M}^{\infty}\left|\sum_{k=j}^{2n+j}\varphi_{-k}\right|^{\alpha-\theta}$$

$$\leq \left(\sum_{j=-\infty}^{-M}|\varphi_k|\right)^{\alpha-\theta-1}\limsup_{n\to\infty} n^{-1}\sum_{j=-\infty}^{\infty}\sum_{k=j}^{2n+j}|\varphi_{-k}|$$

$$= 2\left(\sum_{j=-\infty}^{-M}|\varphi_k|\right)^{\alpha-\theta-1}\sum_{j=-\infty}^{\infty}|\varphi_j|\,,$$

and the last expression converges to zero as $M \to \infty$. If $0 < \alpha \leq 1$, choose in (9.78) $0 < \theta < \epsilon$ for $\epsilon > 0$ for which (9.72) holds. Then for every $M = 1, 2, \ldots$, we have

$$\sum_{j=1}^{\infty} \left| \sum_{k=j}^{2n+j} \varphi_{-k} \right|^{\alpha-\theta} \leq \sum_{j=1}^{\infty} \left(\sum_{k=j}^{2n+j} |\varphi_{-k}|^{\alpha-\epsilon} \right)^{\frac{\alpha-\theta}{\alpha-\epsilon}},$$

and now we proceed as in the case $1 < \alpha < 2$ above.

Finally, by the Cauchy–Schwarz inequality and Proposition 4.2.3, for large n,

$$\begin{aligned} p_{n,2} &\leq \epsilon^{-2} \hat{a}_n^{-2} \sum_{j=1}^{\infty} \left(\sum_{k=j}^{2n+j} \varphi_{-k} \right)^2 E \left[\varepsilon_0 \mathbf{1} \left(\left| \varepsilon_0 \sum_{k=j}^{2n+j} \varphi_{-k} \right| \leq \hat{a}_n \right) \right]^2 \\ &\leq \frac{4}{(2-\alpha)\epsilon^2} \hat{a}_n^{-2} \sum_{j=1}^{\infty} \left(\sum_{k=j}^{2n+j} \varphi_{-k} \right)^2 \hat{a}_n^2 \left(\sum_{k=j}^{2n+j} \varphi_{-k} \right)^{-2} P \left(\left| \varepsilon_0 \sum_{k=j}^{2n+j} \varphi_{-k} \right| > \hat{a}_n \right) \\ &= \frac{4}{(2-\alpha)\epsilon^2} \sum_{j=1}^{\infty} P \left(|\varepsilon_0| > \frac{\hat{a}_n}{\left| \sum_{k=j}^{2n+j} \varphi_{-k} \right|} \right), \end{aligned}$$

reducing the problem to the situation already considered in (9.78). □

Remark 9.5.6. The conclusion from Theorem 9.5.5 is, to a certain extent, similar to the conclusion from Theorem 9.3.1 discussed in Remark 9.3.2: if a symmetric infinite-variance linear process with regularly varying tails with exponent $0 < \alpha < 2$ has coefficients satisfying (9.14) and $a_\varphi \neq 0$, then it satisfies a functional central limit theorem (9.3) with an α-stable Lévy process as a limit. Such a process should, therefore, be regarded as having short memory from the point of view of the behavior of the partial sums. Note, however, that the assumption (9.14) is an additional assumption only in the case $1 < \alpha < 2$, while in the case $0 < \alpha \leq 1$, it is already implied by the assumption (9.72) of the model. Therefore, in the latter case, the only possible appearance of long-range dependence with respect to the behavior of the partial sums is due to the presence of cancellations when $a_\varphi = 0$, and the partial sums grow at a rate slower than that dictated by a_n in (9.62). We see, once again, the phenomenon described in Remark 9.5.4.

If $1 < \alpha < 2$, then the partial sums of the infinite-variance infinite-moving-average process can grow at a rate faster than a_n given by (9.62) if the absolute summability assumption (9.14) on the coefficients fails. As in the finite-variance case considered in Section 9.3, we will impose a balanced regular variation assumption. That is, we will assume that (9.22) holds, but this time, the exponent of regular variation of the sequence (b_n) must be in the range $\beta \in (-1, -1/\alpha)$. We should not be surprised that in this case, the properly normalized partial sums converge to a limit different from the α-stable Lévy process as in (9.63). It is,

perhaps, less clear what self-similar process with stationary increments one obtains as a limit. The marginal tail behavior of the moving-average process leads us to expect that the limiting process might itself be α-stable. However, there are many different α-stable self-similar processes with stationary increments, some of which are described in Section 3.5 and in Example 8.4.1. Interestingly, in the next result, the linear fractional symmetric stable motion of Example 3.5.2 makes an appearance.

Theorem 9.5.7. *Let* $\mathbf{X}$ *be the symmetric infinite-variance infinite-moving-average process of Theorem 9.5.5 with* $1 < \alpha < 2$. *Suppose that the coefficients* (φ_n) *satisfy* (9.22), $-1 < \beta < -1/\alpha$. *Then as* $n \to \infty$,

$$\left(\frac{1}{na_nb_n}S_n(t),\ t \geq 0\right) \Rightarrow \left(C_\alpha^{-1/\alpha}(1+\beta)^{-1}X^{(\mathrm{L})}_{c_+,-c_-,1+1/\alpha+\beta}(t),\ t \geq 0\right) \tag{9.79}$$

weakly in the Skorokhod J_1 *topology on* $D[0,\infty)$, *where* $\mathbf{X}^{(\mathrm{L})}_{c_+,-c_-,1+1/\alpha+\beta}$ *is the linear fractional symmetric stable motion defined by* (3.81) *with* $c_1 = c_+$, $c_2 = -c_-$ *and* $H = 1 + 1/\alpha + \beta$. *Furthermore, the constant* C_α *is given by* (3.31).

Proof. We first prove convergence of the finite-dimensional distributions. Fix $0 < t_1 < t_2 < \ldots < t_d$. We begin with the expression for the partial sum given in (9.25) and proceed through two approximations. Let M and m be two large positive integers. Let

$$T_l^{(m,M)}(n) = \sum_{k=-mM}^{mM-1}\left(\sum_{-kn/m \leq i < nt_l - kn/m}\varphi_i\right)\sum_{kn/m \leq j < (k+1)n/m}\varepsilon_j,\ l = 1,\ldots,d\,.$$

We begin by showing that as $n \to \infty$,

$$\begin{aligned}&\left(\frac{1}{na_nb_n}T_1^{(m,M)}(n),\ldots,\frac{1}{na_nb_n}T_d^{(m,M)}(n)\right) \\ &\Rightarrow \left(C_\alpha^{-1/\alpha}(1+\beta)^{-1}T^{(m,M)}(t_1),\ldots,C_\alpha^{-1/\alpha}(1+\beta)^{-1}T^{(m,M)}(t_d)\right),\end{aligned} \tag{9.80}$$

where

$$\begin{aligned}T^{(m,M)}(t) = \int_{-M}^{M}\Big\{&c_+\Big[\big((t-\phi_m(x))_+\big)^{1+\beta} - \big((-\phi_m(x))_+\big)^{1+\beta}\Big] \\ &- c_-\Big[\big((t-\phi_m(x))_-\big)^{1+\beta} - \big((-\phi_m(x))_-\big)^{1+\beta}\Big]\Big\}\,L(dx),\ t \in \mathbb{R}\,,\end{aligned}$$

and $\phi_m(x) = k/m$ if $k/m \leq x < (k+1)/m$. To this end, for $t > 0$ and $\tau \in \mathbb{R}$, let

$$A_n(t,\tau) = \sum_{-\tau n \leq i \leq (t-\tau)n}\varphi_i,\ = n = 1,2,\ldots\,.$$

An argument similar to the one we used to prove (9.26) shows that

$$\lim_{n\to\infty} \frac{1}{nb_n} A_n(t,\tau) = \frac{1}{1+\beta}\Big\{c_+\Big[\big((t-\tau)_+\big)^{1+\beta} - \big((-\tau)_+\big)^{1+\beta}\Big] - c_-\Big[\big((t-\tau)_-\big)^{1+\beta} - \big((-\tau)_-\big)^{1+\beta}\Big]\Big\}; \tag{9.81}$$

we leave the details to Exercise 9.10.12. Since

$$\frac{1}{na_nb_n} T_l^{(m,M)}(n) = \sum_{k=-mM}^{mM-1} \left(\frac{1}{nb_n} A_n(t_l, k/m)\right)\left(\frac{1}{a_n} \sum_{kn/m \le j < (k+1)n/m} \varepsilon_j\right),$$

$l = 1, \ldots, d$, it follows from (9.81), the central limit theorem (9.63), the regular variation of a_n, and the continuous mapping theorem that the random vector on the left-hand side of (9.80) converges weakly, as $n \to \infty$, to the random vector

$$m^{-1/\alpha} \sum_{k=-mM}^{mM-1} Y_k\, \mathbf{a}(k/m)\,,$$

where (Y_k) are i.i.d. SαS random variables each distributed as the limiting process $\mathbf{Y}_\alpha$ in (9.63) evaluated at time $t = 1$, and for $\tau \in \mathbb{R}$, $\mathbf{a}(\tau)$ is a d-dimensional vector whose lth component is given by

$$\frac{1}{1+\beta}\Big\{c_+\Big[\big((t_l-\tau)_+\big)^{1+\beta} - \big((-\tau)_+\big)^{1+\beta}\Big] - c_-\Big[\big((t_l-\tau)_-\big)^{1+\beta} - \big((-\tau)_-\big)^{1+\beta}\Big]\Big\},$$

$l = 1, \ldots, d$. Elementary properties of the integral of simple functions show that this limit has the same law as the vector on the right-hand side of (9.80). Therefore, the latter claim has been established.

We remove now one level of approximation and prove that

$$\begin{aligned} &\left(\frac{1}{na_nb_n} T_1^{(M)}(n), \ldots, \frac{1}{na_nb_n} T_d^{(M)}(n)\right) \\ &\Rightarrow \left(C_\alpha^{-1/\alpha}(1+\beta)^{-1} T^{(M)}(t_1), \ldots, C_\alpha^{-1/\alpha}(1+\beta)^{-1} T^{(M)}(t_d)\right), \end{aligned} \tag{9.82}$$

where

$$T_l^{(M)}(n) = \sum_{j=-nM}^{nM-1} \left(\sum_{1-j \le i \le nt_l - j} \varphi_i\right) \varepsilon_j,\ l = 1, \ldots, d\,,$$

and

$$T^{(M)}(t) = \int_{-M}^{M} \Big\{ c_+ \Big[\big((t-x)_+\big)^{1+\beta} - \big((-x)_+\big)^{1+\beta} \Big] \\ - c_- \Big[\big((t-x)_-\big)^{1+\beta} - \big((-x)_-\big)^{1+\beta} \Big] \Big\} L(dx), \ t \in \mathbb{R}\,.$$

Since (9.80) has already been proved, we can let $m \to \infty$ in order to use Theorem 3.2 in Billingsley (1999). It follows from Theorem 3.3.2 that the random vector on the right-hand side of (9.80) converges weakly, as $m \to \infty$, to the random vector on the right-hand side of (9.82). Therefore, in order to use Theorem 3.2 in Billingsley (1999), we need to estimate the difference between the random vectors on the left-hand sides of (9.80) and (9.82). Note that for $\epsilon > 0$ and $l = 1, \ldots, d$,

$$P\left(\left| \frac{1}{na_n b_n} T_l^{(m,M)}(n) - \frac{1}{na_n b_n} T_l^{(M)}(n) \right| > \epsilon \right) = P\left(\left| \sum_{j=-nM}^{nM-1} b_{j,m}(n) \varepsilon_j \right| > \epsilon n a_n b_n \right),$$

where

$$b_{j,m}(n) = \left(\sum_{1-j \le i \le nt_l - j} \varphi_i - \sum_{-n[mj/n]/m \le i < nt_l - n[mj/n]/m} \varphi_i \right).$$

Since by Theorem 10.5.6, there is a constant C such that

$$\sup_{j \in \mathbb{Z}} \left| \sum_{i=j+1}^{j+n} \varphi_i \right| \le Cnb_n \ \text{ for } n \ge 1,$$

we see that for all j,

$$\big| b_{j,m}(n) \big| \le 2C(n/m) b_{[n/m]}\,.$$

We write

$$P\left(\left| \sum_{j=-nM}^{nM-1} b_{j,m}(n) \varepsilon_j \right| > \epsilon n a_n b_n \right)$$

$$\le P\left(|\varepsilon_j| > \frac{n a_n b_n}{|b_{j,m}(n)|} \ \text{ for some } -nM \le j < nM \right)$$

$$+ P\left(\left| \sum_{j=-nM}^{nM-1} b_{j,m}(n) \varepsilon_j \mathbf{1}\left(|\varepsilon_j| \le \frac{n a_n b_n}{|b_{j,m}(n)|} \right) \right| > \epsilon n a_n b_n \right).$$

As in the proof of (9.77) above, it is enough to consider the first term on the right-hand side. This term can be bounded for large n by

$$\begin{aligned}\sum_{j=-nM}^{nM-1} P\left(|\varepsilon_0| > \frac{na_nb_n}{|b_{j,m}(n)|}\right) &\le 2nMP\left(|\varepsilon_0| > (2C)^{-1}a_nm(b_n/b_{[n/m]})\right)\\ &\le 2nMP\left(|\varepsilon_0| > (4C)^{-1}a_nm^{1+\beta}\right)\\ &\le 3M\frac{P\left(|\varepsilon_0| > (4C)^{-1}a_nm^{1+\beta}\right)}{P(|\varepsilon_0| > a_n)}\\ &\le 4M\left((4C)^{-1}m^{1+\beta}\right)^{-\alpha/2},\end{aligned}$$

where at the last step, we used the Potter bounds of Corollary 10.5.8. Since $\beta > -1$, the last expression converges to zero as $m \to \infty$. We conclude that

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left(\left|\frac{1}{na_nb_n}T_l^{(m,M)}(n) - \frac{1}{na_nb_n}T_l^{(M)}(n)\right| > \epsilon\right) = 0$$

for every $\epsilon > 0$, and so Theorem 3.2 in Billingsley (1999) applies. That is, (9.82) follows.

In order to remove the second level of approximation and prove that

$$\begin{aligned}&\left(\frac{1}{na_nb_n}S_n(t_1),\dots,\frac{1}{na_nb_n}S_n(t_d)\right) \qquad (9.83)\\ &\Rightarrow \left(C_\alpha^{-1/\alpha}(1+\beta)^{-1}X^{(\mathrm{L})}_{c_+,-c_-,1+1/\alpha+\beta}(t_1),\dots,C_\alpha^{-1/\alpha}(1+\beta)^{-1}X^{(\mathrm{L})}_{c_+,-c_-,1+1/\alpha+\beta}(t_d)\right),\end{aligned}$$

we use, once again, Theorem 3.2 in Billingsley (1999). We know by Theorem 3.3.2 that the random vector on the right-hand side of (9.82) converges weakly, as $m \to \infty$, to the random vector in the right-hand side of (9.83), so it remains to compare the vectors on the respective left-hand sides and prove that for every $\epsilon > 0$,

$$\begin{aligned}&\lim_{M\to\infty}\limsup_{n\to\infty} P\left(\left|\sum_{j=nM}^{\infty}\left(\sum_{1-j\le i\le nt_l-j}\varphi_i\right)\varepsilon_j\right| > \epsilon\right) \qquad (9.84)\\ &= \lim_{M\to\infty}\limsup_{n\to\infty} P\left(\left|\sum_{j=-\infty}^{-nM-1}\left(\sum_{1-j\le i\le nt_l-j}\varphi_i\right)\varepsilon_j\right| > \epsilon\right) = 0\,.\end{aligned}$$

The two statements are similar, so we prove only the first one. As we have seen in the proof of (9.77) above, the argument reduces to proving that

$$\lim_{M\to\infty}\limsup_{n\to\infty}\sum_{j=nM}^{\infty} P\left(|\varepsilon_0| > \frac{na_nb_n}{\left|\sum_{1-j\le i\le nt_l-j}\varphi_i\right|}\right) = 0\,.$$

Let $0 < \gamma < \alpha + 1/\beta$. Using the Potter bounds of Corollary 10.5.8, the sum above can be bounded for large n by

$$2n^{-1} \sum_{j=nM}^{\infty} \frac{1}{P(|\varepsilon_0| > a_n)} P\left(|\varepsilon_0| > \frac{na_n b_n}{\left| \sum_{1-j \le i \le nt_l - j} \varphi_i \right|} \right)$$
$$\le 4n^{-1} \sum_{j=nM}^{\infty} \left(\frac{nb_n}{\left| \sum_{1-j \le i \le nt_l - j} \varphi_i \right|} \right)^{-(\alpha-\gamma)} .$$

Using the regular variation assumption (9.22) and the comparison with a monotone regularly varying function in Exercise 10.9.9 shows that for large n,

$$\sum_{j=nM}^{\infty} \left| \sum_{1-j \le i \le nt_l - j} \varphi_i \right|^{\alpha-\gamma} \le 2 \max(c_-, c_+) \sum_{j=nM}^{\infty} (nb_{j-n})^{\alpha-\gamma}$$
$$\sim 2 \max(c_-, c_+)\, n^{\alpha-\gamma} (nM) \left(b_{n(M-1)}\right)^{\alpha-\gamma}$$

as $n \to \infty$. Therefore,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sum_{j=nM}^{\infty} P\left(|\varepsilon_0| > \frac{na_n b_n}{\left| \sum_{1-j \le i \le nt_l - j} \varphi_i \right|} \right)$$
$$\le 8 \max(c_-, c_+) \lim_{M\to\infty} \limsup_{n\to\infty} M \left(\frac{b_{n(M-1)}}{b_n} \right)^{\alpha-\gamma}$$
$$= 8 \max(c_-, c_+) \lim_{M\to\infty} M\,(M-1)^{\beta(\alpha-\gamma)} = 0$$

by the choice of γ. Therefore, (9.84) follows, and we have proved convergence of the finite-dimensional distributions.

Since $\beta > -1$, the limiting process satisfies $H > 1/\alpha$, and one can prove tightness in the Skorokhod J_1 topology on $D[0, \infty)$ following a similar line of argument as in the proof of Theorem 9.3.3. We leave the details to Exercise 9.10.13. □

Since in the situation of Theorem 9.5.7, a functional central limit theorem of the type (9.3) holds, but the limit is not a, SαS Lévy motion but rather a linear fractional SαS motion with exponent $H > 1/\alpha$ (which is, further, a function of the exponent of the regular variation of the coefficients, in addition to the tail parameter α), we should view such an infinite-variance linear process as a long-memory process from the point of view of behavior of the partial sums. Recall that this phenomenon is possible only in the case $1 < \alpha < 2$. In the case $0 < \alpha \le 1$, long-range dependence from the point of view of the behavior of the partial sums is possible only in the

case of "negative dependence," when the coefficients (φ_n) are summable but their sum vanishes. This situation is possible in the entire range $0 < \alpha < 2$. We will, once again, assume balanced regular variation of the coefficients, but this time, the exponent of regular variation will be in the range $\beta \in (-1-1/\alpha, -1)$. The following theorem can be proved in the same way as Theorem 9.5.7, and its proof is left to Exercise 9.10.14. Note that a linear fractional SαS motion still appears, but this time, we have $H < 1/\alpha$.

Theorem 9.5.8. *Let* $\mathbf{X}$ *be the symmetric infinite-variance infinite-moving-average process of Theorem 9.5.5 with* $0 < \alpha < 2$. *Suppose that the coefficients* (φ_n) *satisfy* (9.22), $-1 - 1/\alpha < \beta < -1$. *Then as* $n \to \infty$, (9.79) *holds in terms of convergence of finite-dimensional distributions.*

Remark 9.5.9. We can summarize the results of Theorems 9.5.5, 9.5.7, and 9.5.8 in a way similar to the discussion in Remark 9.3.5 in the finite-variance case. Summability of the coefficients (9.14) and nonvanishing sum a_φ of the coefficients lead to short memory as far as the behavior of its partial sums of the process is concerned, while the regular variation (9.22) of the coefficients with $\beta \in (-1, -1/\alpha)$, or with $-1 - 1/\alpha < \beta < -1$ but $a_\varphi = 0$, leads, from the same point of view, to long memory. Once again, we have not attempted to draw the exact boundary between short and long memory.

9.6 Partial Sums of Infinite-Variance Infinitely Divisible Processes

In this section, we investigate phase transitions in the behavior of the partial sums of symmetric stationary infinitely divisible processes $\mathbf{X}$ given in the form (9.37). As in Section 9.5, we will assume that the marginal distributions of the process $\mathbf{X}$ have balanced regularly varying tails with exponent $\alpha \in (0, 2)$. Under the assumption of the symmetry of the Lévy measure ρ in the integral representation (9.37) of the process, this regular variation of the marginal tails of $\mathbf{X}$ is equivalent to the following condition:

$$\text{the function } H(y) = (m \times \rho)\Big(\big\{(s,x) \in E \times \mathbb{R} : xf(s) > y\big\}\Big) \tag{9.85}$$

$$\text{is regularly varying with exponent } -\alpha \in (-2, 0).$$

Moreover, under this condition,

$$\lim_{y \to \infty} \frac{P(X_0 > y)}{H(y)} = 1 \tag{9.86}$$

(this is a special case of equivalence results between the tails of functionals of infinitely divisible processes and the corresponding tails with respect to the Lévy measure of Rosiński and Samorodnitsky (1993)).

As in the case of the finite-variance stationary infinitely divisible processes considered in Section 9.4, the length of memory of the process $\mathbf{X}$ is determined by the local Lévy measure ρ, the kernel f, and ergodic-theoretical properties of the map ϕ with respect to the control measure m. It turns out that for stationary processes with regularly varying tails with exponent $\alpha \in (0, 2)$, an important ingredient is the tail behavior of the local Lévy measure ρ, which itself is often regularly varying with exponent $-\alpha$. The "size" of the function f and the ergodic-theoretical properties of the map ϕ are still of crucial importance as far as the length of the memory is concerned.

Once again, we begin with the case that the map ϕ is dissipative with respect to the control measure m. As in the finite-variance case considered in Section 9.4, for the partial sums of the process to behave in a way consistent with short memory, additional "size" constraints on the function f may be required. As we have already seen in the previous section, additional constraints are needed, mostly in the case $1 < \alpha < 2$.

In order to see clearly what happens in the case of a dissipative map ϕ, we will consider the case in which the process $\mathbf{X}$ is already SαS with $0 < \alpha < 2$. This corresponds to the situation in which the local Lévy measure ρ has the form

$$\rho(dx) = \alpha |x|^{-(\alpha+1)}\, dx\,.$$

In this case, the condition (9.38) for the function f to be in $L_0(M)$, and hence for the process $\mathbf{X}$ to be a well-defined stationary process, is $f \in L_\alpha(m)$; see Example 3.3.8. Moreover, in this case, the function H in (9.85) is given by

$$H(y) = y^{-\alpha} \int_E |f(s)|^\alpha\, m(ds),\ y > 0\,,$$

so (9.85) holds. The statement (9.86) then describes a fact already known to us, namely the power decay of the tail of a symmetric stable random variable. This also says that in this case, the sequence (a_n) defined in (9.62) can be taken to be $a_n = n^{1/\alpha}\|f\|_\alpha$.

The following theorem is an infinite-variance analogue of Theorem 9.4.1. Note the difference in the conditions required in the cases $0 < \alpha \le 1$ and $1 < \alpha < 2$.

Theorem 9.6.1. *Let $\mathbf{X}$ be a stationary SαS process given by (9.37) with some $f \in L_\alpha(m)$. Suppose that the map ϕ is dissipative with respect to the measure m.*

(i) Suppose that $0 < \alpha \le 1$. Then the limit

$$b^\alpha = \lim_{n\to\infty} \frac{1}{n} \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) \in [0, \infty) \tag{9.87}$$

exists and is finite, and

$$\big(n^{-1/\alpha} S_n(t),\, t \geq 0\big) \Rightarrow (b\, Y_\alpha(t),\ t \geq 0) \quad \text{as } n \to \infty \tag{9.88}$$

in the sense of convergence of finite-dimensional distributions.

(ii) Suppose that $1 < \alpha < 2$. *Assume that*

$$\int_E |f(s)| \left(\sum_{k=-\infty}^{\infty} |f| \circ \phi^k(s) \right)^{\alpha-1} m(ds) < \infty\,. \tag{9.89}$$

Then the limit b *in* (9.87) *exists and is finite, and* (9.88) *holds in terms of convergence of finite-dimensional distributions.*

Proof. We consider the case $0 < \alpha \leq 1$ first. Let

$$g_n = \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds),\ n = 1, 2, \ldots\,.$$

Notice that for $n, m \geq 1$,

$$\begin{aligned} g_{n+m} &= \int_E \left| \sum_{k=1}^{n+m} f \circ \phi^k(s) \right|^\alpha m(ds) \\ &\leq \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) + \int_E \left| \sum_{k=n+1}^{n+m} f \circ \phi^k(s) \right|^\alpha m(ds) \\ &= \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) + \int_E \left| \sum_{k=1}^{m} f \circ \phi^k(s) \right|^\alpha m(ds) = g_n + g_m\,, \end{aligned}$$

using the fact that the measure m is ϕ-invariant. Therefore, the sequence $(g_n,\ n \geq 1)$ is subadditive, and by Exercise 2.6.14, the limit b in (9.87) exists and is finite. Note that this argument has nothing to do with the fact that ϕ is a dissipative map.

In order to prove (9.88), it is, once again, enough to show that for every $m = 1, 2, \ldots$ and positive integers $j_1, \ldots, j_m$, we have

$$n^{-1/\alpha}\big(S_{j_1 n}, \ldots, S_{j_m n}\big) \Rightarrow b\big(Y_\alpha(j_1), \ldots, Y_\alpha(j_m)\big)\,; \tag{9.90}$$

see Exercise 9.10.7. We may assume that $j_k = k$, $k = 1, \ldots, m$. It follows from (9.87) and the form of the characteristic functions of integrals with respect to SαS random measures in (3.55) that for every $k = 1, \ldots, m$, we have

$$n^{-1/\alpha}\big(S_{kn} - S_{(k-1)n}\big) \Rightarrow b\big(Y_\alpha(k) - Y_\alpha(k-1)\big)\,.$$

Since for integrals with respect to infinitely divisible random measures, pairwise independence implies full independence (this facts follows immediately from part (i) of Corollary 3.3.11), in order to prove (9.90), we need to show only that for every $1 \le k_1 < k_2 \le m$,

$$\begin{aligned} & n^{-1/\alpha}\big(S_{k_1 n} - S_{(k_1-1)n},\, S_{k_2 n} - S_{(k_2-1)n}\big) \\ & \Rightarrow b\big(Y_\alpha(k_1) - Y_\alpha((k_1-1)),\, Y_\alpha(k_2) - Y_\alpha((k_2-1))\big)\,. \end{aligned} \tag{9.91}$$

By stationarity, it is enough to consider the case $k_1 = 1$, and we replace k_2 by k. Since the marginal convergence of the increments has already been established, we need to prove only that the random vector on the left-hand side of (9.91) converges weakly to a limit with independent components. By Theorem 13.14 in Kallenberg (1997), we need to prove that the bivariate Lévy measure μ_n of the SαS random vector on the left-hand side of (9.91) converges vaguely in $[-\infty,\infty]^2 \setminus \{\mathbf{0}\}$ to a measure concentrated on the axes. By symmetry, we need to check that for every $a > 0$,

$$\lim_{n\to\infty} \mu_n\big((a,\infty)\times(a,\infty)\big) = \lim_{n\to\infty} \mu_n\big((-\infty,-a)\times(a,\infty)\big) = 0\,. \tag{9.92}$$

We will prove the first statement in (9.92). The second statement can be checked in the same way. By Theorem 3.3.10,

$$\begin{aligned} & \mu_n\big((a,\infty)\times(a,\infty)\big) \\ & = \int_E \left[\int_{-\infty}^{\infty} \mathbf{1}\left(x n^{-1/\alpha} \sum_{i=1}^{n} f\circ\phi^i(s) > a,\right.\right. \\ & \qquad \left.\left. x n^{-1/\alpha} \sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s) > a \right) \alpha |x|^{-(1+\alpha)}\, dx \right] m(ds) \\ & \le a^{-\alpha}\frac{1}{n}\int_E \min\left(\left|\sum_{i=1}^{n} f\circ\phi^i(s)\right|^\alpha,\, \left|\sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s)\right|^\alpha \right) m(ds)\,. \end{aligned} \tag{9.93}$$

Let $\epsilon > 0$. For $m = 1, 2, \ldots$, let

$$A_m^{(\epsilon)} = \big\{ s \in E :\, |f|\circ\phi^i(s) > \epsilon,\, |f|\circ\phi^j(s) > \epsilon \text{ for some } |i-j| \ge m \big\}\,.$$

Clearly, $A_m^{(\epsilon)}$ is a ϕ-invariant set. Furthermore, since ϕ is dissipative,

$$A_m^{(\epsilon)} \downarrow \emptyset \ \text{ as } m \to \infty \tag{9.94}$$

for every fixed $\epsilon > 0$. We can write for $j \geq 1$,

$$\begin{aligned}
&\frac{1}{n}\int_E \min\left(\left|\sum_{i=1}^{n} f\circ\phi^i(s)\right|^\alpha, \left|\sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s)\right|^\alpha\right) m(ds)\\
&= \frac{1}{n}\int_{A_j^{(\epsilon)}} \min\left(\left|\sum_{i=1}^{n} f\circ\phi^i(s)\right|^\alpha, \left|\sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s)\right|^\alpha\right) m(ds)\\
&+ \frac{1}{n}\int_{(A_j^{(\epsilon)})^c} \min\left(\left|\sum_{i=1}^{n} f\circ\phi^i(s)\right|^\alpha, \left|\sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s)\right|^\alpha\right) m(ds)\\
&:= I_1^{(j,\epsilon)}(n) + I_2^{(j,\epsilon)}(n)\,.
\end{aligned}$$

Since $0 < \alpha \leq 1$, we have

$$I_1^{(j,\epsilon)}(n) \leq \sum_{i=1}^{n} \frac{1}{n}\int_{A_j^{(\epsilon)}} |f|^\alpha \circ \phi^i(s)\, m(ds) = \int_{A_j^{(\epsilon)}} |f(s)|^\alpha\, m(ds),$$

since the set $A_m^{(\epsilon)}$ is ϕ-invariant. Since $f \in L_\alpha(m)$, it follows from (9.94) that

$$\lim_{j\to\infty} \limsup_{n\to\infty} I_1^{(j,\epsilon)}(n) = 0\,. \tag{9.95}$$

Suppose now that $n > j$. On the event $(A_j^{(\epsilon)})^c$, one of the two sums

$$\sum_{i=1}^{n-j} f\circ\phi^i(s) \quad \text{and} \quad \sum_{i=(k-1)n+1}^{kn} f\circ\phi^i(s)$$

has all terms not exceeding ϵ in absolute value. Therefore,

$$\begin{aligned}
I_2^{(j,\epsilon)}(n) &\leq \frac{1}{n}\int_E \sum_{i=n-j+1}^{n} |f|^\alpha \circ \phi^i(s)\, m(ds)\\
&+ \frac{2}{n}\int_E \sum_{i1}^{n} |f|^\alpha \circ \phi^i(s) \mathbf{1}\big(|f|^\alpha \circ \phi^i(s) \leq \epsilon\big)\, m(ds)\\
&= \frac{j}{n}\int_E |f(s)|^\alpha\, m(ds) + 2\int_E |f(s)|^\alpha \mathbf{1}\big(|f(s)| \leq \epsilon\big)\, m(ds)\,.
\end{aligned}$$

We conclude by (9.95) that for every $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \int_E \min\left(\left| \sum_{i=1}^{n} f \circ \phi^i(s) \right|^\alpha , \left| \sum_{i=(k-1)n+1}^{kn} f \circ \phi^i(s) \right|^\alpha \right) m(ds)$$

$$\leq 2 \int_E |f(s)|^\alpha \mathbf{1}\big(|f(s)| \leq \epsilon\big)\, m(ds)\,.$$

Now the first statement in (9.92) follows by letting $\epsilon \to 0$. This completes the proof of part (i) of the theorem.

We begin proving part (ii) by establishing the existence of the limit in (9.87). Since the map ϕ is dissipative, by part (ii) of Theorem 2.4.3 there exists a wandering set W such that $E = \cup_{k=-\infty}^{\infty} \phi^k(W)$ modulo a null set. We will prove that the limit in (9.87) exists and is given by

$$b = \left(\int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} . \tag{9.96}$$

Notice that the assumption (9.89) guarantees that the expression on the right-hand side of (9.96) is finite.

Suppose first that the function f is supported by finitely many translates of W. That is, suppose that there are integers $m_1 < m_2$ such that

$$f(s) = 0 \ \ m\text{-a.e. on } \ \big(\cup_{i=m_1}^{m_2} \phi^i(W)\big)^c \,. \tag{9.97}$$

We have

$$\int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) = \sum_{i=-\infty}^{\infty} \int_{\phi^i(W)} \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds)$$

$$= \sum_{i=-\infty}^{\infty} \int_W \left| \sum_{k=1}^{n} f \circ \phi^{k+i}(s) \right|^\alpha m(ds) = \sum_{i=-\infty}^{\infty} \int_W \left| \sum_{k=i+1}^{i+n} f \circ \phi^k(s) \right|^\alpha m(ds)\,.$$

It follows from (9.97) that if $n > m_2 - m_1$, then in the last sum above, the terms corresponding to $i = m_2 - n, m_2 - n + 1, \ldots, m_1 - 1$ are all equal to

$$\int_W \left| \sum_{k=m_1}^{m_2} f \circ \phi^i(s) \right|^\alpha m(ds) = \int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^i(s) \right|^\alpha m(ds)\,,$$

while for $i = m_1, m_1 + 1, \ldots, m_2 - 1$ and $i = m_1 - n, m_1 - n + 1, \ldots, m_2 - n - 1$, the corresponding terms do not exceed this quantity. We conclude that

$$\left| \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) - (n - m_2 + m_1 - 1) \int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^i(s) \right|^\alpha m(ds) \right|$$

$$\leq 2(m_2 - m_1) \int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^i(s) \right|^\alpha m(ds) .$$

This clearly implies that the limit in (9.87) exists, and b is given by (9.96).

For an f not necessarily supported by finitely many translates of W, we define, for $j = 1, 2, \ldots, f_j = f\mathbf{1}(\cup_{i=-j}^{j} \phi^i(W))$. Clearly, each f_j satisfies (9.97). We have

$$\left| \left(\frac{1}{n} \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} - \left(\int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} \right|$$

$$\leq \left| \left(\frac{1}{n} \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} - \left(\frac{1}{n} \int_E \left| \sum_{k=1}^{n} f_j \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} \right|$$

$$+ \left| \left(\frac{1}{n} \int_E \left| \sum_{k=1}^{n} f_j \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} - \left(\int_W \left| \sum_{k=-\infty}^{\infty} f_j \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} \right|$$

$$+ \left| \left(\int_W \left| \sum_{k=-\infty}^{\infty} f_j \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} - \left(\int_W \left| \sum_{k=-\infty}^{\infty} f \circ \phi^k(s) \right|^\alpha m(ds) \right)^{1/\alpha} \right| .$$

Note that for each j, the middle term on the right-hand side converges to zero as $n \to \infty$, since the function f_j satisfies (9.97). Furthermore, by the triangle inequality in $L_\alpha(m)$, the last term on the right-hand side can be bounded from above by

$$\left(\int_W \left(\sum_{k=-\infty}^{\infty} |f - f_j| \circ \phi^k(s) \right)^\alpha m(ds) \right)^{1/\alpha} .$$

Since $|f - f_j|$ is less than or equal to $|f|$ and goes to zero as $j \to \infty$, the last expression converges to zero as $j \to \infty$. Similarly, the first term on the right-hand side can be bounded from above by

$$\left(\frac{1}{n} \int_E \left(\sum_{k=1}^{n} |f - f_j| \circ \phi^k(s) \right)^\alpha m(ds) \right)^{1/\alpha}$$

$$\leq \left(\frac{1}{n} \int_E \sum_{k=1}^{n} |f - f_j| \circ \phi^k(s) \left(\sum_{i=-\infty}^{\infty} |f - f_j| \circ \phi^i(s) \right)^{\alpha-1} m(ds) \right)^{1/\alpha}$$

$$= \left(\int_E |f - f_j| \left(\sum_{i=-\infty}^{\infty} |f - f_j| \circ \phi^i(s) \right)^{\alpha-1} m(ds) \right)^{1/\alpha},$$

which, by (9.89), converges to zero as $j \to \infty$. This proves that the limit in (9.87) exists, and b is given by (9.96) in full generality.

In order to complete the proof of the theorem in the case $1 < \alpha < 2$, one can use an argument similar to the argument we used above in the case $0 < \alpha \leq 1$. As in the latter case, one needs to prove only that (9.92) holds. We leave the verification to Exercise 9.10.16. □

Remark 9.6.2. Note that the expression (9.96) for the limiting scaling constant b in (9.87) is valid for all $0 < \alpha < 2$; see Exercise 9.10.15. As long as that limiting scaling constant does not vanish, we should view the stationary SαS process satisfying the conditions of Theorem 9.6.1 as having short memory from the point of view of the behavior of its partial sums. Notice that the theorem does not require anything additional from the kernel f (besides the condition $f \in L_\alpha(m)$ necessary for the process to be well defined) if $0 < \alpha \leq 1$. Therefore, if α is in that range, the only possibility for a stationary SαS process, corresponding to a dissipative map ϕ, to have long-range dependence as far as its partial sums are concerned, is the possibility of cancellations resulting in the case $b = 0$. The situation is, therefore, similar to that in Remark 9.5.6, where infinite-variance infinite-moving-average processes were considered.

In the case $1 < \alpha < 2$, the short-memory conclusion of Theorem 9.6.1 requires a "size" condition (9.89) on the values of the function $|f|$ over the trajectories of the flow (ϕ^n). In this sense, the situation is similar to our discussion of finite-variance infinitely divisible processes in Section 9.4. Condition (9.89) may or may not hold when the map ϕ is dissipative, and when that condition does not hold, properly normalized partial sums of the process may have a limit different from the α-stable Lévy process as in (9.63). One example of this is the infinite-variance linear process considered in Section 9.5. We have seen in Example 9.4.4 that when the noise variables are infinitely divisible, then the linear process is itself infinitely divisible. If the noise variables are SαS, then the linear process is itself SαS. Theorem 9.5.7 shows that in this case, the linear fractional symmetric stable motion can appear as the limit. There exist, however, stationary SαS processes with a dissipative map ϕ such that their normalized partial sums converge to other SαS self-similar processes with stationary increments.

Example 9.6.3. Dilated fractional stable motions This class of processes was introduced in Pipiras and Taqqu (2002b). Let $E = G \times \mathbb{R} \times (0, \infty)$, where $(G, \mathcal{G})$ is a measurable space. We equip E with the product σ-field. Let η be a σ-finite measure on $(G, \mathcal{G})$. If M is an SαS random measure on E with modified control measure $\tilde{m}$ given by

$$\tilde{m}(dy,\, du,\, dx) = \eta(dy)\, du\, x^{-1}\, dx,\ (y, u, x) \in E,$$

and $H > 0$ is a number satisfying the restrictions (8.16) of Corollary 8.2.7, then for a measurable function $F : G \times \mathbb{R} \to \mathbb{R}$ satisfying the appropriate integrability condition, we can define

$$Y(t) = \int_G \int_{\mathbb{R}} \int_0^\infty x^{-(H-1/\alpha)} \big(F(y, x(t+u)) - F(y, xu)\big)\, M(dy, du, dx)\,, \tag{9.98}$$

$t \in \mathbb{R}$. The integrability condition on F for the process $\mathbf{Y}$ to be well defined is

$$\int_G \int_{\mathbb{R}} \int_0^\infty x^{-\alpha H-1} \big|F(y, x+u) - F(y, u)\big|^\alpha \, \eta(dy)\, du\, dx < \infty\,; \tag{9.99}$$

see Example 3.3.8. Under this condition, the process $\mathbf{Y}$ is an SαS process. Moreover, for every $k = 1, 2, \ldots$, real numbers $t_1, \ldots, t_k$, real numbers $\theta_1, \ldots, \theta_k$, and $c > 0$, we have, by (3.55),

$$\begin{aligned} & E \exp\left\{ i \sum_{j=1}^k \theta_j Y(ct_j) \right\} \qquad (9.100) \\ &= \exp\left\{ - \int_G \int_{\mathbb{R}} \int_0^\infty x^{-\alpha H} \left| \sum_{j=1}^k \theta_j \big(F(y, x(ct_j + u)) - F(y, xu)\big) \right|^\alpha \eta(dy)\, du\, dx \right\} \\ &= \exp\left\{ -c^{\alpha H} \int_G \int_{\mathbb{R}} \int_0^\infty x^{-\alpha H} \left| \sum_{j=1}^k \theta_j \big(F(y, x(t_j + u)) - F(y, xu)\big) \right|^\alpha \eta(dy)\, du\, dx \right\} \\ &= E \exp\left\{ i \sum_{j=1}^k c^H \theta_j Y(t_j) \right\}\,. \end{aligned}$$

Therefore, the process $\mathbf{Y}$ is self-similar with exponent H. It is equally simple to check that it has stationary increments. Therefore, the increment process $X_n = Y(n) - Y(n-1)$, $n \in \mathbb{Z}$, is a stationary SαS process. This process is of the form (9.37) with

$$f(y, u, x) = x^{-(H-1/\alpha)} \big(F(y, xu)) - F(y, x(u-1))\big),\ (y, u, x) \in E\,,$$

and $\phi : E \to E$ given by $\phi(y, u, x) = (y, u+1, x)$. This is a dissipative map, since the set $W = G \times (0, 1] \times \mathbb{R}$ is wandering and $E = \cup_n \phi^n(W)$. We saw in Example 3.6.9 that the increment process of the linear fractional stable motion also corresponds to a dissipative flow. Nonetheless, the two processes are different.

If $\mathbf{X}$ is the increment process of a dilated fractional stable motion, then according to Proposition 9.2.6, its partial sums satisfy a functional central limit theorem (9.11). The limit is, of course, the dilated fractional stable motion itself. This both tells us

that the process $\mathbf{X}$ has long-range dependence from the point of view of the behavior of its partial sums and shows that limits other than a linear fractional symmetric stable motion can appear.

An example of a dilated fractional stable motion is the *telecom process*. It is defined for $1 < \alpha < 2$ and $1/\alpha < H < 1$. Here $G = \{0\}$ is a singleton (with η a unit point mass at the single element), and $F(0,u) = \big(\min(u,0)+1\big)_+$, $u \in \mathbb{R}$. We leave checking the integrability condition to the reader (Exercise 9.10.17).

As in the finite-variance case of Section 9.4, a conservative map in (9.37) makes it hard for a stationary infinitely divisible process to have short memory from the point of view of the behavior of its partial sums. The following result goes in the same direction as Proposition 9.4.5.

Proposition 9.6.4. *Let* $\mathbf{X}$ *be a stationary* $S\alpha S$ *process given by* (9.37) *with some* $f \in L_\alpha(m)$. *Assume that the map* ϕ *is conservative with respect to the measure* m.

(i) Suppose that $0 < \alpha < 1$. *Then* (9.87) *holds with a zero limit, and hence the normalized partial sums process on the left-hand side of* (9.88) *converges weakly to the zero process.*

(ii) Suppose that $1 < \alpha < 2$. *Assume that* $f \geq 0$ m*-a.e. on the set* E_f *in* (9.42). *If* $m(E_f) > 0$, *then* (9.87) *holds with an infinite limit, and hence the finite-dimensional distributions of the normalized partial sums process on the left-hand side of* (9.88) *are not tight.*

Proof. For both parts of the theorem, we may assume that $E_f = E$ modulo a null set. We begin with the case $0 < \alpha < 1$. It is clearly enough to show that (9.87) holds with a zero limit. If $f = 0$ modulo a null set, then this is trivial, so assume that $m(E_f) > 0$. For $K = 1, 2, \ldots$ and $M > 0$, define

$$B_{K,M} = \left\{ s \in S : \sum_{k=-K}^{K} |f| \circ \phi^k(s) \leq M \right\} . \tag{9.101}$$

As in the proof of Proposition 9.4.5, it follows from the fact that ϕ is conservative that (9.54) holds, so that

$$B_{K,M} \downarrow \emptyset \ \text{ as } K \to \infty \text{ for every } M > 0. \tag{9.102}$$

Suppose first that the function f is bounded. Let $M > 0$ be a large number. Write

$$\begin{aligned} &\frac{1}{n} \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha m(ds) \\ &= \frac{1}{n} \int_E \left| \sum_{k=1}^{n} f \circ \phi^k(s) \right|^\alpha \mathbf{1}\left(\sum_{k=1}^{n} |f| \circ \phi^k(s) \ > M \right) m(ds) \end{aligned}$$

$$+\frac{1}{n}\int_E \left|\sum_{k=1}^n f\circ\phi^k(s)\right|^\alpha \mathbf{1}\left(\sum_{k=1}^n |f|\circ\phi^k(s)\le M\right) m(ds)$$

$$\le M^{-(1-\alpha)}\frac{1}{n}\int_E \sum_{k=1}^n |f|\circ\phi^k(s)\, m(ds)$$

$$+\frac{1}{n}\sum_{j=1}^n \int_E |f(s)|^\alpha \mathbf{1}\left(\sum_{k=1}^n |f|\circ\phi^{k-j}(s)\le M\right) m(ds)\,.$$

Since we have assumed that f is bounded, the assumption $f\in L_\alpha(m)$ implies that $f\in L_1(m)$, and the first term on the right-hand side is $M^{-(1-\alpha)}\|f\|_1\to 0$ as $M\to\infty$. We now show that for every $M>0$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n \int_E |f(s)|^\alpha \mathbf{1}\left(\sum_{k=1}^n |f|\circ\phi^{k-j}(s)\le M\right) m(ds)=0\,. \tag{9.103}$$

To this end, let $K=1,2,\ldots$. For $n>2K$, we can write

$$\frac{1}{n}\sum_{j=1}^n \int_E |f(s)|^\alpha \mathbf{1}\left(\sum_{k=1}^n |f|\circ\phi^{k-j}(s)\le M\right) m(ds)$$

$$\le\frac{2K}{n}\int_E |f(s)|^\alpha\, m(ds)+\frac{1}{n}\sum_{j=K+1}^{n-K}\int_E |f(s)|^\alpha \mathbf{1}\left(\sum_{k=-K}^{K} |f|\circ\phi^{k}(s)\le M\right) m(ds)$$

$$\to\int_E |f(s)|^\alpha \mathbf{1}\left(s\in B_{K,M}\right)\, m(ds)$$

as $n\to\infty$. Since $f\in L_\alpha(m)$, we can let $K\to\infty$ and use (9.102) to check that (9.103) holds.

We have proved that (9.87) holds with a zero limit if the function f is bounded. In the general case, for $D>0$ we can write

$$\frac{1}{n}\int_E \left|\sum_{k=1}^n f\circ\phi^k(s)\right|^\alpha m(ds)$$

$$\le\frac{1}{n}\int_E \left|\sum_{k=1}^n f\mathbf{1}\big(|f|\le D\big)\circ\phi^k(s)\right|^\alpha m(ds)+\frac{1}{n}\int_E \left|\sum_{k=1}^n f\mathbf{1}\big(|f|>D\big)\circ\phi^k(s)\right|^\alpha m(ds)\,.$$

Since the function $f\mathbf{1}\big(|f|\le D\big)$ is bounded, the first term on the right-hand side vanishes in the limit. Furthermore,

$$\limsup_{n\to\infty}\frac{1}{n}\int_E \left|\sum_{k=1}^n f\mathbf{1}\big(|f|>D\big)\circ\phi^k(s)\right|^\alpha m(ds)\le\int_E |f(s)|^\alpha\mathbf{1}\big(|f(s)|>D\big)\, m(ds)\,,$$

which converges to zero as $D \to \infty$. This completes the proof of part (i) of the proposition.

The proof of part (ii) is similar to the proof of part (i) and is left to Exercise 9.10.18. □

A variety of limits can be obtained in the functional central limit theorem for partial sums of heavy-tailed stationary symmetric infinitely divisible processes defined by (9.37) in which the map ϕ is conservative. This variety is greater than for the corresponding finite-variance processes because of the large number of different SαS self-similar processes with stationary increments. We finish this section by demonstrating one such situation. We consider the partial sums of a heavy-tailed version of a symmetric stationary infinitely divisible process defined by (9.37) in which the representation corresponds to a null recurrent Markov chain of the setup of Theorem 9.4.7. We will no longer assume that the process is SαS. Instead, we will assume that the tail of the Lévy measure ρ of the infinitely divisible random measure M in (9.37) is regularly varying. The following theorem, which we state without proof, shows that in this case, one obtains, as a limit, a process not yet seen in the limit theorems of this section.

Theorem 9.6.5. *Let* $\mathbf{X}$ *be a stationary symmetric infinitely divisible process given in the form* (9.37)*. Assume that the local Lévy measure* ρ *satisfies*

$$\rho\big((x,\infty)\big) \text{ is regularly varying with exponent } -\alpha \in (-2,0) \text{ as } x \to \infty \tag{9.104}$$

and that for some $\theta < 2$*,*

$$x^{\theta}\rho\big((x,\infty)\big) \to 0 \text{ as } x \to 0.$$

Let $E = \mathbb{Z}^{\mathbb{Z}}$*, and let* m *be a shift-invariant measure on* E *generated by an invariant measure of an irreducible null recurrent Markov chain on* $\mathbb{Z}$ *given in* (9.55)*, and* f *the indicator function given by* (9.56)*. Let* ϕ *be the left shift on* E*. Assume that the sequence* $P_0(\tau_1 > n)$*,* $n = 1, 2, \ldots$*, is regularly varying with exponent* $-\beta \in (-1, 0)$*. Then the partial sums of the process* $\mathbf{X}$ *satisfy*

$$\big(c_n^{-1} S_n(t),\ t \geq 0\big) \Rightarrow \big(c_{\alpha,\beta} Y_{\alpha,\beta}(t),\ t \geq 0\big) \text{ as } n \to \infty \tag{9.105}$$

in the Skorokhod J_1 *topology on* $D[0,\infty)$*, where* $\mathbf{Y}_{\alpha,\beta}$ *is the* β*-Mittag-Leffler fractional S*α*S motion of Example 8.4.1. The normalizing sequence is given by*

$$c_n = \frac{\rho^{\leftarrow}\big((nP_0(\tau_1 > n))^{-1}\big)}{P_0(\tau_1 > n)},\ n = 1, 2, \ldots,$$

with

$$\rho^{\leftarrow}(u) = \inf\big\{x > 0 :\ \rho\big((x,\infty)\big) \leq u\big\},\ u > 0\,.$$

Furthermore,

$$c_{\alpha,\beta} = \pi_0 \frac{\Gamma(1-\beta)}{(1-\beta)^{1/\beta}} C_\alpha^{1/\alpha},$$

where C_α is given by (3.31).

We refer the reader to Owada and Samorodnitsky (2015a) for a proof. We also note that under the conditions of the theorem, the statement (9.85) holds, and hence the marginal distributions of the process $\mathbf{X}$ are regularly varying with exponent $-\alpha \in (-2, 0)$. See Exercise 9.10.19.

Remark 9.6.6. Since the limiting process in Theorem 9.6.5, the β-Mittag-Leffler fractional SαS motion, is different from the SαS Lévy motion in (9.63), the symmetric stationary infinitely divisible process in the theorem should be regarded as having long-range dependence from the point of view of the behavior of its partial sums. It is also interesting to compare the effect of the tail of the first return time of the Markov chain on the normalizing constant (c_n) in the theorem with the analogous situation in the finite-variance case considered in Remark 9.4.8. The sequence (c_n) is, of course, regularly varying with exponent $(1-\beta)/\alpha + \beta$, and $0 < \beta < 1$ describes how heavy the tail of the first return time τ_1 is. This exponent is smaller than $1/\alpha$ (recall that in the short-memory case, the exponent of regular variation of the normalizing constants is equal to $1/\alpha$) if $0 < \alpha < 1$ and larger than $1/\alpha$ if $1 < \alpha < 2$. In both cases, the exponent of regular variation of (c_n) approaches $1/\alpha$ as $\beta \to 0$, the Markov chain becomes closer to being transient, and the left shift ϕ on $\mathbb{Z}^{\mathbb{Z}}$ becomes closer to being dissipative. It is also interesting to note that in the case $\alpha = 1$, the exponent of regular variation of the sequence (c_n) is equal to $1 = 1/\alpha$ regardless of the value of β. In spite of this fact, we should still view the process in Theorem 9.6.5 as being long-range dependent as far as its partial sums are concerned, because the limiting process is different from the SαS Lévy motion. It is useful to recall at this point that the increment process of the β-Mittag-Leffler fractional SαS motions is generated by a conservative flow, while the increment process of the SαS Lévy motion is generated by a dissipative flow.

9.7 Phase Transitions in Partial Maxima

In this section, we will address the question of long-range dependence by studying the partial maxima of a stationary process and looking for possible phase transitions. Both the notation and the approach are similar here and in our study of the partial sums in Section 9.2. Let $\mathbf{X} = (X_n, n \in \mathbb{Z})$ be a stationary stochastic process. We will use the notation

$$M_n = \max(X_1, \ldots, X_n), \; n = 0, 1, 2, \ldots \tag{9.106}$$

(with $M_0 = 0$) for the sequence of partial sums of the process. We will introduce certain assumptions on the marginal distributions of the process $\mathbf{X}$ and investigate whether the behavior of the partial maxima is qualitatively different from the behavior of the corresponding partial maxima of an i.i.d. sequence with the same (or similar) marginal distributions. As in the case of the partial sums, by qualitative difference we mean a change in the order of magnitude of the partial maxima, or a different limit obtained in a limit theorem.

We would like to allow the partial maximum M_n of the process to have a chance to become large as n becomes large. This requires assuming that the right endpoint of the support of the distribution of X is infinite: $P(X > x) > 0$ for all $x \in \mathbb{R}$. We will consider two specific assumptions on the marginal distribution of the process $\mathbf{X}$. They are easiest to state in terms of the partial maxima of an i.i.d. sequence with the same marginal distribution as $\mathbf{X}$, so let $(\tilde{X}_n,\ n \in \mathbb{Z})$ be an i.i.d. sequence such that $\tilde{X}_0 \stackrel{d}{=} X_0$.

Assumption 9.7.1. *Either*

- *for some positive sequences* (a_n) *and* (b_n) *with* $b_n \to \infty$ *and* $b_n/a_n \to \infty$ *as* $n \to \infty$,

$$P\left(\frac{1}{a_n}\big(\max(\tilde{X}_1, \dots, \tilde{X}_n) - b_n\big) \le x\right) \to \exp\{-e^{-x}\} := G_0(x) \tag{9.107}$$

  as $n \to \infty$ *for* $x \in \mathbb{R}$, *or*
- *for some* $\alpha > 0$ *and a positive sequence* (a_n) *that is regularly varying with exponent* $1/\alpha$,

$$P\left(\frac{1}{a_n}\max(\tilde{X}_1, \dots, \tilde{X}_n) \le x\right) \to \exp\{-x^{-\alpha}\} := \Phi_\alpha(x) \tag{9.108}$$

  as $n \to \infty$ *for* $x > 0$.

The limiting distribution G_0 on the right-hand side of (9.107) is the standard Gumbel distribution, and one-dimensional distributions (of $\tilde{X}$) that satisfy (9.107) are said to be in the maximum domain of attraction of the Gumbel distribution. The limiting distribution Φ_α on the right-hand side of (9.108) is the standard Fréchet (or α-Fréchet) distribution, and distributions that satisfy (9.108) are said to be in the maximum domain of attraction of the Fréchet distribution. Details on this and other facts from the one-dimensional extreme value theory can be found, for example, in deHaan and Ferreira (2006).

A necessary and sufficient condition for the distribution of a random variable X to belong to the Gumbel domain of attraction is existence of a positive differentiable function f with $f'(t) \to 0$ as $t \to \infty$ such that

$$\lim_{t\to\infty} \frac{P\big(X > t + xf(t)\big)}{P(X > t)} = e^{-x} \quad \text{for all } x \in \mathbb{R}. \tag{9.109}$$

Notice that (9.109) is satisfied by both the standard exponential distribution ($f \equiv 1$) and the standard normal distribution ($f(t) = 1/t$). If the distribution of a random variable X is in the Gumbel domain of attraction, and a function f satisfies (9.109) for that random variable, then convergence in (9.107) holds with $a_n = f(\tilde{a}_n)$, $b_n = \tilde{a}_n$, where

$$\tilde{a}_n = \inf\{x > 0 : P(X > x) \le 1/n\}, \ n = 1, 2, \dots \tag{9.110}$$

(this is the same sequence as in (9.62) in a different notation).

A necessary and sufficient condition for the distribution of a random variable X to belong to the Fréchet domain of attraction is

$$\text{the right tail } P(X > t) \text{ is regularly varying with exponent } -\alpha. \tag{9.111}$$

If the distribution of a random variable X is in the Fréchet domain of attraction, then convergence in (9.108) holds with a_n given by (9.62).

Remark 9.7.2. The "size" of the partial maxima is one of the important characteristics of a stationary process in which we will be interested when we study long-range dependence. In this connection, it is important to note that under (9.107), the "size" of the partial maxima is described by the sequence (b_n), while the sequence (a_n) describes the second-order behavior of the partial maxima. On the other hand, under (9.108), the "size" of the partial maxima is clearly given by the sequence (a_n).

As in the case of our discussion of phase transitions in the behavior of the partial sums of stationary processes, we are also interested in functional limit theorems for the partial maxima of stationary processes. Versions of the functional central limit theorems (9.4) and (9.63) for partial maxima hold as long as one-dimensional limit theorems of the type (9.107) and (9.108) hold. For a one-dimensional distribution F, one can define a consistent system of finite-dimensional distributions on $(0, \infty)$ by

$$\begin{aligned} &F_{t_1,t_2,\dots,t_k}(x_1, x_2, \dots, x_k) \\ &= \big(F\big(\min(x_1, x_2, \dots, x_k)\big)\big)^{t_1} \big(F\big(\min(x_2, \dots, x_k)\big)\big)^{t_2 - t_1} \dots \big(F(x_k)\big)^{t_k - t_{k-1}} \end{aligned} \tag{9.112}$$

for $0 < t_1 < t_2 < \dots < t_k$ and real numbers $x_1, x_2, \dots, x_k$, so we can define a stochastic process $\big(Y_F(t), t > 0\big)$ with finite-dimensional distributions given by (9.112). This process is called an F-extremal process. It is continuous in probability and has a version with nondecreasing sample paths that are right continuous and have left limits; see Dwass (1964) and Resnick and Rubinovitch (1973). If $x_L := \inf\{x \in \mathbb{R} : F(x) > 0\} > -\infty$, then the domain of the F-extremal process can be extended to $t = 0$ by setting $Y_F(0) = x_L$.

For a stationary process $\mathbf{X}$, we define the nth partial maxima process by $M_n(t) = M_{[nt]}$ for $t \ge 0$. It was shown in Lamperti (1964) that if $\mathbf{X}$ is an i.i.d. sequence

and (9.107) holds (i.e., the marginal distribution of the process is in the Gumbel domain of attraction), then as $n \to \infty$,

$$\left(\frac{1}{a_n}(M_n(t) - b_n),\ t > 0\right) \Rightarrow \left(Y_{G_0}(t),\ t > 0\right) \tag{9.113}$$

weakly in the Skorokhod J_1-topology on $D(0, \infty)$ (recall that the latter topology is defined as the J_1-topology on each compact subinterval of $(0, \infty)$). On the other hand, if $\mathbf{X}$ is an i.i.d. sequence and (9.108) holds (i.e., the marginal distribution of the process is in the α-Fréchet domain of attraction), then as $n \to \infty$,

$$\left(\frac{1}{a_n}M_n(t),\ t \geq 0\right) \Rightarrow \left(Y_{\Phi_\alpha}(t),\ t \geq 0\right) \tag{9.114}$$

weakly in the Skorokhod J_1-topology on $D[0, \infty)$.

When studying the behavior of the partial sums of stationary processes, we looked for limits in a functional central limit theorem different from a Brownian motion as in (9.4), or from an α-stable Lévy motion as in (9.63). The presence of such a limit can be viewed as a phase transition, and hence it indicates long-range dependence from the point of view of the behavior of the partial sums. Similarly, when studying the behavior of the partial maxima of a stationary processes, we can look at functional limit theorems for the maxima. The presence of a limit different from an extremal process as in (9.113) or in (9.114) can be viewed as an indication of long-range dependence from the point of view of the behavior of the partial maxima.

Let $\mathbf{X} = (X_n,\ n \in \mathbb{Z})$ be a stationary stochastic process, whose sequence (M_n) of partial maxima we would like to study with the purpose of detecting possible long-range dependence. We begin by describing a general phenomenon: the partial maxima of a stationary process cannot grow faster than the partial maxima of an i.i.d. sequence with the same marginal distributions. This phenomenon is similar to what we observed in Proposition 9.5.1 for the partial sums of very heavy tailed stationary processes.

Proposition 9.7.3. *Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a stochastic process with each $X_i \stackrel{d}{=} X$, for some random variable X. Let $\mathbf{Y} = (Y_1, Y_2, \ldots)$ be a sequence of independent copies of X. Let (a_n) be a sequence of positive numbers. If*

$$\limsup_{n\to\infty} P\big(\max(X_1, X_2, \ldots, X_n) > a_n\big) > 0, \tag{9.115}$$

then also

$$\limsup_{n\to\infty} P\big(\max(Y_1, Y_2, \ldots, Y_n) > a_n\big) > 0\,. \tag{9.116}$$

Proof. We have

$$\begin{aligned} P\big(\max(Y_1, Y_2, \ldots, Y_n) > a_n\big) &= 1 - \big(1 - P(X > a_n)\big)^n \\ &\geq 1 - \exp\{-nP(X > a_n)\} \\ &\geq 1 - \exp\big\{-P\big(\max(X_1, X_2, \ldots, X_n) > a_n\big)\big\}\ , \end{aligned}$$

and (9.116) follows from (9.115). □

An immediate conclusion from Proposition 9.7.3 is that every deviation of the rate of growth of the partial maxima of a stationary process from the rate of growth of the partial maxima of the corresponding i.i.d. sequence (specified by a_n) must be in the direction of a slower rate of growth. We will need to keep this in mind when we investigate phase transitions between short and long memory from the point of view of the behavior of the partial maxima.

In general, the boundary between short memory and long memory from the point of view of the behavior of the partial maxima is different from the boundary between short memory and long memory from the point of view of the behavior of the partial sums. This point is already visible when we consider stationary Gaussian processes.

Let $\mathbf{X} = (X_n, n \in \mathbb{Z})$ be a zero-mean stationary Gaussian process with covariance function R_X. Recall that by Proposition 9.2.3, a necessary and sufficient condition for the process to have short memory from the point of view of the behavior of the partial sums is that the sequence (s_n^2) of the variances of the partial sums of the process be regularly varying with exponent 1. We will show now that much weaker conditions imply short memory from the point of view of the behavior of the partial maxima of a stationary Gaussian process. We begin with the following result.

Proposition 9.7.4. *Let* $\mathbf{X} = (X_n, n \in \mathbb{Z})$ *be a zero-mean unit-variance stationary Gaussian process satisfying* $R_X(n) \to 0$ *as* $n \to \infty$. *Then*

$$\frac{1}{\sqrt{2\log n}} M_n \to 1 \ \textit{in probability as } n \to \infty. \tag{9.117}$$

Proof. Let $(Y_n, n = 1, 2, \ldots)$ be a sequence of i.i.d. standard normal random variables, and set $M_n^{(Y)} = \max(Y_1, \ldots, Y_n)$ for $n = 1, 2, \ldots$. The standard normal random variable is in the Gumbel domain of attraction, it satisfies (9.107), and hence as $n \to \infty$,

$$\frac{1}{b_n} M_n^{(Y)} \to 1$$

in probability. Since $b_n = \tilde{a}_n$ given by (9.110), we see that $b_n \sim \sqrt{2\log n}$ as $n \to \infty$, and hence (9.117) holds for the i.i.d. sequence $(Y_n, n = 1, 2, \ldots)$ (i.e., with M_n replaced by $M_n^{(Y)}$). We immediately conclude by Proposition 9.7.3 that as $n \to \infty$,

$$P\left(\frac{1}{\sqrt{2\log n}} M_n > 1 + \varepsilon\right) \to 0$$

for every $\varepsilon > 0$. Therefore, it remains to prove that for every $0 < \varepsilon < 1$,

$$P\left(\frac{1}{\sqrt{2\log n}} M_n > 1 - \varepsilon\right) \to 1 \tag{9.118}$$

as $n \to \infty$. To this end, for $0 < \theta < 1$, let

$$k_\theta = \inf\left\{k = 0, 1, 2, \ldots : R_X(m) \le \theta \text{ for all } m \ge k\right\}.$$

Since the covariance function is assumed to converge to zero, we see that $k_\theta < \infty$ for all $0 < \theta < 1$. Fix $0 < \theta < 2\varepsilon - \varepsilon^2$. We begin with an obvious observation that for every $x \in \mathbb{R}$,

$$P(M_n > x) \ge P\left(\max_{i=1,\ldots,\lfloor n/k_\theta \rfloor} X_{ik_\theta} > x\right).$$

By definition, the Gaussian vector $\left(X_{ik_\theta},\, i = 1, \ldots, \lfloor n/k_\theta \rfloor\right)$ has a covariance matrix will all off-diagonal terms not exceeding θ. We now use the Slepian lemma of Theorem 10.7.9. To this end, let Y_0 be a standard normal random variable independent of the i.i.d. sequence $(Y_n,\, n = 1, 2, \ldots)$. Note that the random vector

$$(1-\theta)^{1/2} Y_i + \theta^{1/2} Y_0,\; i = 1, \ldots, \lfloor n/k_\theta \rfloor$$

is a centered Gaussian random vector with unit variance and covariance matrix whose off-diagonal elements equal θ. By the Slepian lemma,

$$\begin{aligned} P\left(\frac{1}{\sqrt{2\log n}} M_n > 1 - \varepsilon\right) &\ge P\left(\frac{1}{\sqrt{2\log n}} \max_{i=1,\ldots,\lfloor n/k_\theta \rfloor} X_{ik_\theta} > 1 - \varepsilon\right) \\ \ge P\left(\frac{1}{\sqrt{2\log n}} \left(\theta^{1/2} Y_0 + (1-\theta)^{1/2} M^{(Y)}_{\lfloor n/k_\theta \rfloor}\right) > 1 - \varepsilon\right) &\to 1 \end{aligned}$$

as $n \to \infty$, since (9.117) holds for the i.i.d. sequence $(Y_n,\, n = 1, 2, \ldots)$ and by the choice of θ. Therefore, (9.118) holds, and the proposition has been proved. □

Remark 9.7.5. We see that the size of the partial maxima of a stationary Gaussian sequence is exactly the same as that of the partial maxima of an i.i.d. Gaussian sequence, as long as the covariance function of the Gaussian sequence converges to zero as the lag goes to infinity. Therefore, it is reasonable to say that every stationary Gaussian process whose covariance function asymptotically vanishes has short memory from the point of view of the behavior of the partial maxima. Recall, on the other hand, that certain stationary Gaussian processes with an asymptotically vanishing covariance function should be viewed as long-range dependent as far as the behavior of their partial sums is concerned; see, for example, Proposition 9.2.4. This provides examples of processes that have long-range dependence from the point of view of partial sums but not partial maxima.

It turns out that just slightly stronger assumptions on a stationary Gaussian process ensure even greater similarity between the behavior of its partial maxima and the behavior of those of an i.i.d. Gaussian sequence. This is the message of the following result, which we present without proof, which can be found in Berman (1964). Recall that the standard normal satisfies (9.107), where we can choose

$$a_n = (2\log n)^{-1/2}, \tag{9.119}$$
$$b_n = (2\log n)^{1/2} - \frac{1}{2}(2\log n)^{-1/2}\big(\log\log n + \log 4\pi\big),\ n \geq 2\,;$$

see Exercise 9.10.20.

Proposition 9.7.6. *Let* $\mathbf{X} = (X_n,\, n \in \mathbb{Z})$ *be a zero-mean unit-variance stationary Gaussian process satisfying* $R_X(n)\log n \to 0$ *as* $n \to \infty$. *Then with the sequences* (a_n) *and* (b_n) *given in* (9.119), *we have*

$$P\left(\frac{1}{a_n}(M_n - b_n) \leq x\right) \to \exp\{-e^{-x}\} \tag{9.120}$$

as $n \to \infty$ *for* $x \in \mathbb{R}$.

Remark 9.7.7. We see that as long as the covariance function of a stationary Gaussian process converges to zero faster than $(\log n)^{-1}$, the asymptotic behavior of its partial maxima is nearly unchanged from the case of an i.i.d. Gaussian sequence. The situation is, of course, very different if we consider the behavior of the partial sums instead. This shows that the boundary between short memory and long memory for stationary Gaussian processes is very different if we look at the behavior of the partial sums of the process from what we see if we look at the behavior of the partial maxima of the process.

9.8 Partial Maxima of Stationary Stable Processes

The phenomenon described in Remark 9.7.7 occurs in other families of stationary processes. In order to observe this phenomenon in the case of heavy tails, we will consider, as in Section 9.6, stationary SαS processes given in the form (9.37). We will see that for these processes, the boundary between short memory and long memory from the point of view of the behavior of the partial maxima is also different from the boundary between short memory and long memory from the point of view of the behavior of the partial sums. Specifically, from the point of view of the partial maxima, the length of memory turns out to be determined exclusively by whether the map ϕ in (9.37) is dissipative or conservative. In comparison, we saw in Section 9.6 that while whether the map ϕ is dissipative or conservative is also relevant for the length of memory from the point of view of the partial sums, the properties of the

kernel f play a role as well, particularly in the case $1 < \alpha < 2$. The properties of the kernel play no role in the behavior of the partial maxima.

Recall that an SαS random variable is in the maximum domain of attraction of the Fréchet distribution. Moreover, the normalizing constants (a_n) in (9.108) can be taken to be proportional to $n^{1/\alpha}$. Specifically, if an SαS random variable is an observation of an SαS process given by (9.37), then one can take $a_n = n^{1/\alpha}\|f\|_\alpha$. The next theorem shows that if the map ϕ in (9.37) is dissipative, then the SαS process has short memory from the point of view of the behavior of the partial maxima.

Theorem 9.8.1. *Let* $\mathbf{X}$ *be a stationary SαS process given by* (9.37) *with some* $f \in L_\alpha(m)$. *Suppose that the map* ϕ *is dissipative with respect to the measure* m.

(i) Let W *be a wandering set such that* $E = \cup_{k=-\infty}^{\infty} \phi^k(W)$ *modulo a null set. Then*

$$\lim_{n\to\infty} \frac{1}{n} \int_E \left(\max_{k=1,\ldots,n} f_+^\alpha \circ \phi^k(s) \right) m(ds) = \int_W \left(\sup_{k\in\mathbb{Z}} f_+^\alpha \circ \phi^k(s) \right) m(ds) \in [0, \infty) \,. \tag{9.121}$$

(ii) Denote by $m_+(f)$ *the limit in* (9.121), *and define*

$$m_\pm(f) = \frac{1}{2}\big(m_+(f) + m_+(-f)\big) \,.$$

Then

$$\big(n^{-1/\alpha} M_n(t),\ t \ge 0\big) \Rightarrow \Big(\big(m_\pm(f)\big)^{1/\alpha}\, Y_{\Phi_\alpha}(t),\ t \ge 0\Big) \quad \text{as } n \to \infty \tag{9.122}$$

in the sense of convergence of finite-dimensional distributions, where $\mathbf{Y}_{\Phi_\alpha}$ *is the extremal process corresponding to the standard Fréchet distribution.*

Proof. For part (i), we begin by noticing that the expression in (9.121) is finite, because $f \in L_\alpha(m)$. The proof of (9.121) is very similar to the proof of (9.96) in Theorem 9.6.1. We leave checking the details to Exercise 9.10.21.

In order to prove the convergence in (9.122), we have to prove that for every $k = 1, 2, \ldots$ and $0 = t_0 < t_1 < \ldots < t_k$,

$$\begin{aligned} &\left(n^{-1/\alpha} \max_{nt_{j-1} < i \le nt_j} X_i,\ j = 1, \ldots, k \right) \\ &\Rightarrow \left(\big(m_\pm(f)\big)^{1/\alpha} (t_j - t_{j-1})^{1/\alpha}\, Y_j,\ j = 1, \ldots, k \right) \end{aligned} \tag{9.123}$$

as $n \to \infty$, where $Y_1, \ldots, Y_k$ on the right-hand side are i.i.d. standard Fréchet random variables. It is not hard to notice that it is enough to prove that (9.123) holds if $t_j = j$, $j = 1, \ldots, k$. Indeed, if (9.123) holds for $t_j = j$, $j = 1, \ldots, k$, then it

also holds for all choices of (t_j) that are nonnegative integers. This will then imply just by a change of notation that (9.123) holds for (t_j) that are nonnegative rational numbers, and then, using the obvious approximation, for all $0 = t_0 < t_1 < \ldots < t_k$.

We will prove now (9.123) with $t_j = j$, $j = 1, \ldots, k$. We begin by assuming that the function f is supported by finitely many translates of W, i.e., by assuming that (9.97) holds for some $m_2 = -m_1 = 1, 2, \ldots$. Write

$$X_i = \sum_{d=m_2-nk}^{-m_2-1} \int_{\phi^d(W)} f \circ \phi^i(s)\, M(ds) + \left[\sum_{d=-m_2-nk}^{m_2-nk-1} \int_{\phi^d(W)} f \circ \phi^i(s)\, M(ds) + \sum_{d=-m_2}^{m_2-1} \int_{\phi^d(W)} f \circ \phi^i(s)\, M(ds) \right],$$

$i = 1, \ldots, nk$. Define

$$Z_{i,d} = \int_{\phi^d(W)} f \circ \phi^i(s)\, M(ds),\ i, d \in \mathbb{Z}\,.$$

The claim (9.123) with $t_j = j$, $j = 1, \ldots, k$, will follow once we prove the following two statements. As $n \to \infty$,

$$\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} \sum_{d=m_2-nk}^{-m_2-1} Z_{i,d},\ j = 1, \ldots, k \right) \tag{9.124}$$
$$\Rightarrow \left(\left(m_{\pm}(f)\right)^{1/\alpha} Y_j,\ j = 1, \ldots, k \right),$$

$$n^{-1/\alpha} \max_{0<i\le n} \sum_{d=-m_2-n}^{m_2-n-1} Z_{i,d} \to 0,\quad n^{-1/\alpha} \max_{0<i\le n} \sum_{d=-m_2}^{m_2-1} Z_{i,d} \to 0\,. \tag{9.125}$$

We will begin with the first claim in (9.125). For $-m_2 - n \le d \le m_2 - n - 1$, let $I_d = [1, \ldots, n] \cap [-m_2 - d, \ldots, m_2 - d]$. Note that I_d contains at most $2m_2 + 1$ integers, and $Z_{i,d} = 0$ for $i = 1, \ldots, n$ unless $i \in I_d$. Therefore,

$$\left| \max_{0<i\le n} \sum_{d=-m_2-n}^{m_2-n-1} Z_{i,d} \right| \le \sum_{d=-m_2-n}^{m_2-n-1} \max_{0<i\le n} |Z_{i,d}| \tag{9.126}$$

is bounded by a sum of at most $2m_2(2m_2+1)$ random variables each of which is the absolute value of an SαS random variable with the scale not exceeding $\|f\|_\alpha$. This clearly establishes the first claim in (9.125), and the second claim can be proved in

the same way. It remains, therefore, to prove (9.124), and to do so, we will prove that for every $x_1 > 0, \ldots, x_j > 0$,

$$\lim_{n\to\infty} P\left(\max_{n(j-1)<i\le nj} \sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} \le n^{1/\alpha} x_j,\ j = 1, \ldots, k\right)$$
$$= \exp\left\{-m_{\pm}(f) \sum_{j=1}^{k} x_j^{-\alpha}\right\}, \tag{9.127}$$

and the first step toward doing so is to prove that

$$\lim_{n\to\infty} P\left(\max_{n(j-1)<i\le nj} Z_{i,d} \le n^{1/\alpha} x_j,\ j = 1, \ldots, k,\ d = m_2 - nk, \ldots, -m_2 - 1\right)$$
$$= \exp\left\{-m_{\pm}(f) \sum_{j=1}^{k} x_j^{-\alpha}\right\}. \tag{9.128}$$

Since the set W is wandering, the sets $\phi^d(W)$ are disjoint for different d. Therefore, the stochastic processes $\left(Z_{i,d},\ i \in \mathbb{Z}\right)$, $d \in \mathbb{Z}$, are independent. We conclude that the probability on the left-hand side of (9.128) is equal to

$$\prod_{d=m_2-nk}^{-m_2-1} P\left(\max_{n(j-1)<i\le nj} Z_{i,d} \le n^{1/\alpha} x_j,\ j = 1, \ldots, k\right)$$
$$= \prod_{d=m_2-nk}^{-m_2-1} P\left(\max_{n(j-1)<i\le nj} Z_{i+d,0} \le n^{1/\alpha} x_j,\ j = 1, \ldots, k\right),$$

with the last step following from the fact that the control measure m of the SαS random measure M is preserved under the map ϕ. Since (9.97) is assumed to hold, we know that $Z_{i,0} = 0$ for $|i| > m_2$. Set

$$D_n = \{m_2 - nk, \ldots, -m_2 - 1\} \setminus \bigcup_{j=1}^{k} \{m_2 - nj, \ldots, -m_2 - 1 - n(j-1)\}$$

and notice that the set D_n has $2m_2k$ elements. Furthermore, for all $j = 1, \ldots, k$ and $d \in \{m_2 - nj, \ldots, -m_2 - 1 - n(j-1)\}$, all such i that $|i + d| \le m_2$ are in the interval $n(j-1) < i \le nj$. We conclude that the probability on the left-hand side of (9.128) can be written in the form

$$\prod_{j=1}^{k} \left(P\left(\max_{|i|\le m_2} Z_{i,0} \le n^{1/\alpha} x_j\right)\right)^{n-2m_2}$$
$$\prod_{d\in D_n} P\left(\max_{n(j-1)<i\le nj} Z_{i+d,0} \le n^{1/\alpha} x_j,\ j = 1, \ldots, k\right).$$

Note that the second product above converges to 1 as $n \to \infty$, since the number of terms in the product is constant, and each probability in that product cannot be smaller than the probability

$$P\left(\max_{|i|\le m_2} Z_{i,0} \le n^{1/\alpha} \min_{j=1,\dots,k} x_j\right),$$

which converges to 1 as $n \to \infty$. Therefore, in order to prove (9.128), it is enough to check that

$$\lim_{x\to\infty} x^\alpha P\left(\max_{|i|\le m_2} Z_{i,0} > x\right) = m_\pm(f).$$

This, however, is also a special case of equivalence results between the tails of functionals of infinitely divisible processes and the corresponding tails with respect to the Lévy measure of Rosiński and Samorodnitsky (1993). Therefore, (9.128) follows.

The second ingredient for the proof of (9.127) is the claim that for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P\Big(\text{for some } i = 1,\dots,nk,\ Z_{i,d} > n^{1/\alpha}\varepsilon \text{ for more than one } d\Big) = 0. \tag{9.129}$$

This follows immediately by noticing that the probability in (9.129) does not exceed

$$\begin{aligned}
&\sum_{i=1}^{nk} P\Big(Z_{i,d} > n^{1/\alpha}\varepsilon \text{ for more than one } d\Big)\\
&= nk\, P\Big(Z_{0,d} > n^{1/\alpha}\varepsilon \text{ for more than one } d\Big)\\
&= nk\, P\Big(Z_{0,d} > n^{1/\alpha}\varepsilon \text{ for at least 2 different } d \in \{-m_2,\dots,m_2\}\Big)\\
&\le nk \binom{2m_2+1}{2} \max_{d\in\{-m_2,\dots,m_2\}} \Big(P\Big(Z_{0,d} > n^{1/\alpha}\varepsilon\Big)\Big)^2 \to 0
\end{aligned}$$

as $n \to \infty$, where we have taken into account the fact that $Z_{0,d} = 0$ for $|d| > 2m_2$.

Let $\theta > 0$ be a small number (smaller than $\min_j x_j$), and let $\varepsilon > 0$ be an even smaller number. Define the event

$$B_{n,\theta,\varepsilon} = \Big\{\max_{n(j-1)<i\le nj} Z_{i,d} \le n^{1/\alpha}(x_j - \theta),\ j = 1,\dots,k,\ d = m_2 - nk,\dots,-m_2 - 1,$$
$$\text{and for each } i = 1,\dots,nk,\ Z_{i,d} > n^{1/\alpha}\varepsilon \text{ for at most one } d\Big\}.$$

It follows from (9.129) that

$$P\left(\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right) \tag{9.130}$$

$$\le P\left(\max_{n(j-1)<i\le nj} Z_{i,d} > n^{1/\alpha}(x_j-\theta) \text{ for some } j=1,\dots,k,\right.$$

$$\left. + P\left(\left\{\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right\} \cap B_{n,\theta,\varepsilon}\right)\right.\,.$$

Note that on the event described in the second term on the right-hand side of (9.130), we necessarily have

$$\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d}\mathbf{1}\big(0\le Z_{i,d}\le n^{1/\alpha}\varepsilon\big) > n^{1/\alpha}\theta \text{ for some } j=1,\dots,k\,.$$

Therefore, by (9.128),

$$\limsup_{n\to\infty} P\left(\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right)$$

$$\le 1-\exp\left\{-m_{\pm}(f)\sum_{j=1}^{k}(x_j-\theta)^{-\alpha}\right\}$$

$$+\limsup_{n\to\infty} P\left(\max_{i=1,\dots,nk}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d}\mathbf{1}\big(0\le Z_{i,d}\le n^{1/\alpha}\varepsilon\big) > n^{1/\alpha}\theta\right)\,.$$

The last term above does not exceed

$$nk\,P\left(\sum_{d=-m_2}^{m_2} Z_{0,d}\mathbf{1}\big(0\le Z_{0,d}\le n^{1/\alpha}\varepsilon\big) > n^{1/\alpha}\theta\right)$$

$$\le nk\sum_{d=-m_2}^{m_2} P\Big(Z_{0,d}\mathbf{1}\big(0\le Z_{0,d}\le n^{1/\alpha}\varepsilon\big) > n^{1/\alpha}\theta/(2m_2+1)\Big)$$

$$\le n^{1-2/\alpha}\frac{k(2m_2+1)^{\alpha}}{\theta^{\alpha}}\sum_{d=-m_2}^{m_2} E\Big(Z_{0,d}^2\mathbf{1}\big(0\le Z_{0,d}\le n^{1/\alpha}\varepsilon\big)\Big)$$

$$\sim n^{1-2/\alpha}\frac{k(2m_2+1)^{\alpha}}{\theta^{\alpha}}\sum_{d=-m_2}^{m_2}\frac{\alpha}{2-\alpha}\big(n^{1/\alpha}\varepsilon\big)^2 P\big(Z_{0,d}>n^{1/\alpha}\varepsilon\big)$$

$$\to c_{k,m_2}\theta^{-\alpha}\varepsilon^{2-\alpha}\,,$$

where c_{k,m_2} is a constant depending only on k and m_2. In this computation we used Proposition 4.2.3 to estimate a truncated second moment of an SαS random variable. Therefore,

$$\limsup_{n\to\infty} P\left(\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right)$$
$$\le 1-\exp\left\{-m_{\pm}(f)\sum_{j=1}^{k}(x_j-\theta)^{-\alpha}\right\} + c_{k,m_2}\theta^{-\alpha}\varepsilon^{2-\alpha}\,.$$

Letting first $\varepsilon\to 0$ and then $\theta\to 0$ shows that

$$\limsup_{n\to\infty} P\left(\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right)$$
$$\le 1-\exp\left\{-m_{\pm}(f)\sum_{j=1}^{k}x_j^{-\alpha}\right\}\,.$$

The matching lower bound,

$$\liminf_{n\to\infty} P\left(\max_{n(j-1)<i\le nj}\sum_{d=m_2-nk}^{-m_2-1} Z_{i,d} > n^{1/\alpha}x_j \text{ for some } j=1,\dots,k\right)$$
$$\ge 1-\exp\left\{-m_{\pm}(f)\sum_{j=1}^{k}x_j^{-\alpha}\right\}\,, \tag{9.131}$$

can be proved in essentially the same way. We leave the details to Exercise 9.10.22. This proves (9.124), and hence we have established (9.123) (with $t_j = j, j = 1,\dots,k$) in the case that the function f is supported by finitely many translates of W.

In the case of a general $f\in L_\alpha(m)$, we define for $m_2 = 1,2,\dots$ a function in $L_\alpha(m)$ that satisfies (9.97) by $f_{m_2} = f\mathbf{1}(\cup_{i=-m_2}^{m_2}\phi^i(W))$. Notice that $\|f-f_{m_2}\|_\alpha\to 0$ as $m_2\to\infty$. With the same SαS random measure M as in (9.37) that defined the process $\mathbf{X}$, we define a new stationary SαS process by

$$X_n^{(m_2)} = \int_E f_{m_2}\circ\phi^n(s)\,M(ds),\ n\in\mathbb{Z}\,.$$

Let $0 < \varepsilon < 1$. Obviously,

$$
\begin{aligned}
& P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} X_i^{(m_2)} > x_j(1+\varepsilon) \text{ for some } j = 1,\dots,k\right) \\
& - P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} \left(X_i^{(m_2)} - X_i\right) > x_j\varepsilon \text{ for some } j = 1,\dots,k\right) \\
& \le P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} X_i > x_j \text{ for some } j = 1,\dots,k\right) \\
& \le P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} X_i^{(m_2)} > x_j(1-\varepsilon) \text{ for some } j = 1,\dots,k\right) \\
& + P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} \left(X_i - X_i^{(m_2)}\right) > x_j\varepsilon \text{ for some } j = 1,\dots,k\right).
\end{aligned}
$$

Since $m_\pm\left(f_{m_2}\right) \to m_\pm(f)$ as $m_2 \to \infty$ by the monotone convergence theorem, (9.123) (with $t_j = j$, $j = 1,\dots,k$) will follow once we prove that for every $0 < \varepsilon < 1$,

$$
\lim_{m_2\to\infty} \limsup_{n\to\infty} P\left(n^{-1/\alpha} \max_{n(j-1)<i\le nj} \left|X_i - X_i^{(m_2)}\right| > x_j\varepsilon \text{ for some } j = 1,\dots,k\right) = 0\,.
$$

However, the probability above does not exceed

$$
\begin{aligned}
& P\left(\max_{1\le i\le nk} \left|X_i - X_i^{(m_2)}\right| > n^{1/\alpha} \min_{j=1,\dots,k} x_j\varepsilon\right) \\
& \le \sum_{i=1}^{nk} P\left(\left|X_i - X_i^{(m_2)}\right| > n^{1/\alpha} \min_{j=1,\dots,k} x_j\varepsilon\right).
\end{aligned}
$$

As $n \to \infty$, this bound converges to

$$
k\left(\min_{j=1,\dots,k} x_j\varepsilon\right)^{-\alpha} \|f - f_{m_2}\|_\alpha^\alpha\,,
$$

and this expression converges to 0 as $m_2 \to \infty$. Therefore, the proof of the theorem is complete. □

Theorem 9.8.1 says that from the point of view of the partial maxima, all stationary SαS processes of the form (9.37) with a dissipative map ϕ have short memory. We show next that if the map ϕ is conservative, then the SαS process has long memory from the point of view of the partial maxima. Recall that by Proposition 9.7.3, long-range dependence can result only in a slower rate of increase of the partial maxima.

Theorem 9.8.2. *Let* $\mathbf{X}$ *be a stationary SαS process given by* (9.37) *with some* $f \in L_\alpha(m)$. *Suppose that the map* ϕ *is conservative with respect to the measure* m. *Then*

$$n^{-1/\alpha} \max_{i=1,\dots,n} |X_i| \to 0 \ \textit{in probability} \tag{9.132}$$

as $n \to \infty$.

Proof. We begin with the case that f is an indicator function. Let $A \in \mathcal{E}$, with $0 < m(A) < \infty$, and let $f = \mathbf{1}_A$. Define

$$B_n = \cup_{i=1}^n \phi^{-i}(A), \ n = 1, 2, \dots .$$

We claim that

$$\lim_{n\to\infty} \frac{1}{n} m(B_n) = 0 . \tag{9.133}$$

To see this, let

$$A_k = \left\{x \in E : \phi^k(x) \in A, \ \phi^j(x) \notin A, \ j = 1, \dots, k-1\right\}, \ k = 1, 2, \dots .$$

Since these sets are disjoint and $B_n = \cup_{k=1}^n A_k$, we need to show only that

$$\lim_{k\to\infty} m(A_k) = 0 . \tag{9.134}$$

Let $\psi : E \to E$ be the inverse map, $\psi = \phi^{-1}$. Clearly, ψ is conservative if ϕ is. Since ϕ and ψ preserve the measure m, we have $m(A_k) = m(\tilde{A}_k)$, where

$$\tilde{A}_k = \left\{x \in A : \psi^i(x) \notin A, \ i = 1, \dots, k-1\right\} .$$

However, since ψ is a conservative map, it follows by Exercise 2.6.12 that $\tilde{A}_k \downarrow \emptyset$ modulo a null set. Since $\tilde{A}_k \subset A$ for all k, we conclude that $m(\tilde{A}_k) \to 0$ as $k \to \infty$, and so (9.134) follows.

Set $\tilde{M}_n = \max_{i=1,\dots,n} |X_i|$, $n = 1, 2, \dots$. For $\varepsilon > 0$, we write

$$P\big(\tilde{M}_n > \varepsilon n^{1/\alpha}\big) = \sum_{k=1}^n P\Big(|X_k| > \varepsilon n^{1/\alpha}, \ |X_i| \le \varepsilon n^{1/\alpha}, \ i = 1, \dots, k-1\Big) .$$

Since $f = \mathbf{1}_A$, we know that $X_k = M\big(\phi^{-k}(A)\big)$, $k = 1, \dots, n$, and we can write for $k = 1, \dots, n$,

$$\begin{aligned} &P\left(|X_k| > \varepsilon n^{1/\alpha}, \ |X_i| \le \varepsilon n^{1/\alpha}, \ i = 1, \dots, k-1\right) \\ &= P\left(|X_k| > \varepsilon n^{1/\alpha}, \ |X_i| \le \varepsilon n^{1/\alpha}, \ i = 1, \dots, k-1,\right. \end{aligned}$$

$$\begin{aligned}
&\left|M\big(B_{k-1}\setminus\phi^{-k}(A)\big)\right|\le\varepsilon n^{1/\alpha}/3,\ \left|M\big(B_n\setminus B_k\big)\right|\le\varepsilon n^{1/\alpha}/3\Big)\\
&+P\Big(|X_k|>\varepsilon n^{1/\alpha},\ |X_i|\le\varepsilon n^{1/\alpha},\ i=1,\dots,k-1,\\
&\qquad\left|M\big(B_{k-1}\setminus\phi^{-k}(A)\big)\right|>\varepsilon n^{1/\alpha}/3\ \text{ or }\ \left|M\big(B_n\setminus B_k\big)\right|>\varepsilon n^{1/\alpha}/3\Big)\\
&:=p_{1,k,n}+p_{2,k,n}\,.
\end{aligned}$$

Since the SαS random measure M assigns independent values to disjoint sets, it follows that

$$\begin{aligned}
p_{2,k,n}\le&P\Big(\left|M\big(\phi^{-k}(A)\big)\right|>\varepsilon n^{1/\alpha}\Big)P\Big(\left|M\big(B_{k-1}\setminus\phi^{-k}(A)\big)\right|>\varepsilon n^{1/\alpha}/3\Big)\\
&+P\Big(\left|M\big(\phi^{-k}(A)\big)\right|>\varepsilon n^{1/\alpha}\Big)P\Big(\left|M\big(B_n\setminus B_k\big)\right|>\varepsilon n^{1/\alpha}/3\Big)\\
\le&\,2P\big(|X_1|>\varepsilon n^{1/\alpha}\big)P\big(|M(B_n)|>\varepsilon n^{1/\alpha}/3\big),
\end{aligned}$$

and hence

$$\begin{aligned}
\sum_{k=1}^{n}p_{2,k,n}\le&2nP\big(|X_1|>\varepsilon n^{1/\alpha}\big)P\big(|M(B_n)|>\varepsilon n^{1/\alpha}/3\big)\\
\sim&2n\,\left(\varepsilon^{-\alpha}n^{-1}m(A)\right)\left((\varepsilon/3)^{-\alpha}n^{-1}m(B_n)\right)\\
\to&\,0
\end{aligned}$$

as $n\to\infty$ by (9.133). On the other hand,

$$p_{1,k,n}\le P\Big(|X_k|>\varepsilon n^{1/\alpha},\ |X_i|\le\varepsilon n^{1/\alpha},\ i=1,\dots,k-1,\ |M(B_n)|>\varepsilon n^{1/\alpha}/3\Big)\,,$$

and hence

$$\begin{aligned}
\sum_{k=1}^{n}p_{1,k,n}&\le P\big(|M(B_n)|>\varepsilon n^{1/\alpha}/3\big)\\
&\sim(\varepsilon/3)^{-\alpha}n^{-1}m(B_n)\to0
\end{aligned}$$

as $n\to\infty$, once again by (9.133). Therefore,

$$P\left(\tilde{M}_n>\varepsilon n^{1/\alpha}\right)\to0$$

as $n\to\infty$ for every $\varepsilon>0$. That is, we have proved (9.132) in the case that f is an indicator function.

If the function f is a simple function of the type $f = \sum_1^k a_j \mathbf{1}_{A_j}$, with $a_i \in \mathbb{R}$ and $0 < m(A_j) < \infty$ for $j = 1, \ldots, k$, then by the linearity of the integral,

$$\max_{i=1,\ldots,n} |X_i| = \max_{i=1,\ldots,n} \left| \sum_{j=1}^{k} a_j \int_E \mathbf{1}_{A_j} \circ \phi^i(s)\, M(ds) \right|$$

$$\leq \sum_{j=1}^{k} |a_j| \max_{i=1,\ldots,n} \left| \int_E \mathbf{1}_{A_j} \circ \phi^i(s)\, M(ds) \right|,$$

and hence (9.132) follows from its validity in the case that f is an indicator function.

Finally, in the case of a general $f \in L_\alpha(m)$, for every $\delta > 0$ we can find a simple function f_δ as above with $\|f - f_\delta\|_\alpha \leq \delta$. Then for every $\varepsilon > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\max_{i=1,\ldots,n} |X_i| > \varepsilon n^{1/\alpha}\right) \\
&\leq \limsup_{n\to\infty} P\left(\max_{i=1,\ldots,n} \left| \int_E f_\delta \circ \phi^i(s)\, M(ds) \right| > \varepsilon n^{1/\alpha}/2\right) \\
&+ \limsup_{n\to\infty} P\left(\max_{i=1,\ldots,n} \left| \int_E (f - f_\delta) \circ \phi^i(s)\, M(ds) \right| > \varepsilon n^{1/\alpha}/2\right) \\
&= \limsup_{n\to\infty} P\left(\max_{i=1,\ldots,n} \left| \int_E (f - f_\delta) \circ \phi^i(s)\, M(ds) \right| > \varepsilon n^{1/\alpha}/2\right),
\end{aligned}$$

since (9.132) holds when f is a simple function. Since the probability in the last line does not exceed

$$\sum_{i=1}^{n} P\left(\left| \int_E (f - f_\delta) \circ \phi^i(s)\, M(ds) \right| > \varepsilon n^{1/\alpha}/2\right),$$

we conclude that

$$\begin{aligned}
&\limsup_{n\to\infty} P\Big(\max_{i=1,\ldots,n} |X_i| > \varepsilon n^{1/\alpha}\Big) \\
&\leq n\big((\varepsilon/2)^{-\alpha} n^{-1} \|f - f_\delta\|_\alpha^\alpha\big) \leq (\varepsilon/2)^{-\alpha} \delta^\alpha,
\end{aligned}$$

and we obtain (9.132) by letting $\delta \to 0$. □

Remark 9.8.3. Theorems 9.8.1 and 9.8.2 provide a clear classification of stationary SαS processes into those with short memory according to the behavior of the partial maxima and those with long-range dependence in the same sense. Indeed, if $\mathbf{X}$ is a stationary SαS process given by (9.37) with some $f \in L_\alpha(m)$, then $\mathbf{X}$ has long-range dependence according to the behavior of the partial maxima if and only if the dissipative part in the Hopf decomposition of the map ϕ (i.e., the dissipative part

of ϕ) vanishes. To see this, note that we can write

$$X_n = X_n^{(\mathrm{d})} + X_n^{(\mathrm{c})},\ n \in \mathbb{Z},$$

where $\mathbf{X}^{(\mathrm{d})}$ and $\mathbf{X}^{(\mathrm{c})}$ are two independent stationary processes, the former corresponding to a dissipative map, and the latter corresponding to a conservative map. If the dissipative part of ϕ does not vanish, then the process $\mathbf{X}^{(\mathrm{d})}$ is a nonzero process, and it follows from Theorems 9.8.1 and 9.8.2 that

$$\left(n^{-1/\alpha} M_n(t),\ t \geq 0\right) \Rightarrow \left(\left(m_{\pm}(f)\right)^{1/\alpha} Y_{\Phi_\alpha}(t),\ t \geq 0\right) \ \text{ as } n \to \infty$$

in the sense of convergence of finite-dimensional distributions, where $\mathbf{Y}_{\Phi_\alpha}$ is the extremal process corresponding to the standard Fréchet distribution. Here $m_{\pm}(f)$ is as in Theorem 9.8.1, but now $\cup_{k=-\infty}^{\infty} \phi^k(W) = \mathcal{D}(\phi)$. On the other hand, if the dissipative part of ϕ does vanish, then the process $\mathbf{X}$ corresponds to a conservative map, and Theorem 9.8.2 applies. In this case, the partial maxima grow at a rate strictly lower than the rate of $n^{1/\alpha}$, which is an indication of long-range dependence with respect to the partial maxima.

It is important to consider for a moment what happens when both the dissipative part of ϕ and the conservative part of ϕ do not vanish. In that case, both processes $\mathbf{X}^{(\mathrm{d})}$ and $\mathbf{X}^{(\mathrm{c})}$ are nontrivial stationary SαS processes. Our discussion in this section shows that the process $\mathbf{X}^{(\mathrm{d})}$ has short memory with respect to the behavior of the partial maxima, while the process $\mathbf{X}^{(\mathrm{c})}$ has long memory. In this situation, the process $\mathbf{X}$, which is the sum of $\mathbf{X}^{(\mathrm{d})}$ and $\mathbf{X}^{(\mathrm{c})}$, still has short memory with respect to the behavior of the partial maxima, even though it may be counterintuitive that the long memory of the process $\mathbf{X}^{(\mathrm{c})}$ disappears in the sum.

We will demonstrate what limits may be obtained for partial maxima of stationary SαS processes defined by (9.37) in which the map ϕ is conservative by considering the partial maxima of a stationary SαS process with a representation in (9.37) given, once again, by the null recurrent Markov chain of the setup of Theorem 9.4.7. In the next theorem, which we do not prove, the limit is no longer the extremal process corresponding to the Fréchet distribution. However, the marginal distributions of the limiting process are still Fréchet.

Theorem 9.8.4. *Let* $\mathbf{X}$ *be a stationary SαS process given in the form* (9.37)*, with* $E = \mathbb{Z}^{\mathbb{Z}}$*,* m *a shift-invariant measure on* E *generated by an invariant measure of an irreducible null recurrent Markov chain on* $\mathbb{Z}$ *given in* (9.55)*, and* f *the indicator function given by* (9.56)*. Let* ϕ *be the left shift on* E*. Assume that the sequence* $P_0(\tau_1 > n)$*,* $n = 1, 2, \ldots$*, is regularly varying with exponent* $-\beta \in (-1/2, 0)$*. Then the partial maxima of the process* $\mathbf{X}$ *satisfy*

$$\left(b_n^{-1} M_n(t),\ t \geq 0\right) \Rightarrow \left(C_\alpha^{1/\alpha} Y_{\Phi_\alpha}\left(t^{1-\beta}\right),\ t \geq 0\right) \ \text{as } n \to \infty \tag{9.135}$$

in the Skorokhod J_1 topology on $D[0,\infty)$, where $\mathbf{Y}_{\Phi_\alpha}$ is the extremal process corresponding to the standard Fréchet distribution. Here C_α is given by (3.31), and the normalizing sequence satisfies

$$b_n^{1/\alpha} = \sum_{i\in\mathbb{Z}} \pi_i P_i(\tau_1 \le n),\ n = 1, 2, \dots ,$$

with (π_i) being the invariant law of the Markov chain defining the measure m. It is regularly varying with exponent $(1-\beta)/\alpha$.

We refer the reader to Owada and Samorodnitsky (2015b) for a proof. Note that the assumptions of Theorem 9.8.4 restrict the parameter β to the subinterval $(0, 1/2)$ of the full interval $(0, 1)$ allowed both in Theorem 9.4.7 and in Theorem 9.6.5.

9.9 Comments on Chapter 9

Comments on Section 9.1

The phase transition approach to long-range dependence was proposed in Samorodnitsky (2004).

Comments on Section 9.2

The classical invariance principle for a sequence of i.i.d. random variables with finite variance was proved in Donsker (1951).

For general stationary processes with a finite variance, drawing a precise boundary between short and long memory, as far as the behavior of the partial sums is concerned, is not easy. A number of results have been proved establishing functional central limit theorems with a Brownian limit under various strong mixing conditions. This is the source of common association between strong mixing and short memory. The following result is in Merlevéde et al. (2006), Proposition 34.

Theorem 9.9.1. *Assume that $\mathbf{X}$ is a zero-mean strongly mixing process such that $E|X_0|^p < \infty$ for some $p > 2$, and (9.7) holds. If the variance of the partial sums satisfies $s_n \to \infty$ as $n \to \infty$, then*

$$\left(s_n^{-1} S_n(t),\ t \ge 0\right) \Rightarrow \left(B(t),\ t \ge 0\right) \ \text{as } n \to \infty$$

weakly in the Skorokhod J_1 topology on $D[0,\infty)$.

It follows from Theorem 8.1.5 that the sequence (s_n) in Theorem 9.9.1 must be regularly varying with exponent $1/2$. Because of this and the Brownian limit, we should view a process satisfying the assumptions of Theorem 9.9.1 as having short memory. Under certain strong mixing conditions (some of which are stronger than the mixing conditions considered in Section 2.3), one can get rid of the extra moment assumption in Theorem 9.9.1, and the sequence (s_n^2) of the variances of the partial sums grows linearly, so that (9.4) holds. See Peligrad (1998) and Bradley

(1993) for a discussion of the so-called *interlaced strongly mixing* condition. For a stationary Gaussian process to be strongly mixing, it is necessary that its spectral measure be absolutely continuous with respect to the Lebesgue measure on $(-\pi, \pi]$. If the spectral density (i.e., the derivative of the spectral measure with respect to the Lebesgue measure on $(-\pi, \pi]$) is continuous and positive, then the process is interlaced strong mixing (see Kolmogorov and Rozanov (1960) and Rosenblatt (1985)).

An interesting phase transition phenomenon was discovered in a series of papers published in the 1970s and 1980s treating the partial sums of stationary processes of the type considered in Section 6.3. Recall that those are stationary processes of the type $Y_n = g(X_n)$, $n = 1, 2, \ldots$, where $\mathbf{X}$ is a stationary centered Gaussian process with unit variance, and $g : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $Eg(X_n)^2 < \infty$. Let k_g be the Hermite index of g; assume it to be finite. Let R_X be the covariance function of the Gaussian process $\mathbf{X}$. It was proved in Breuer and Major (1983) that if

$$\sum_{n=0}^{\infty} |R_X(n)|^{k_g} < \infty\,, \tag{9.136}$$

then (9.4) holds (in terms of convergence of the finite-dimensional distributions) with some $\sigma \geq 0$. If (9.136) holds and $\sigma > 0$, then the process $\mathbf{Y}$ has short memory with respect to the behavior of its partial sums.

On the other hand, suppose that the covariance function R_X is regularly varying, i.e., that (6.9) holds with $-1 < d < 0$. The following result was proved (separately) in Dobrushin and Major (1979) and in Taqqu (1979). Assume that

$$k_g < -\frac{1}{d}\,. \tag{9.137}$$

Then the partial sum process of the stationary process $\mathbf{Y}$ satisfies

$$\left(\frac{1}{nR_X(n)^{k_g/2}} S_n(t),\, t \geq 0\right) \Rightarrow \big(\sigma Z(t),\, t \geq 0\big)\,,$$

as $n \to \infty$, in terms of convergence of the finite-dimensional distributions, where $\sigma > 0$ and $\mathbf{Z}$ is the self-similar process with stationary increments corresponding to the Taqqu kernel of Example 8.3.3 with $H = 1 - dk_g/2$. That is, under the assumption of the regular variation of R_X and condition (9.137), the process $\mathbf{Y}$ has long-range dependence with respect to the behavior of its partial sums.

Comments on Section 9.3

A setup that is supposed to contain finite-variance infinite-moving-average processes with both Brownian and non-Brownian limits in the functional central limit theorem situation is in Davydov (1970). Functional limit theorems with Brownian limits for such processes were subsequently derived a number of times,

in situations in which the noise process $(\varepsilon_n,\ n \in \mathbb{Z})$ is not necessarily an i.i.d. sequence but may be a more general process, often a martingale difference (see, e.g., Hannan (1979), Merlevéde and Peligrad (2006), and Merlevéde et al. (2006)), under conditions that are sometimes weaker than the summability of the coefficients (9.14). On the long-memory side, a functional central limit theorem under the regular variation assumption (9.22) with $\beta \in (-1, -1/2)$, for one-sided infinite moving averages, was proven in Peligrad and Sang (2012) (this theorem even allows an infinite second moment of the noise variables, as long as the truncated second moment is slowly varying).

Comments on Section 9.5

Rigorous topological convergence in the heavy-tailed invariance principle (9.63) was established in Skorohod (1957). Extending convergence in Theorem 9.5.5 to weak convergence in $D[0,\infty)$ requires additional assumptions on the coefficients of the moving-average process. If more than one of the coefficients is different from zero (i.e., if the process $\mathbf{X}$ is not an i.i.d. sequence), then weak convergence in the J_1 topology on $D[0,\infty)$ is impossible, as was shown in Avram and Taqqu (1992), in which it was also shown that if the coefficients are nonnegative and only finitely many of them are different from zero, then convergence in Theorem 9.5.5 holds as weak convergence in $D[0,\infty)$ with the Skorokhod M_1 topology. The fact that M_1 weak convergence holds even if infinitely many coefficients are different from zero (but all are still nonnegative) was proved in Louhichi and Rio (2011). If the coefficients are not necessarily nonnegative, than an even weaker topology on the space $D[0,\infty)$ is needed, and weak convergence in the so-called S topology was shown in Balan et al. (2014).

Comments on Section 9.6

In a series of papers beginning with Pipiras and Taqqu (2002a) and Pipiras and Taqqu (2002b), the authors studied SαS processes of the form

$$Y(t) = \int_G \int_{\mathbb{R}} \big(F(y, t+u) - F(y,u)\big)\, M(dy, du),\ t \in \mathbb{R}\,, \tag{9.138}$$

where M is an SαS random measure on the space $E = G \times \mathbb{R}$, $(G, \mathcal{G}, \eta)$ is a σ-finite measure space, and the control measure of M is $\eta \times \text{Leb}$ (the function F has to satisfy the appropriate integrability condition). Under this setup, the process $\mathbf{Y}$ has stationary increments, and the stationary increment process

$$X_n = Y(n) - Y(n-1) = \int_G \int_{\mathbb{R}} \big(F(y, n+u) - F(y, n-1+u)\big)\, M(dy, du),\ n \in \mathbb{Z}$$

always corresponds to a dissipative map $\phi : E \to E$ given by $\phi(y,u) = (y, u+1)$. In some cases, the process $\mathbf{Y}$ in (9.138) is also self-similar, and Pipiras and Taqqu classified the self-similar processes with stationary increments that can be obtained in this way. They accomplished this by studying the multiplicative flows appearing naturally through the self-similarity of the process $\mathbf{Y}$ in its integral representation,

similarly to the classification of stationary SαS processes of Rosiński (1995). The notions of a dissipative map and a conservative map appear once again, but this time, they are applied to the multiplicative action related to self-similarity. In particular, the linear fractional stable motion of Example 3.5.2 turns out to be of the form (9.138), with a conservative multiplicative action, while the dilated fractional stable motions of Example 9.6.3 are also of the form (9.138), but with a dissipative multiplicative action. The latter processes are studied in detail in Pipiras and Taqqu (2004a). The telecom process was introduced in Pipiras and Taqqu (2000). Further details on self-similar processes with stationary increments of the form (9.138) are in Pipiras and Taqqu (2004b, 2007, 2008).

The statement of Theorem 9.6.5 is a special case of a more general result in Owada and Samorodnitsky (2015a). Another functional central limit theorem for the partial sums of an infinite-variance stationary infinitely divisible process with regularly varying tails and corresponding to a conservative map ϕ in (9.37) is in Jung et al. (2016). In that case, the limiting process is yet another self-similar stationary increments SαS process.

Comments on Section 9.7

The classical extreme value theory studies limit theorems of the type

$$\frac{1}{a_n}\big(\max(X_1,\dots,X_n)-b_n\big) \Rightarrow Y \tag{9.139}$$

for an i.i.d. sequence $X_1, X_2, \dots$, a positive sequence (a_n), and a real sequence (b_n). If the distribution of the limiting random variable Y is nondegenerate, then up to a linear transformation of the type $Y \to aY + b$, $a > 0$ and b real, this limiting distribution can take one of the three possible forms (Fisher and Tippett 1928; Gnedenko 1943). Two of the possible limiting distributions are the Gumbel distribution in (9.107) and the Fréchet distribution in (9.108). The third possible type of limiting distribution is the reverse Weibull distribution

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\} & \text{if } x < 0, \\ 1 & \text{if } x \geq 0\,, \end{cases}$$

$\alpha > 0$. The distributions in the domain of attraction of the reverse Weibull distribution have support bounded on the right. Since we are interested in the "size" of the maxima of a process, we are not considering the case of the Weibull domain of attraction. Some distributions in the Gumbel domain of attraction also have support bounded on the right, and our main interest lies in those distributions in the Gumbel domain of attraction whose support is unbounded on the right. There are many texts on the classical extreme value theory; two of them are Resnick (1987) and deHaan and Ferreira (2006).

The fact that for an i.i.d. sequence, the weak convergence in (9.139) implies functional weak convergence

$$\left(\frac{1}{a_n}(M_n(t)-b_n),\ t>0\right) \Rightarrow \big(Y_F(t),\ t>0\big)$$

(where F is the law of Y in (9.139)) weakly in the Skorokhod J_1-topology on $D(0,\infty)$ was proved in Lamperti (1964), with several subsequent additions and clarifications, including Resnick (1975) and Weissman (1975).

Comments on Section 9.8

The restriction of the parameter β to the interval $(0, 1/2)$ in Theorem 9.8.4 is a consequence of the fact that the result in Owada and Samorodnitsky (2015b) applies only in this case. We expect the memory in the process $\mathbf{X}$ to become longer as β increases from 0 to 1; see Remark 9.4.8. Therefore, Theorem 9.8.4 does not address the behavior of the partial maxima when the memory in the process $\mathbf{X}$ is the longest.

9.10 Exercises to Chapter 9

Exercise 9.10.1. *Let* $\mathbf{Y} = \big(Y(t),\, t \in \mathbb{R}\big)$ *be the standard Ornstein–Uhlenbeck process, i.e., a centered stationary Gaussian process with covariance function* $R_Y(t) = e^{-|t|}$ *(see Example 1.2.4). Let* V *be a gamma random variable independent of* $\mathbf{Y}$ *with unit scale and shape* $0 < \alpha < 1$. *Define* $X_n = Y(nV)$, $n \in \mathbb{Z}$.

(a) *Prove that* $\mathbf{X}$ *is a stationary zero-mean finite-variance process with covariance function* $R(n) = (n+1)^{-\alpha}$, $n = 0, 1, 2, \ldots$.

(b) *Prove that the standard deviations (9.5) of the partial sums of* $\mathbf{X}$ *satisfy* $s_n \sim c_\alpha n^{1-\alpha/2}$ *for some* $c_\alpha \in (0,\infty)$ *as* $n \to \infty$.

(c) *Prove that*

$$n^{-1/2} S_n \Rightarrow Y(0) \left(\frac{e^V + 1}{e^V - 1} \right)^{1/2} \quad \text{as } n \to \infty,$$

so that $S_n / s_n \to 0$ *in probability.*

Exercise 9.10.2. *Let* X *be a nonnegative random variable such that* $EX^p < \infty$ *for some* $p > 1$. *Prove the inequality*

$$\big(P\big(X > \delta EX\big)\big)^{p-1} \geq (1-\delta)^p \frac{(EX)^p}{EX^p}$$

for $0 < \delta < 1$.

Exercise 9.10.3. *Let* $f \in D[0,\infty)$, *and for* $n = 1, 2, \ldots$, *define* $f_n \in D[0,\infty)$ *by*

$$f_n(t) = f\big(n^{-1}[nt]\big),\ t \geq 0\,.$$

Prove that $f_n \to f$ *as* $n \to \infty$ *in the Skorokhod* J_1 *topology.* ***Hint:*** *A useful property of functions in* $D[0,\infty)$ *is that they can be approximated uniformly, on compact intervals, by piecewise constant functions. That is, for every* $T \in (0,\infty)$ *and* $\varepsilon > 0$,

there are points $0 = t_0 < t_1 < \ldots < t_m = T$ *such that* $|f(t) - f(t_i)| \le \varepsilon$ *for every* $t_i \le t < t_{i+1}$, $i = 0, \ldots, m-1$. *See, e.g., Whitt (2002).*

Exercise 9.10.4. *Prove the second part of* (9.19)*:*

$$\lim_{n\to\infty} P\Big(n^{-1/2} \sup_{0\le t\le 1} \big|C_{[nt]}\big| > \epsilon\Big) = 0\,.$$

Exercise 9.10.5. *Prove* (9.32) *and* (9.33).

Exercise 9.10.6. *Proof Theorem 9.3.4.* ***Hint:*** *Use the assumption* (9.35) *to reduce the sums of the coefficients to infinite sums of the type* $\sum_{-\infty}^{\cdot}$ *and* $\sum_{\cdot}^{\infty}$. *To prove tightness under a stronger moment assumption, use the Marcinkiewicz–Zygmund inequalities of Theorem 10.7.2.*

Exercise 9.10.7. *Consider the partial sums of a stationary process* $\mathbf{X}$. *Let* (a_n) *be a sequence of positive numbers with* $a_n \to \infty$. *Assume that for every* $m = 1, 2, \ldots$ *and positive integers* $j_1, \ldots, j_m$,

$$a_n^{-1}\big(S_{j_1 n}, \ldots, S_{j_m n}\big) \Rightarrow \big(Y(j_1), \ldots, Y(j_m)\big)\,,$$

where $\big(Y(t),\ t \ge 0\big)$ *is a continuous in probability process. Prove that*

$$\big(a_n^{-1} S_n(t),\ t \ge 0\big) \Rightarrow \big(Y(t),\ t \ge 0\big) \ \text{ as } n \to \infty$$

in finite-dimensional distributions.

Exercise 9.10.8. *Prove that the left shift on the space* $E = \mathbb{Z}^{\mathbb{Z}}$ *is conservative with respect to the measure* m *defined by* (9.55).

Exercise 9.10.9. *Prove that the first return time defined in* (9.57) *satisfies*

$$\sum_{i=-\infty}^{\infty} \pi_i P_i(\tau_1 = n) = \pi_0 P_0(\tau_1 \ge n),\ n = 1, 2, \ldots\,.$$

Exercise 9.10.10. *In the Markov chain setup of Theorem 9.4.7, let* N_n *be the number of times the Markov chain visits state 0 in the time interval* $\{1, \ldots, n\}$. *Prove that starting in state 0 (and hence in any other initial state), the Markov chain satisfies*

$$P_0(\tau_1 > n) N_n \Rightarrow Z_\beta^{-\beta} \ \text{ as } n \to \infty,$$

where Z_β *is a positive* β*-stable random variable with characteristic triple* $(0, \mu, 0)$, *where the Lévy measure is given by* $\mu(dx) = \beta x^{-(1+\beta)}\,dx$ *for* $x > 0$. *Moreover, prove that*

$$E_0\Big[\big(P_0(\tau_1 > n) N_n\big)^p\Big] \to E\big(Z_\beta^{-p\beta}\big) \ \text{ as } n \to \infty$$

for every $p > 0$.

Exercise 9.10.11. *Still in the Markov chain setup of Theorem 9.4.7, use Exercise 3.8.2 to show that*

$$\mathrm{Var}(S_n) = \int_{-\infty}^{\infty} x^2 \, \rho(dx) \sum_{i=-\infty}^{\infty} \pi_i E_i(N_n^2) \, .$$

Use an argument similar to the one used in the proof of Theorem 9.3.3 to establish the tightness claim of the theorem.

Exercise 9.10.12. *Prove (9.81).*

Exercise 9.10.13. *Prove tightness in Theorem 9.5.7.*

Exercise 9.10.14. *Prove Theorem 9.5.8.*

Exercise 9.10.15. *In the proof of Theorem 9.6.1, we proved that the limit b in (9.87) is given by (9.96) in the case $1 < \alpha < 2$. Prove that (9.96) is valid in the case $0 < \alpha \leq 1$ as well.*

Exercise 9.10.16. *Prove that (9.92) holds in the case $1 < \alpha < 2$ under the conditions of Theorem 9.6.1.*

Exercise 9.10.17. *Prove that the telecom process in Example 9.6.3 is well defined.*

Exercise 9.10.18. *Prove part (ii) of Proposition 9.6.4.*

Exercise 9.10.19. *Check that (9.85) holds under the conditions of Theorem 9.6.5. Conclude that*

$$\lim_{x \to \infty} \frac{P(X_0 > x)}{\rho\big((x, \infty)\big)} = \pi_0 \, .$$

Exercise 9.10.20. *Prove that the standard normal random distribution satisfies (9.107) with the sequences (a_n) and (b_n) given by (9.119).*

Exercise 9.10.21. *Prove that (9.121) holds by following the lines of the argument for (9.96).*

Exercise 9.10.22. *Prove the lower bound (9.131) in Theorem 9.8.1, assuming that the function f is supported by finitely many translates of the set W.*

Chapter 10
Appendix

The main reason for this chapter is to give the reader easy access to some of the notions and the results used throughout this book. Most of the material can be found in book form elsewhere. The reader will notice that the quantity of detail is different from topic to topic, depending on how accessible the alternative references are.

10.1 Topological Groups

For more details on the notions presented in this section, see Chapter 1 in Rudin (1962) and Chapter 6 in Hewitt and Ross (1979).

Let G be a group whose group operation is written multiplicatively. Suppose that G, viewed as a set, is given a topology such that both

$$\text{the mapping } P: G \times G \to G: \quad P(x, y) = xy \text{ is continuous, and}$$

$$\text{the mapping } I: G \to G: \quad I(x) = x^{-1} \text{ is continuous.}$$

Then G is called a *topological group*. This name is often further modified both by the properties of the group and those of the topology. For example, if the group is abelian, then G is an abelian topological group. If the topology is generated by a complete separable metric, making G, as a set, into a complete separable metric space, otherwise known as a Polish space, then G is called a Polish group. A *topological isomorphism* of two topological groups G_1 and G_2 is a group isomorphism $\phi: G_1 \to G_2$ such that both ϕ and its inverse ϕ^{-1} are continuous.

Let G be a locally compact abelian group. A *character* of G is a function $\gamma: G \to \mathbb{C}$ such that $|\gamma(x)| = 1$ for every $x \in G$, and $\gamma(x + y) = \gamma(x)\gamma(y)$ for every $x, y \in G$ (we have switched to the additive notation, as is usual with abelian groups). The set Γ of all *continuous* characters of G equipped with the product of complex-

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4_10

valued functions is called the *dual group* of G. It is clearly an abelian group. The dual group of G is itself a locally compact abelian group when it is equipped with a sort of weak topology, called the *Gelfand topology*.

Example 10.1.1. 1. Let $G = \mathbb{R}^d$, $d \geq 1$. Then the continuous characters of G are of the form $\gamma(x) = e^{i(a,x)}$, $x \in \mathbb{R}^d$, for some $a \in \mathbb{R}^d$, and the dual group is topologically isomorphic to $\mathbb{R}^d$, with the usual topology (see Example (f), in Hewitt and Ross (1979, p. 368)).

2. Let $G = \mathbb{Z}^d$, $d \geq 1$. Then the continuous characters of G are of the form $\gamma(x) = \prod_{j=1}^{d} z_j^{x_j}$, $x = (x_1, \ldots, x_d) \in \mathbb{Z}^d$, for some complex numbers $z_1, \ldots, z_d$ of norm 1. The dual group of G is topologically isomorphic to the d-dimensional torus, which is, in turn, topologically isomorphic to $(-\pi, \pi]^d$, with the usual topology (see Example (b), p. 366, and Theorem 23.18 in Hewitt and Ross (1979)).

The following theorem is an important characterization of nonnegative definite functions on locally compact abelian groups; see Rudin (1962, p. 19).

Theorem 10.1.2. *Let G be a locally compact abelian group, with dual group Γ. Let $\varphi : G \to \mathbb{C}$ be a continuous function. Then φ is nonnegative definite if and only if there is a finite measure μ on Γ such that*

$$\varphi(x) = \int_{\Gamma} \gamma(x)\, \mu(d\gamma), \; x \in G\,.$$

Furthermore, this measure is unique.

10.2 Weak and Vague Convergence

This section presents certain basic facts related to the two important notions of convergence of measures on metric spaces: weak convergence and vague convergence. The classical introduction to weak convergence is in Billingsley (1999).

Let (S, ρ) be a metric space, and let $\mathcal{S}$ be the Borel σ-field on S.

Definition 10.2.1. A sequence (P_n) of probability measures on $(S, \mathcal{S})$ is said to converge weakly to a probability measure P on $(S, \mathcal{S})$ if for every bounded and continuous function f on S,

$$\lim_{n\to\infty} \int_S f(x)\, P_n(dx) = \int_S f(x)\, P(dx)\,. \tag{10.1}$$

The typical notation is

$$P_n \Rightarrow P \;\text{ or }\; P_n \xrightarrow{w} P\,.$$

A sequence (X_n) of S-valued random variables is sometimes said to converge weakly to an S-valued random variable X if the sequence of the laws on $(S, \mathcal{S})$ of the random variables (X_n) converges weakly to the law of X in the sense of Definition 10.2.1. One sometimes writes

$$X_n \Rightarrow X \text{ or } X_n \xrightarrow{w} X .$$

It is important to keep in mind that even when weak convergence is stated in the language of random variables, it is not necessary for the random variables to be defined on the same probability space.

If S is the real line equipped with its usual topology, then weak convergence $X_n \Rightarrow X$ is the same as the usual convergence in distribution

$$F_n(x) := P(X_n \le x) \to F(x) := P(X \le x) \text{ as } n \to \infty$$

for every $x \in \mathbb{R}$ at which the distribution function F is continuous. In the case of a general metric space S, weak convergence can also be characterized by the limiting behavior of certain sequences of probabilities.

The following theorem is a collection of such characterizations. It is usually called the *portmanteau* (i.e., containing several things in one package) *theorem*.

Theorem 10.2.2. *Each of the following three assertions is equivalent to the weak convergence $P_n \Rightarrow P$:*

(a) $\limsup_n P_n(F) \le P(F)$ *for every closed set F;*
(b) $\liminf_n P_n(G) \ge P(G)$ *for every open set G;*
(c) $\lim_n P_n(A) = P(A)$ *for every P-continuity set A.*

Recall that a subset $A \subset S$ is called a P-continuity set if P does not charge the boundary of A, that is, if $P(\partial A) = 0$. See Exercise 10.9.1.

An important tool for establishing weak convergence of probability measures is the notion of *relative compactness*. We say that a set H of probability measures on S is relatively compact if every sequence of probability measures in H has a weakly convergent subsequence. The following criterion is known as *Prokhorov's theorem*.

Theorem 10.2.3. (i) *Suppose that for every $\varepsilon > 0$, there is a compact set K_ε such that $P(K_\varepsilon) \ge 1 - \varepsilon$ for each $P \in H$ (this property of H is called tightness). Then the set H is relatively compact.*
(ii) *Suppose that the metric space S is complete and separable. If a set H of probability measures on S is relatively compact, then it is also tight.*

One can use weak convergence of a sequence of probability measures on one metric space to establish weak convergence of sequences of probability measures on other metric spaces via the following theorem, usually called the *continuous mapping theorem*.

Theorem 10.2.4. *Let S and S_1 be metric spaces, and $h : S \to S_1$ a Borel measurable map. Let (P_n), P be probability measures on the Borel σ-field in S.*

If $P_n \Rightarrow P$ in S and P does not charge the set of the discontinuities of h, then $P_n \circ h^{-1} \Rightarrow P \circ h^{-1}$ in S_1.

The set of discontinuities of h is always measurable (even if h is not); see Exercise 10.9.1.

Vague convergence of measures is a counterpart of weak convergence that is sometimes natural to use in dealing with measures whose total mass is not a priori bounded. It applies to sequences of Radon measures on a metric space S. For discussion of vague convergence, we assume that the metric space S is a complete separable metric space and in addition, that it is locally compact. That is, for every point $x \in S$, there is an $\varepsilon = \varepsilon(x) > 0$ such that the closed ball centered at x with radius $\varepsilon(x)$ is compact. It is not difficult to check that every such metric space is also σ-compact (i.e., is a union of countably many compact sets). Section 3.4 of Resnick (1987) can be consulted for more details on the discussion presented below.

Definition 10.2.5. A measure m on $(S, \mathcal{S})$ is said to be *Radon* if $m(K) < \infty$ for every compact set K.

In particular, every finite measure is Radon. Since the space S is σ-compact, every Radon measure is necessarily σ-finite.

Definition 10.2.6. A sequence (m_n) of Radon measures on $(S, \mathcal{S})$ is said to converge vaguely to a Radon measure m on $(S, \mathcal{S})$ if for every continuous function f on S with compact support,

$$\lim_{n\to\infty} \int_S f(x)\, m_n(dx) = \int_S f(x)\, m(dx)\,. \tag{10.2}$$

Note that a continuous function with compact support is automatically bounded. The common notation is

$$m_n \xrightarrow{v} m\,.$$

Let (P_n) be a sequence of *probability measures* (P_n) on a complete separable locally compact metric space S. It is obvious from the definitions that weak convergence $P_n \xrightarrow{w} P$ to some probability measure P implies also vague convergence $P_n \xrightarrow{v} P$. The converse statement also turns out to be true; see Exercise 10.9.2.

The following is a version of the portmanteau theorem corresponding to the vague convergence.

Theorem 10.2.7. *Each of the following two assertions is equivalent to the vague convergence $m_n \xrightarrow{v} m$:*

(a) $\limsup_n m_n(K) \le m(K)$ *for every compact set K and* $\liminf_n m_n(G) \ge m(G)$ *for every open relatively compact set G;*

(b) $\lim_n m_n(A) = m(A)$ *for every m-continuity set A relatively compact in S.*

Note that the two equivalent conditions (a) and (b) in the portmanteau theorem for weak convergence (Theorem 10.2.2) have become a single condition (i.e., both have to be assumed separately) in the portmanteau theorem for vague convergence (Theorem 10.2.7).

Relative compactness in the vague topology of a set of measures is defined analogously to its weak convergence counterpart. The following is a criterion for relative compactness for vague convergence.

Theorem 10.2.8. *Relative compactness of a set H of Radon measures on a complete separable locally compact metric space is equivalent to each of the following conditions:*

(i) *For each compact set $B \subset S$, $\sup_{m\in H} m(B) < \infty$.*
(ii) *For each nonnegative continuous function f on S with compact support, $\sup_{m\in H} \int f\, dm < \infty$.*

There is also a version of the continuous mapping theorem suitable for vague convergence.

Theorem 10.2.9. *Let S and S_1 be complete separable locally compact metric spaces, and $h : S \to S_1$ a Borel measurable map, such that*

$$h^{-1}(K_1) \text{ is relatively compact in } S \text{ whenever } K_1 \text{ is compact in } S_1.$$

Let (m_n), m be Radon measures on the Borel σ-field in S. Then $(m_n \circ h^{-1})$, $m \circ h^{-1}$ are Radon measures on the Borel σ-field in S_1. Further, if $m_n \xrightarrow{v} m$ in S and m does not charge the set of the discontinuities of h, then $m_n \circ h^{-1} \xrightarrow{v} m \circ h^{-1}$ in S_1.

Let $M_+^{\mathrm{R}}(S)$ be the space of all Radon measures on a locally compact complete separable metric space S. There is a metric d_v on $M_+^{\mathrm{R}}(S)$ under which this space is a complete separable metric space and is such that for measures $(m_n), m$ in $M_+^{\mathrm{R}}(S)$,

$$m_n \xrightarrow{v} m \text{ if and only if } d_v(m_n, m) \to 0\,. \tag{10.3}$$

The metric d_v can be chosen to be of the following form. One can find a countable family (h_i) of nonnegative continuous functions $S \to \mathbb{R}$ with compact support such that for $m_1, m_2 \in M_+^{\mathrm{R}}(S)$,

$$d_v(m_1, m_2) = \sum_{i=1}^{\infty} 2^{-i} \left(1 - \exp\left\{-\left|\int_S h_i(x)\, m_1(dx) - \int_S h_i(x)\, m_2(dx)\right|\right\}\right). \tag{10.4}$$

This is Proposition 3.17 in Resnick (1987), which can also be consulted for much of the discussion below.

A *random Radon measure* is a measurable map from some probability space into $M_+^{\mathrm{R}}(S)$ equipped with the Borel σ-field corresponding to any metric satisfying (10.3). The following proposition gives an alternative characterization of this

Borel σ-field; it is useful in checking measurability of certain mappings, such as certain infinitely divisible random measures of Section 3.2.

Proposition 10.2.10. *The Borel σ-field on $M_+^{\mathrm{R}}(S)$ coincides with the σ-field generated by the evaluation maps $M_+^{\mathrm{R}}(S) \to [0,\infty)$ given by $m \to m(K)$, K a compact subset of S.*

Proof. Denote the Borel σ-field on $M_+^{\mathrm{R}}(S)$ by $\mathcal{B}$, and the σ-field generated by the evaluation maps by $\mathcal{E}$. Let K be a compact set in S. By the portmanteau theorem for the vague convergence in Theorem 10.2.7, the mapping $m \to m(K)$ is upper semicontinuous on $M_+^{\mathrm{R}}(S)$, hence measurable with respect to the Borel σ-field on $M_+^{\mathrm{R}}(S)$. Therefore, $\mathcal{E} \subseteq \mathcal{B}$.

In the other direction, for every nonnegative continuous function h with compact support on S, the mapping $M_+^{\mathrm{R}}(S) \to [0,\infty)$, $m \to \int h\,dm$, is $\mathcal{E}$-measurable, as can be easily seen by approximating h by simple functions. Therefore, for a fixed $m_2 \in M_+^{\mathrm{R}}(S)$, the distance $d_v(m_1, m_2)$ in (10.4) is an $\mathcal{E}$-measurable function of $m_1 \in M_+^{\mathrm{R}}(S)$. Therefore, every open ball in $M_+^{\mathrm{R}}(S)$ is $\mathcal{E}$-measurable, and hence $\mathcal{B} \subseteq \mathcal{E}$. □

Example 10.2.11. Poisson random measure as a random Radon measure

We construct a version of a Poisson random measure of Example 3.2.5 that is a random Radon measure. Let m be a Radon measure on S. Let (S_i) be a partition of S into measurable relatively compact sets; then the restriction m_i of m to each set S_i is finite. Restricting ourselves to those i for which $m_i(S_i) > 0$, begin with a double array of independent S-valued random variables, $\big(X_j^{(i)}\big)$, such that for each i, the random variables in the ith row are i.i.d., with a common law $m_i/m_i(S_i)$. Let (K_i) be a sequence of independent Poisson random variables, independent of the double array, with $E(K_i) = m_i(S_i)$, $i \geq 1$. For every measurable set $B \subseteq S$, define

$$M(B) = \sum_i \sum_{j=1}^{K_i} \mathbf{1}\big(X_j^{(i)} \in B\big)\,. \tag{10.5}$$

Letting $\mathcal{S}_0$ be the collection of all measurable sets of finite measure m, it follows from Exercise 3.8.19 that $\big(M(B),\ B \in \mathcal{S}_0\big)$ is a Poisson random measure in the sense of Example 3.2.5.

On the other hand, (10.5) clearly defines, for every $\omega \in \Omega$, a σ-finite measure on S. In order to guarantee that this measure is actually Radon, let (c_i) be a countable dense subset of S, and let (r_i) be positive numbers such that each of the open balls $B_{c_i}(r_i)$ has a compact closure, and their union is S. By construction, for each i, $M\big(B_{c_i}(r_i)\big)$ is a Poisson random variable, hence a.s. finite. Let Ω_0 be the subset of Ω on which all these Poisson random variables are finite. Then $P(\Omega_0) = 1$, and we restrict the definition of M in (10.5). If K is a compact set, then there is a finite subcollection of the open balls $\big(B_{c_i}(r_i)\big)$ that covers it. Since each of these balls has finite M-measure, so does K. Therefore, this version of M is a Radon measure.

Finally, by definition, for each measurable set B with compact closure, $M(B)$ is a well-defined random variable. By Proposition 10.2.10, we conclude that M is a random Radon measure.

The following definition introduces in the realm of random Radon measures an object analogous to the Laplace transform of a nonnegative finite-dimensional random vector.

Definition 10.2.12. The Laplace functional of a random Radon measure M is

$$\Psi_M(f) = E \exp\left\{ - \int_S f(x)\, M(dx) \right\} ,$$

f a nonnegative continuous function $S \to \mathbb{R}$ with compact support.

Since M is Radon, it assigns a finite mass to the support of f, which is, in turn, automatically bounded. Therefore, the integral in the exponent is a.s. finite. By the definition of vague convergence, the integral is a continuous, hence Borel measurable, function of M, and so the Laplace functional of a random Radon measure is well defined. The Laplace functional of a Poisson random measure of Example 10.2.11 is calculated in Exercise 10.9.4.

Weak convergence of nonnegative finite-dimensional random vectors is equivalent to convergence of their Laplace transforms. Similarly, weak convergence of random Radon measures (i.e., weak convergence in the vague topology) turns out to be equivalent to convergence of their Laplace functionals.

Theorem 10.2.13. *Let (M_n), M be random Radon measures on a locally compact complete separable metric space S. Then $M_n \xrightarrow{w} M$ if and only if $\Psi_{M_n}(f) \to \Psi_M(f)$ for every nonnegative continuous function f on S with compact support.*

10.3 Signed Measures

This section contains a brief description of signed measures taking values in $[-\infty, \infty]$. Most texts prohibit one of the two infinite values (e.g., Billingsley (1995), Dudley (1989)).

Let $(S, \mathcal{S})$ be a measurable space. Recall that a signed measure is a σ-additive finite-valued set function on $\mathcal{S}$.

Let (B_n) be events in $\mathcal{S}$ such that $B_n \uparrow S$. Let (ν_n) be a sequence of signed measures on $(S, \mathcal{S})$ such that

- for every $n \geq 1$, the signed measure ν_n is supported on the event B_n, i.e., $\nu_n(A) = 0$ for each $A \in \mathcal{S}$ such that $A \cap B_n = \emptyset$;
- the sequence (ν_n) is consistent in the sense that for each $n \geq 1$, the restriction of ν_{n+1} to B_n coincides with ν_n.

The Hahn decomposition of signed measures (see, e.g., Theorem 32.1 in Billingsley (1995)) says that for each $n \geq 1$, there exist a measurable partition of $B_n = B_n^+ \cup B_n^-$ and finite measures m_n^+ and m_n^- supported by B_n^+ and B_n^- respectively such that $\nu_n = m_n^+ - m_n^-$. By consistency of the sequence (ν_n), we see that the two sequences of measures $\left(m_n^+\right)$ and $\left(m_n^-\right)$ are consistent as well. Therefore, there are σ-finite measures m^+ and m^- on $\left(S, \mathcal{S}\right)$ such that for each $n \geq 1$, m_n^+ is the restriction of m^+ to B_n, and m_n^- is the restriction of m^- to B_n. By construction, the measures m^+ and m^- are mutually singular (i.e., have disjoint supports) as well.

Let

$$\mathcal{S}_0 = \left\{B \in \mathcal{S} : \ \min\left(m^+(B), m^-(B)\right) < \infty.\right\}.$$

For every set $B \in \mathcal{S}_0$, we define

$$\nu(B) = m^+(B) - m^-(B)\,,$$

and we call ν a $[-\infty, \infty]$-valued σ-finite measure on $\left(S, \mathcal{S}\right)$ generated by the sequence (ν_n). We call the measure m^+ the positive part of ν, and the measure m^- the negative part of ν. The σ-finite measure $\|\nu\| = m_1 + m_2$ is the total variation measure of ν. Note that even though ν is not, in general, defined on the entire σ-field $\mathcal{S}$, its restriction to each event B_n is the signed measure ν_n, which is defined on the restriction of $\mathcal{S}$ to B_n.

Note that the specific collection of events $B_n \uparrow S$ used to define a $[-\infty, \infty]$-valued σ-finite measure is not important. In fact, given mutually singular σ-finite measures m^+ and m^-, any sequence $B_n \uparrow S$ such that $m^+(B_n) < \infty$, $m^-(B_n) < \infty$ leads to the same family $\mathcal{S}_0$ and the same set function ν on that family. In the sequel we will refer to ν simply as a signed measure, but it may be both σ-finite and $[-\infty, \infty]$-valued.

Clearly, the family $\mathcal{S}_0$ has the following property: $B \in \mathcal{S}_0$ and $A \in \mathcal{S}$, $A \subset B$ implies $A \in \mathcal{S}_0$, which allows us to make the following definition.

Definition 10.3.1. A set $B \in \mathcal{S}_0$ is called a null set of a signed measure ν if $\nu(A) = 0$ for each $A \in \mathcal{S}$, $A \subset B$.

Obviously, a set B is a null set of a signed measure if and only if both its positive and negative parts vanish on B.

If $f : S \to \mathbb{R}$ is a measurable function such that the integrals $\int f\, dm^+$ and $\int f\, dm^-$ exist, and

$$\max\left(\int_S f(s)\, m^+(ds), \int_S f(s)\, m^-(ds)\right) > -\infty,$$

$$\min\left(\int_S f(s)\, m^+(ds), \int_S f(s)\, m^-(ds)\right) < \infty,$$

then we say that the integral $\int f\,d\nu$ exists, and

$$\int_S f(s)\,\nu(ds) = \int_S f(s)\,m^+(ds) - \int_S f(s)\,m^-(ds)\,.$$

As usual, if $f : S \to \mathbb{R}$ is a measurable function and $B \in \mathcal{S}$, then we use the notation

$$\int_B f(s)\,\nu(ds) := \int_S f(s)\mathbf{1}\big(s \in B\big)\,\nu(ds)\,,$$

assuming that the integral on the right-hand side exists.

Let ν be a signed measure with positive part m^+ and negative part m^-, and let $f : S \to \mathbb{R}$ be a measurable function. We can define another signed measure as follows. Define σ-finite measures on $\big(S, \mathcal{S}\big)$ by

$$\begin{aligned}\hat{m}^+(B) &= \int_B f(s)_+\,m^+(ds) + \int_B f(s)_-\,m^-(ds),\\ \hat{m}^-(B) &= \int_B f(s)_-\,m^+(ds) + \int_B f(s)_+\,m^-(ds)\,,\end{aligned}$$

and notice that the two measures are mutually singular. Let $\hat{B}_n \uparrow S$ be any sequence of events such that $\hat{m}^+(B_n) < \infty$, $\hat{m}^-(B_n) < \infty$ for each $n \geq 1$, and construct a signed measure μ that is finite on each $\hat{B}_n$ and for which $\hat{m}^+$ is the positive part and $\hat{m}^-$ the negative part. If

$$\hat{\mathcal{S}}_0 = \Big\{B \in \mathcal{S} :\ \min\big(\hat{m}^+(B), \hat{m}^-(B)\big) < \infty.\Big\}$$

is the set of $B \in \mathcal{S}$ for which the signed measure μ is defined, then it is straightforward to check that

$$\hat{\mathcal{S}}_0 = \Big\{B \in \mathcal{S} :\ \int_B f(s)\,\nu(ds) \text{ is well defined}\Big\}$$

and for every $B \in \hat{\mathcal{S}}_0$,

$$\mu(B) = \int_B f(s)\,\nu(ds)\,. \tag{10.6}$$

If ν and μ are two signed measures connected via (10.6), we will say, by analogy with positive measures, that f is the derivative of μ with respect to ν. This terminology is justified by the fact that if both f_1 and f_2 satisfy the relation (10.6), then $f_1 = f_2$ a.e. with respect to the total variation measure $\|\nu\|$ (see Exercise 10.9.6). We write $d\mu/d\nu = f$.

By analogy with absolute continuity of positive measures, we introduce the following definition.

Definition 10.3.2. Let ν and μ be two signed measures on $(S, \mathcal{S})$. If every null set of ν is also a null set of μ, we say that the signed measure μ is absolutely continuous with respect to the signed measure ν and write $\mu \ll \nu$.

Clearly, absolute continuity $\mu \ll \nu$ of signed measures is equivalent to absolute continuity $\|\mu\| \ll \|\nu\|$ of the corresponding total variation measures.

If ν and μ are two signed measures satisfying (10.6), then it follows immediately from the above discussion that μ is absolutely continuous with respect to ν. A version of the Radon–Nikodym theorem also holds for signed measures.

Theorem 10.3.3. *Let ν and μ be two signed measures on $(S, \mathcal{S})$, and assume that $\mu \ll \nu$. Then there is a measurable function $f : S \to \mathbb{R}$ such that (10.6) holds.*

Proof. The assumptions of the theorem imply absolute continuity $\|\mu\| \ll \|\nu\|$ of the positive σ-finite measures, and therefore, by the Radon–Nikodym theorem for such measures (see, e.g., Theorem 32.2 in Billingsley (1995)), there exists a measurable function $g : S \to \mathbb{R}^+$ such that

$$\|\mu\|(B) = \int_B g(s)\, \|\nu\|(ds) \quad \text{for every } B \in \mathcal{F}. \tag{10.7}$$

Let m^+ and m^- be the positive and negative parts of the signed measure ν, and m_1^+ and m_1^- the positive and negative parts of the signed measure μ (so that $\|\nu\| = m^+ + m^-$ and $\|\mu\| = m_1^+ + m_1^-$). Let $S = S^+ \cup S^-$ be a measurable partition of S such that m^+ is supported by S^+ and m^- is supported by S^-, and let $S = S_1^+ \cup S_1^-$ be a similar partition of S into supports of the measures m_1^+ and m_1^-. Define a measurable function $f : S \to \mathbb{R}$ by

$$f(s) = \begin{cases} g(s) & \text{if } s \in S^+ \cap S_1^+ \text{ or } s \in S^- \cap S_1^-, \\ -g(s) & \text{if } s \in S^+ \cap S_1^- \text{ or } s \in S^- \cap S_1^+. \end{cases}$$

Let $\hat{\mu}$ be the signed measure such that $d\hat{\mu}/d\nu = f$. We will prove that $\hat{\mu} = \mu$.

Let $B_n \uparrow S$ be a sequence of measurable sets such that both $\|\nu\|(B_n) < \infty$ and $\|\mu\|(B_n) < \infty$ for every $n \geq 1$, and for $k \geq 1$, let $A_k = \{s : g(s) \leq k\}$. Take any increasing sequence (I_n) of finite collections of pairs (m, k), $m, k \geq 1$ such that $I_n \uparrow \mathbb{N}^2$, and define

$$\hat{B}_n = \bigcup_{(m,k) \in I_n} (B_m \cap A_k), \; n \geq 1\,.$$

Clearly, $\hat{B}_n \uparrow S$, and both signed measures μ and $\hat{\mu}$ are finite on each set $\hat{B}_n$. The claim of the theorem will follow once we check that μ and $\hat{\mu}$ coincide on the measurable subsets of $\hat{B}_n$ for each $n \geq 1$.

Let $B \subseteq \hat{B}_n$ be measurable. Using (10.7) and the definition of f, we have

$$\begin{aligned}
&\mu(B) = m_1^+(B) - m_1^-(B) \\
&= m_1^+\big(B \cap S_1^+\big) - m_1^-\big(B \cap S_1^-\big) = \|\mu\|\big(B \cap S_1^+\big) - \|\mu\|\big(B \cap S_1^-\big) \\
&= \int_{B\cap S_1^+} g(s)\, \|\nu\|(ds) - \int_{B\cap S_1^-} g(s)\, \|\nu\|(ds) \\
&= \int_{B\cap S_1^+\cap S^+} g(s)\, m^+(ds) + \int_{B\cap S_1^+\cap S^-} g(s)\, m^-(ds) \\
&\quad - \int_{B\cap S_1^-\cap S^+} g(s)\, m^+(ds) - \int_{B\cap S_1^-\cap S^-} g(s)\, m^-(ds) \\
&= \int_{B\cap S_1^+\cap S^+} f(s)\, m^+(ds) - \int_{B\cap S_1^+\cap S^-} f(s)\, m^-(ds) \\
&\quad + \int_{B\cap S_1^-\cap S^+} f(s)\, m^+(ds) - \int_{B\cap S_1^-\cap S^-} f(s)\, m^-(ds) \\
&= \int_{B\cap S^+} f(s)\, m^+(ds) - \int_{B\cap S^-} f(s)\, m^-(ds) = \hat{\mu}(B)\,,
\end{aligned}$$

as required. □

10.4 Occupation Measures and Local Times

Let $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ be a measurable stochastic process. For a Borel set $D \in \mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$, we define

$$\mu_{\mathbf{X}}(D) = \lambda_d\Big(\Big\{\mathbf{t} \in \mathbb{R}^d :\, \big(\mathbf{t}, X(\mathbf{t})\big) \in D\Big\}\Big)\,. \tag{10.8}$$

Clearly, $\mu_{\mathbf{X}}$ is a σ-finite measure on $\mathbb{R}^{d+1}$; it is the *occupation measure* of the stochastic process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$. For Borel sets $A \in \mathbb{R}^d$ and $B \in \mathbb{R}$, the value of $\mu_{\mathbf{X}}(A \times B)$ describes, informally, the amount of time in the set A the process spends in the set B.

Fix a "time set" $A \in \mathbb{R}^d$ of a finite positive Lebesgue measure and consider the measure on $\mathbb{R}$ defined by

$$\mu_{\mathbf{X},A}(B) = \mu_{\mathbf{X}}(A \times B),\ B \in \mathbb{R},\ \text{Borel}.$$

By the definition of the occupation measure, we have the following identity, valid for every measurable nonnegative function f on $\mathbb{R}$:

$$\int_A f\big(X(\mathbf{t})\big)\, \lambda_d(d\mathbf{t}) = \int_{\mathbb{R}} f(x)\, \mu_{\mathbf{X},A}(dx)\,. \tag{10.9}$$

If on an event of probability 1, $\mu_{\mathbf{X},A}(B)$ is absolutely continuous with respect to the Lebesgue measure λ, we say that the process has a *local time* over the set A. A local time is a version of the Radon–Nikodym derivative

$$l_{\mathbf{X},A}(x) = \frac{d\mu_{\mathbf{X},A}}{d\lambda}(x),\ x \in \mathbb{R}\,. \tag{10.10}$$

A local time is otherwise known as an *occupation density*. The basic property of the local time follows from (10.9): for every measurable nonnegative function f on $\mathbb{R}$,

$$\int_A f\big(X(\mathbf{t})\big)\,\lambda_d(d\mathbf{t}) = \int_{\mathbb{R}} f(x) l_{\mathbf{X},A}(x)\,dx\,. \tag{10.11}$$

If a process has a local time over a set A, the local time can also be computed by

$$l_{\mathbf{X},A}(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_A \mathbf{1}_{[x-\varepsilon,x+\varepsilon]}(X(\mathbf{t}))\,\lambda_d(d\mathbf{t})\,, \tag{10.12}$$

and the limit exists for almost every $x \in \mathbb{R}$. This useful representation of the local time has also an attractive intuitive meaning. It implies, in particular, that one can choose a version of a local time such that $l_{\mathbf{X},A}(x) = l_{\mathbf{X},A}(\omega; x)$ is a measurable function $\Omega \times \mathbb{R} \to \mathbb{R}$. In particular, for every $x \in \mathbb{R}$, $l_{\mathbf{X},A}(x)$ is a well-defined random variable.

An immediate conclusion from (10.12) is the following monotonicity property of the local times: if a process $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ has local times over sets A and B, then

$$A \subset B \text{ implies that } l_{\mathbf{X},A}(x) \le l_{\mathbf{X},B}(x) \text{ a.s.} \tag{10.13}$$

Let $\big(X(t),\ t \in \mathbb{R}\big)$ be a measurable stochastic process with a one-dimensional time. If the process has a local time over each interval $[0, t]$ in some range $t \in [0, T]$, then it is common to use the two-variable notation

$$l_{\mathbf{X}}(x, t) = l_{\mathbf{X},[0,t]}(x),\ 0 \le t \le T,\ x \in \mathbb{R}\,.$$

Using (10.12) shows that there is a version of $\big(l_{\mathbf{X}}(x, t)\big)$ that is jointly measurable in all three variables, ω, x, t.

As expected, existence and finite-dimensional distributions of a local time are determined by the finite-dimensional distributions of the underlying process.

Proposition 10.4.1. *(i) The finite-dimensional distributions of the local time are determined by the finite-dimensional distributions of the process. That is, let* $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ *and* $\big(Y(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ *be measurable stochastic processes with the same finite-dimensional distributions. If* $\big(X(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$ *has a local time over a set A, then so does the process* $\big(Y(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^d\big)$. *Moreover, there is a Borel set* $S \subset \mathbb{R}$ *of null Lebesgue measure such that the finite-dimensional distributions*

of $\big(l_{\mathbf{X},A}(x),\, x \in S^c\big)$ *coincide with the finite-dimensional distributions of* $\big(l_{\mathbf{Y},A}(x),\, x \in S^c\big)$.

(ii) If $\big(X(t),\, t \in \mathbb{R}\big)$ *and* $\big(Y(t),\, t \in \mathbb{R}\big)$ *are measurable stochastic processes with the same finite-dimensional distributions, and if* $\big(X(t),\, t \in \mathbb{R}\big)$ *has a local time over each interval* $[0,t]$ *in some range* $t \in [0,T]$*, then so does the process* $\big(Y(t),\, t \in \mathbb{R}\big)$*. Moreover, for every* $t_1,\dots,t_k$ *in* $[0,T]$*, there is a Borel set* $S \subset \mathbb{R}$ *of null Lebesgue measure such that the finite-dimensional distributions of* $\big(l_{\mathbf{X},A}(x,t_j),\, x \in S^c,\, j = 1,\dots,k\big)$ *coincide with the finite-dimensional distributions of* $\big(l_{\mathbf{Y},A}(x,t_j),\, x \in S^c,\, j = 1,\dots,k\big)$.

We first prove a useful lemma.

Lemma 10.4.2. *Let* $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ *and* $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ *be measurable stochastic processes with the same finite-dimensional distributions. Let* $A \subset \mathbb{R}^d$ *be a measurable set of finite positive Lebesgue measure, and* $f : A \to \mathbb{R}$ *a bounded measurable function. Then*

$$\int_A f\big(X(\mathbf{t})\big)\,\lambda_d(d\mathbf{t}) \stackrel{d}{=} \int_A f\big(Y(\mathbf{t})\big)\,\lambda_d(d\mathbf{t})\,.$$

Proof. Suppose that the process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is defined on some probability space $\big(\Omega_1, \mathcal{F}_1, P_1\big)$, while the process $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is defined on some other probability space $\big(\Omega_2, \mathcal{F}_2, P_2\big)$. Let $\mathbf{T}_1, \mathbf{T}_2, \dots$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ whose common law is the normalized Lebesgue measure on A, and suppose that the sequence is defined on yet another probability space $\big(\Omega_3, \mathcal{F}_3, P_3\big)$. Note that for every $\omega_1 \in \Omega_1$,

$$\frac{1}{n}\sum_{j=1}^{n} f\big(X(\mathbf{T}_j)\big) \to \int_A f\big(X(\mathbf{t})\big)\,\lambda_d(d\mathbf{t}) \tag{10.14}$$

as $n \to \infty$ P_3-a.s. by the law of large numbers. By Fubini's theorem, we see that on the product probability space $\big(\Omega_1 \times \Omega_3, \mathcal{F}_1 \times \mathcal{F}_3, P_1 \times P_3\big)$, there is an event $\Omega_3^{(1)} \in \mathcal{F}_3$ of full P_3-probability such that (10.14) holds P_1-a.s. for every $\omega_3 \in \Omega_3^{(1)}$.

Repeating the argument, we see that there is an event $\Omega_3^{(2)} \in \mathcal{F}_3$ of full P_3-probability such that for every $\omega_3 \in \Omega_3^{(2)}$,

$$\frac{1}{n}\sum_{j=1}^{n} f\big(Y(\mathbf{T}_j)\big) \to \int_A f\big(Y(\mathbf{t})\big)\,\lambda_d(d\mathbf{t}) \tag{10.15}$$

as $n \to \infty$ P_2-a.s. The event $\Omega_3^{(1)} \cap \Omega_3^{(2)}$ has full P_3-probability, so it must contain a point ω_3, which we fix. This gives us a fixed sequence $\big(\mathbf{T}_j\big)$, and for this sequence, the expressions on the left-hand sides of (10.14) and (10.15) have the same distributions. Since we have convergence in both (10.14) and (10.15), the claim of the lemma follows. □

Proof of Proposition 10.4.1. For part (i), let $A \in \mathbb{R}^d$ be a set of finite positive Lebesgue measure. It follows from Lemma 10.4.2 applied to indicator functions of Borel sets and linear combinations of such indicator functions that

$$\big(\mu_{\mathbf{X},A}(B),\ B \text{ Borel}\big) \stackrel{d}{=} \big(\mu_{\mathbf{Y},A}(B),\ B \text{ Borel}\big) \tag{10.16}$$

in the sense of equality of finite-dimensional distributions. Suppose that $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ has a local time over the set A. Then on an event of probability 1, the probability measure $\mu_{\mathbf{X},A}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. This implies that on that event, for every $m_1 \geq 1$ there is $m_2 \geq 1$ such that for all $k = 1, 2, \ldots$ and rational numbers $\tau_1 < \tau_1' < \tau_2 < \tau_2' < \ldots < \tau_k < \tau_k'$ with $\sum_{-=1}^{k}(\tau_i' - \tau_i) < 1/m_2$, we have

$$\sum_{i=1}^{k} \mu_{\mathbf{X},A}\big((\tau_i, \tau_i']\big) < 1/m_1\,.$$

Then (10.16) implies that the same is true for the probability measure $\mu_{\mathbf{Y},A}$, and so on an event of probability 1, the probability measure $\mu_{\mathbf{Y},A}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ as well; see, e.g., Royden (1968). This means that the process $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ has a local time over the set A.

Next, suppose that the process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is defined on some probability space $\big(\Omega_1, \mathcal{F}_1, P_1\big)$, while the process $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ is defined on some other probability space $\big(\Omega_2, \mathcal{F}_2, P_2\big)$, and let $\Omega_i^{(1)} \in \mathcal{F}_1$, $i = 1, 2$, be events of full probability on which $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ and $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ have local times over the set A. Let $S \subset \mathbb{R}$ be a Borel set of null Lebesgue measure such that for every $x \in S^c$, the relation (10.12) holds for P_1-almost every $\omega_1 \in \Omega_i^{(1)}$, and the version of (10.12) written for the process $\big(Y(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ holds for P_2-almost every $\omega_2 \in \Omega_i^{(2)}$. Then the fact that the finite-dimensional distributions of $\big(l_{\mathbf{X},A}(x),\, x \in S^c\big)$ coincide with the finite-dimensional distributions of $\big(l_{\mathbf{Y},A}(x),\, x \in S^c\big)$ follows from Lemma 10.4.2. This proves part (i) of the proposition.

For part (ii) of the proposition, the fact that the process $\big(Y(t),\, t \in \mathbb{R}\big)$ has a local time over each interval $[0, t]$ in the range $t \in [0, T]$ follows from part (i), while the equality of the finite-dimensional distributions follows from (10.12) in the same way as the corresponding statement in part (i). □

The next, and basic, property of the local time follows from its definition.

Proposition 10.4.3. *(i) Suppose that a process $\big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ has a local time over a set A. Let $B \subset \mathbb{R}$ be a Borel set. If $\Omega_0 \in \mathcal{F}$ is an event such that for every $\omega \in \Omega_0$, $X(t) \in B^c$ for each $t \in A$, then for every $\omega \in \Omega_0$, $l_{\mathbf{X},A}(x) = 0$ for almost every $x \in B$.*

(ii) Let $t > 0$ and suppose that the local time $l_{\mathbf{X}}(\cdot, t)$ of the process $\big(X(t),\, t \in \mathbb{R}\big)$ over the interval $[0, t]$ exists. Let $y > 0$. If $\Omega_0 \in \mathcal{F}$ is an event such that for

every $\omega \in \Omega_0$, $\sup_{s\in[0,t]} |X(s)| < y$, *then for every* $\omega \in \Omega_0$, $l_{\mathbf{X}}(x,t) = 0$ *for almost every* x *with* $|x| \geq y$.

Proof. Letting f be the indicator function of the set B and appealing to (10.11) proves the first statement of the proposition. The second statement follows from the first with $B = (-\infty, -y] \cup [y, \infty)$. □

When do local times exist? An easy-to-check criterion for existence of a local time is due to Berman (1969). It is based on the following classical result on characteristic functions of random vectors.

Lemma 10.4.4. *(i) Let* $\mathbf{X}$ *be a random vector, and let* $\varphi_{\mathbf{X}}(\boldsymbol{\theta}) = Ee^{i(\boldsymbol{\theta},\mathbf{X})}$, $\boldsymbol{\theta} \in \mathbb{R}^d$ *be its characteristic function. Then* $\mathbf{X}$ *has a square integrable density if and only if*

$$\int_{\mathbb{R}^d} |\varphi_{\mathbf{X}}(\boldsymbol{\theta})|^2 \, \lambda_d(d\boldsymbol{\theta}) < \infty \, .$$

(ii) If

$$\int_{\mathbb{R}^d} |\varphi_{\mathbf{X}}(\boldsymbol{\theta})| \, \lambda_d(d\boldsymbol{\theta}) < \infty \, ,$$

then $\mathbf{X}$ *has a bounded uniformly continuous density.*

Proof. The second part of the lemma appears in about every book on probability; see, e.g., Corollary 2, p. 149, in Laha and Rohatgi (1979)). The statement of the first part of the lemma is less common in the probabilistic literature, so we include a proof.

Suppose that $\varphi_{\mathbf{X}}$ is square integrable. The general theory of L^2 Fourier transforms tells us that the function

$$f(\mathbf{x}) = \lim_{h\uparrow\infty} \frac{1}{(2\pi)^{d/2}} \int_{\|\boldsymbol{\theta}\|\leq h} e^{i(\boldsymbol{\theta},\mathbf{x})} \varphi_{\mathbf{X}}(\boldsymbol{\theta}) \, \lambda_d(d\boldsymbol{\theta}), \ \mathbf{x} \in \mathbb{R}^d$$

exists in $L^2(\lambda_d)$, and moreover, the function

$$f_1(\mathbf{x}) = \int_0^{x_1} \dots \int_0^{x_d} f(y_1, \dots, y_d) \, dy_1 \, \dots \, dy_d, \ \mathbf{x} = (x_1, \dots, x_d) \in (0,\infty)^d$$

satisfies the relation

$$f_1(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}}(\boldsymbol{\theta}) \prod_{j=1}^{d} \frac{e^{-it_j\theta_j} - 1}{-it_j} \, \lambda_d(d\boldsymbol{\theta}), \ \mathbf{x} = (x_1, \dots, x_d) \in (0,\infty)^d \, ;$$

see Section VI.2 in Yosida (1965). On the other hand, by the inversion theorem for characteristic functions (see, e.g., Theorem 3.3.3 in Laha and Rohatgi (1979)) we

know that for every $\mathbf{x} = (x_1, \ldots, x_d) \in (0, \infty)^d$ such that each x_j is a continuity point of the marginal distribution of the jth component of $\mathbf{X}$,

$$\frac{1}{\pi^d} \int_{\mathbb{R}^d} \varphi_{\mathbf{X}}(\boldsymbol{\theta}) \prod_{j=1}^{d} \frac{e^{-it_j\theta_j} - 1}{-it_j} \lambda_d(d\boldsymbol{\theta}) = P\Big(\mathbf{X} \in \prod_{j=1}^{d} (0, x_j]^d\Big).$$

We conclude that

$$P\Big(\mathbf{X} \in \prod_{j=1}^{d} (0, x_j]^d\Big) = \Big(\frac{\pi}{2}\Big)^{d/2} \int_0^{x_1} \ldots \int_0^{x_d} f(y_1, \ldots, y_d)\, dy_1 \ldots dy_d$$

a.e. on $(0, \infty)^d$. This means that the function f is real and nonnegative a.e. on $(0, \infty)^d$, the law of $\mathbf{X}$ is absolutely continuous on this set, and its density is square integrable. Since this argument can be repeated with only notational changes for other quadrants of $\mathbb{R}^d$, this proves the "if" part of the lemma. The other direction is easy, since the usual Fourier transform of a function in $L^1(\lambda_d) \cap L^2(\lambda_d)$ is in $L^2(\lambda_d)$; see once again Section VI.2 in Yosida (1965). □

The following proposition, due to Berman (1969), is an easy consequence of the lemma.

Proposition 10.4.5. *Let* $\big(X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\big)$ *be a measurable stochastic process. Let* $A \in \mathbb{R}^d$ *be a measurable set of a finite d-dimensional Lebesgue measure. A sufficient condition for the process to have a local time over the set A satisfying*

$$\int_{\mathbb{R}} l_{\mathbf{X},A}(x)^2\, dx < \infty \text{ with probability 1}$$

is

$$\int_{\mathbb{R}} \int_A \int_A E e^{i\theta(X(t) - X(s))}\, \lambda_d(dt)\, \lambda_d(ds)\, d\theta < \infty. \tag{10.17}$$

A sufficient condition for the process to have a bounded and uniformly continuous local time over the set A is

$$\int_{\mathbb{R}} \left(\int_A \int_A E e^{i\theta(X(t) - X(s))}\, \lambda_d(dt)\, \lambda_d(ds) \right)^{1/2} d\theta < \infty. \tag{10.18}$$

Proof. For a fixed $\omega \in \Omega$, consider $X = X(t)$, $t \in A$ as a random variable on the probability space A with the Borel σ-field restricted to A, and the probability measure $Q = (\lambda_d(A))^{-1} \lambda_d$. Then the occupation measure $\mu_{\mathbf{X},A}(\cdot)$ is, up to a constant, the probability law of X, and the existence of a square integrable local time over the set A is equivalent to the existence of a square integrable density of the probability law of X. Since the characteristic function of X at a point $\theta \in \mathbb{R}$ is

given by

$$\frac{1}{\lambda_d(A)} \int_A e^{i\theta X(t)}\, \lambda_d(dt)\,,$$

by part (i) of Lemma 10.4.4, the existence of such a square integrable density is equivalent to the finiteness of the expression

$$\int_{\mathbb{R}} \left| \int_A e^{i\theta X(t)}\, \lambda_d(dt) \right|^2 d\theta\,.$$

The expectation of this expression coincides with the left-hand side of (10.17), and if the expectation is finite, then the expression itself is finite on a set of probability 1. This proves the first statement.

Similarly, the existence of a bounded and uniformly continuous local time over the set A is equivalent to the existence of a bounded and uniformly continuous density of the probability law of X above. By part (ii) of Lemma 10.4.4, the existence of such a density follows from the a.s. finiteness of the integral

$$\int_{\mathbb{R}} \left| \int_A e^{i\theta X(t)}\, \lambda_d(dt) \right| d\theta\,.$$

Taking the expectation and using the Cauchy–Schwartz inequality

$$E \left| \int_A e^{i\theta X(t)}\, \lambda_d(dt) \right| \le \left(E \left| \int_A e^{i\theta X(t)}\, \lambda_d(dt) \right|^2 \right)^{1/2}$$

proves the second statement. □

Even the simple tools of Proposition 10.4.5 already guarantee the existence and regularity of the local times of certain self-similar SαS processes with stationary increments; see Exercise 10.9.7. Stronger results have been obtained for certain Gaussian processes using the theory of local nondeterminism introduced in Berman (1973), and later extended to non-Gaussian stable processes in Nolan (1982). The following proposition shows the existence of jointly continuous local times for certain self-similar processes with stationary increments.

Proposition 10.4.6. *Let* $\big(X(t),\ t \in \mathbb{R}\big)$ *be a fractional Brownian motion, or the real harmonizable SαS motion with exponent of self-similarity* $0 < H < 1$*, or a linear fractional SαS motion with* $\alpha > 1$*,* $1/\alpha < H < 1$*, and* $c_2 = 0$*. Then the process has a local time over every interval* $[0, t]$*,* $t > 0$*, and moreover, there is a version of the local time that is jointly continuous in time and space. That is, there is a random field*

$$l_{\mathbf{X}}(x, t) = l_{\mathbf{X}}(x, t, \omega),\ 0 \le t \le T,\ x \in \mathbb{R},\ \omega \in \Omega$$

such that every $\omega \in \Omega$, $l_{\mathbf{X}}(x,t)$, is jointly continuous in $x \in \mathbb{R}$ and $t \geq 0$ (with $l_{\mathbf{X}}(x,0) = 0$ for all $x \in \mathbb{R}$), and for each $t > 0$, $l_{\mathbf{X}}(x,t)$, $x \in \mathbb{R}$ is a version of the local time $l_{\mathbf{X},[0,t]}(x)$, $x \in \mathbb{R}$.

Proof. For the fractional Brownian motion, the claim follows from Section 7 in Pitt (1978) and Theorem 8.1 in Berman (1973). For the real harmonizable SαS motion, the claim follows from Theorem 4.11 in Nolan (1989). For the linear fractional SαS motion, the claim follows from Ayache et al. (2008). □

Very precise estimates on the size of the time increments of the local time of the fractional Brownian motion are due to Xiao (1997). Some of them are summarized in the following proposition.

Proposition 10.4.7. *(i) Let $\left(l_{\mathbf{X}}(x,t),\ x \in \mathbb{R},\ t \geq 0\right)$ be the jointly continuous local time of a fractional Brownian motion with exponent $0 < H < 1$ of self-similarity. Then the supremum*

$$\sup_{\substack{x\in\mathbb{R}\\ 0\leq s<t\leq 1/2}} \frac{l(x,t)-l(x,s)}{(t-s)^{1-H}\left(\log\frac{1}{t-s}\right)^{H}}$$

is a.s. finite and has finite moments of all orders.

(ii) For every $t > 0$ and $p > 0$,

$$E\sup_{x\in\mathbb{R}} l(x,t)^p < \infty\,.$$

Proof. The finiteness of the supremum in the first part of the proposition follows from Corollary 1.1 in Xiao (1997). The finiteness of the moments is a very slight modification of the argument leading to the above corollary. The second part of the proposition follows from the first part by breaking the interval $[0,t]$ into parts of length less than $1/2$. □

It is, perhaps, not surprising that certain properties of a stochastic process, such as self-similarity, stationarity, and stationarity of the increments, are reflected in an appropriate way in the properties of the local time, assuming that the latter exists. In order to simplify the formulation of these relationships, we will assume that the local time is continuous.

Proposition 10.4.8. *Let $\left(X(t),\ t \in \mathbb{R}\right)$ be a measurable stochastic process, and assume that it has local time $\left(l_{\mathbf{X}}(x,t),\ x \in \mathbb{R},\ t \geq 0\right)$ that is jointly continuous in time and space.*

(i) If the process is self-similar with exponent H of self-similarity, then for every $c > 0$,

$$\left(l_{\mathbf{X}}(c^H x, ct),\ x \in \mathbb{R},\ t \geq 0\right) \stackrel{d}{=} \left(c^{1-H} l_{\mathbf{X}}(x,t),\ x \in \mathbb{R},\ t \geq 0\right). \tag{10.19}$$

(ii) If the process is stationary, then for every $h > 0$,

$$\big(l_{\mathbf{X}}(x, t+h) - l_{\mathbf{X}}(x, h),\, x \in \mathbb{R},\, t \geq 0\big) \stackrel{d}{=} \big(l_{\mathbf{X}}(x, t),\, x \in \mathbb{R},\, t \geq 0\big)\,. \tag{10.20}$$

(iii) Suppose that the process $\big(X(t),\, t \in \mathbb{R}\big)$ is defined on some probability space $\big(\Omega, \mathcal{F}, P\big)$. Suppose that the process has stationary increments and sample paths that are bounded on compact intervals satisfying

$$E \sup_{0 \leq t \leq T} |X(t)| < \infty\,.$$

Then for every $h > 0$, the infinite "law" of

$$\big(l_{\mathbf{X}}(x+u, t+h)(\omega) - l_{\mathbf{X}}(x+u, h)(\omega),\, x \in \mathbb{R},\, t \geq 0\big)$$

under the infinite measure $P \times \lambda$ does not depend on the shift h.

Proof. Note, first of all, that when the local times are continuous, the exceptional set S in Proposition 10.4.1 may be taken to be the empty set, and we will do that throughout this proof.

For part (i), let $c > 0$, and define a new stochastic process by

$$Y(t) = c^{-H} X(ct),\, t \in \mathbb{R}\,.$$

By self-similarity, the new process has the same finite-dimensional distributions as the original process $\big(X(t),\, t \in \mathbb{R}\big)$. Let f be a nonnegative measurable function on $\mathbb{R}$. For $t > 0$, we change the variable of integration twice, using in between (10.11) for $A = [0, ct]$, to write

$$\begin{aligned}
\int_0^t f(Y(s))\, ds &= \int_0^t f\big(c^{-H} X(cs)\big)\, ds \\
= c^{-1} \int_0^{ct} f\big(c^{-H} X(s)\big)\, ds &= c^{-1} \int_{\mathbb{R}} f\big(c^{-H} x\big)\, l_{\mathbf{X}}(x, ct)\, dx \\
&= c^{H-1} \int_{\mathbb{R}} f\big(x\big)\, l_{\mathbf{X}}(c^H x, ct)\, dx\,.
\end{aligned}$$

Therefore, $\big(c^{H-1} l_{\mathbf{X}}(c^H x, ct),\, x \in \mathbb{R},\, t \geq 0\big)$ is a version of the local time $\big(l_{\mathbf{Y}}(x, t),\, x \in \mathbb{R},\, t \geq 0\big)$. By Proposition 10.4.1, the latter has the same finite-dimensional distributions as the local time $\big(l_{\mathbf{X}}(x, t),\, x \in \mathbb{R},\, t \geq 0\big)$, and this proves (10.19).

In a similar manner, for part (ii) we take $h > 0$, define a new stochastic process by $Y(t) = X(t+h)$, $t \in \mathbb{R}$, and write for a nonnegative measurable function f and $t > 0$,

$$\int_0^t f(Y(s))\,ds = \int_0^t f\big(X(s+h)\big)\,ds$$
$$= \int_h^{t+h} f\big(X(s)\big)\,ds = \int_0^{t+h} f\big(X(s)\big)\,ds - \int_0^h f\big(X(s)\big)\,ds$$
$$= \int_{\mathbb{R}} f(x) l_{\mathbf{X}}(x, t+h)\,dx - \int_{\mathbb{R}} f(x) l_{\mathbf{X}}(x, h)\,dx$$
$$= \int_{\mathbb{R}} f(x)\big(l_{\mathbf{X}}(x, t+h) - l_{\mathbf{X}}(x, h)\big)\,dx\,,$$

so that

$$\big(l_{\mathbf{X}}(x, t+h) - l_{\mathbf{X}}(x, h),\ x \in \mathbb{R},\ t \geq 0\big)$$

is a version of the local time $\big(l_{\mathbf{Y}}(x,t),\ x \in \mathbb{R},\ t \geq 0\big)$. Now another appeal to Proposition 10.4.1 proves (10.20).

The proof of part (iii) of the proposition, which we now commence, has the same idea as the proof of part (ii), except that now we have to deal with infinite measures. Fix $h > 0$. Let f be a nonnegative measurable function. As before, there is an event of full probability such that on this event, for every $u \in \mathbb{R}$ and $t > 0$,

$$\int_0^t f\big(u + X(s+h)\big)\,ds = \int_{\mathbb{R}} f(x)\big(l_{\mathbf{X}}(x-u, t+h) - l_{\mathbf{X}}(x-u, h)\big)\,dx\,.$$

Applying this to a function $f = \mathbf{1}_{[x-\varepsilon, x+\varepsilon]}/2\varepsilon$ for $\varepsilon > 0$ and using the continuity of the local times gives us

$$l_{\mathbf{X}}(x-u, t+h) - l_{\mathbf{X}}(x-u, h) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbf{1}_{[x-\varepsilon, x+\varepsilon]}\big(u + X(s+h)\big)\,ds \qquad (10.21)$$

for every $u \in \mathbb{R}$, $t > 0$, and $x \in \mathbb{R}$.

Denote the expression on the left-hand side of (10.21) by $A_h(x, t; u, \omega)$ and the expression under the limit on the right-hand side of (10.21) by $A_{h,\varepsilon}(x, t; u, \omega)$. Choose pairs (x_j, t_j), $j = 1, \ldots, k$. Fix ω, and note that by Fubini's theorem, there is a measurable set $F \in (0, \infty)^k$ of full Lebesgue measure such that for all $(a_1, \ldots, a_k) \in F$, we have

$$\mathbf{1}\Big(A_{h,\varepsilon}(x_j, t_j; u, \omega) > a_j,\ j = 1, \ldots, k\Big) \to \mathbf{1}\Big(A_h(x_j, t_j; u, \omega) > a_j,\ j = 1, \ldots, k\Big)$$

for almost every $u \in \mathbb{R}$. Let now $M > |x_1| + 1 + \sup_{h \leq s \leq t_1 + h} |X(s)|$. We have, by the dominated convergence theorem,

$$\lambda\big\{u \in \mathbb{R} :\ A_{h,\varepsilon}(x_j, t_j; u, \omega) > a_j,\ j = 1, \ldots, k\big\}$$

$$= \lambda\{u \in [-M, M] : A_{h,\varepsilon}(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\}$$
$$\to \lambda\{u \in [-M, M] : A_h(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\}$$

for every $(a_1, \dots, a_k) \in F$.

Further, for every $\varepsilon > 0$,

$$\lambda\{u : A_{h,\varepsilon}(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\}$$
$$\le \lambda\{u : A_{h,\varepsilon}(x_1, t_1; u, \omega) > a_1\}$$
$$\le \frac{1}{2a_1\varepsilon} \int_{\mathbb{R}} \int_0^{t_1} \mathbf{1}_{[x_1-\varepsilon, x_1+\varepsilon]}\big(u + X(s+h)\big)\, ds\, du, = \frac{t_1}{a_1},$$

where at the last step we used Fubini's theorem. Finally, using once again the dominated convergence theorem, we conclude that

$$P \times \lambda\{(\omega, u) \in \Omega \times \mathbb{R} : A_{h,\varepsilon}(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\} \tag{10.22}$$
$$\to P \times \lambda\{(\omega, u) \in \Omega \times \mathbb{R} : A_h(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\}$$

for every $(a_1, \dots, a_k) \in F$. In particular, the "law" of $\big(A_h(x_j, t_j),\ j = 1, \dots, k\big)$ under $P \times \lambda$ is σ-finite on $(0, \infty)^k$.

Suppose that we show that for every $\varepsilon > 0$, the "law" of $\big(A_{h,\varepsilon}(x_j, t_j),\ j = 1, \dots, k\big)$ under $P \times \lambda$ is independent of $h > 0$. Then for every $h_1, h_2 > 0$, we can find a subset of $(0, \infty)^k$ of full Lebesgue measure such that (10.22) holds for h_1, h_2 and all $(a_1, \dots, a_k)$ in that set. This means that for such $(a_1, \dots, a_k)$,

$$P \times \lambda\{(\omega, u) \in \Omega \times \mathbb{R} : A_{h_1}(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\}$$
$$= P \times \lambda\{(\omega, u) \in \Omega \times \mathbb{R} : A_{h_2}(x_j, t_j; u, \omega) > a_j,\ j = 1, \dots, k\},$$

and hence the equality holds for all $a_1 > 0, \dots, a_k > 0$. This will establish the claim of part (iii) of the proposition.

It remains to show that for every $\varepsilon > 0$, the "law" of $\big(A_{h,\varepsilon}(x_j, t_j),\ j = 1, \dots, k\big)$ under $P \times \lambda$ is independent of $h > 0$. It is, of course, enough to consider there laws restricted to the "punctured" set $[0, \infty)^d \setminus \{0\}$. Assume without loss of generality that $t_1 < t_2 < \dots < t_d$, and notice that only those pairs (ω, u) for which

$$u + \inf_{0 \le s \le t_d} X(s+h) \le x + \varepsilon \text{ and } u + \sup_{0 \le s \le t_d} X(s+h) \ge x - \varepsilon$$

contribute to the values of $\big(A_{h,\varepsilon}(x_j, t_j),\ j = 1, \dots, k\big)$ in the set $[0, \infty)^d \setminus \{0\}$. Call this set $\mathcal{V}_h$. It follows from the assumption that the supremum of the process over compact intervals is integrable that the measure $P \times \lambda$ restricted to $\mathcal{V}_h$ is finite. Moreover, it follows from Proposition 1.1.11 that the total mass of this restricted measure is independent of $h > 0$. Normalizing this restricted measure

to be a probability measure, the required independence of $h > 0$ of the "law" of $\big(A_{h,\varepsilon}(x_j, t_j),\, j = 1, \ldots, k\big)$ follows from Proposition 1.1.11 and Lemma 10.4.2 applied to the linear combinations of $\big(A_{h,\varepsilon}(x_j, t_j),\, j = 1, \ldots, k\big)$. □

10.5 Karamata Theory for Regularly Varying Functions

This section contains a brief exposition of regularly varying functions of a single variable.

Definition 10.5.1. A measurable function $f : [0, \infty) \to \mathbb{R}$ is called regularly varying at infinity with exponent $\beta \in \mathbb{R}$ if f is either eventually positive or eventually negative, and for each $b > 0$,

$$\lim_{x \to \infty} \frac{f(bx)}{f(x)} = b^{\beta}\,. \tag{10.23}$$

It is not immediately obvious exactly what role the requirement of measurability plays in the definition of a regularly varying function. We will see that it provides a modicum of regularity necessary for developing Karamata theory, beginning with the uniform convergence statement in Proposition 10.5.5 below.

Remark 10.5.2. Our discussion in this section concerns only functions that are regularly varying at infinity. Regular variation at other points can be defined analogously, and much of the theory of functions regularly varying at infinity has obvious counterparts for functions that are regularly varying elsewhere. Such connections can be easily established, for example, by changing the variable. For example, a function f is regularly varying with exponent β as $x \downarrow 0$ if and only if the function $f(1/\cdot)$ is regularly varying at infinity with exponent $-\beta$.

The next statement is obvious.

Lemma 10.5.3. *Let f_1, f_2 be two positive regularly varying functions with exponents β_1 and β_2 respectively. Then for any two real numbers a_1, a_2, the function $g = f_1^{a_1} f_2^{a_2}$ is regularly varying with exponent $a_1\beta_1 + a_2\beta_2$.*

A regularly varying function L with exponent $\beta = 0$ is called slowly varying. An immediate consequence of Lemma 10.5.3 is the following representation of a regularly varying function.

Corollary 10.5.4. *Every regularly varying function f with exponent β can be represented in the form $f(x) = x^{\beta} L(x)$, $x > 0$, where L is a slowly varying function.*

The pointwise convergence assumption in the definition of regular variation plus the measurability assumption turns out to guarantee certain uniformity of this convergence. For slowly varying functions, this convergence is uniform on compact intervals in $(0, \infty)$, as shown in the following proposition.

Proposition 10.5.5. *Let L be a slowly varying function. Then for every $0 < a < b < \infty$ and $\delta > 0$, there is $x_0 \in (0, \infty)$ such that for all $x \geq x_0$,*

$$\left| \frac{L(cx)}{L(x)} - 1 \right| \leq \delta \tag{10.24}$$

for every $a \leq c \leq b$.

Proof. It is enough to consider the case in which L is eventually positive. Since the proposition describes the behavior of L for large values of its argument, we are free to change L arbitrarily on bounded sets, so we may assume that L is strictly positive on $[0, \infty)$.

Things are more transparent in the additive form as opposed to the multiplicative form, so we define

$$g(x) = \log L(e^x), \ x \in \mathbb{R}\,.$$

The function g is a measurable function, and the slow variation assumption on L translates into the following statement for g: for every $y \in \mathbb{R}$,

$$\lim_{x \to \infty} \big(g(x + y) - g(x)\big) \to 0 \tag{10.25}$$

as $x \to \infty$. It is clear that if convergence in (10.25) is uniform on compact y-intervals, then convergence in (10.24) is uniform on compact c-intervals in $(0, \infty)$. We will therefore prove that convergence in (10.25) is uniform on compact y-intervals. For this purpose, it is enough to check that this convergence is uniform for $y \in [0, 1]$. Indeed, if for every $\delta > 0$, there is $x_0 \in \mathbb{R}$ such that for all $x \geq x_0$,

$$\big|g(x + y) - g(x)\big| \leq \delta \text{ for each } y \in [0, 1], \tag{10.26}$$

then for every $k = 1, 2, \ldots$ and $y \in [0, 1]$,

$$\begin{aligned} &\big|g(x + y - k) - g(x)\big| \\ &\leq \big|g(x + y - k) - g(x - k)\big| + \sum_{j=1}^{k} \big|g(x - j) - g(x - j + 1)\big| \leq (k + 1)\delta \end{aligned}$$

for $x \geq x_0 + k$. That is, convergence in (10.25) is uniform on intervals of the form $[-k, -k + 1]$ for $k = 1, 2, \ldots$. We can similarly show uniform convergence on intervals of the form $[k, k + 1]$ for $k = 1, 2, \ldots$. Since every compact interval can be covered by a finite number of intervals of the above form, this will show that convergence in (10.25) is uniform on all compact y-intervals.

For the purpose of proving (10.26), let $\delta > 0$ and note that by (the pointwise version of) (10.25) and the bounded convergence theorem, for every $x \in \mathbb{R}$,

$$w(x) := \int_x^{x+2} \mathbf{1}\big(|g(t) - g(x)| > \delta/2\big)\, dt$$
$$= \int_0^2 \mathbf{1}\big(|g(t+x) - g(x)| > \delta/2\big)\, dt \to 0$$

as $x \to \infty$, so that there is $x_0 \in \mathbb{R}$ such that $w(x) < 1/2$ for all $x \geq x_0$. For such x and any $y \in [0,1]$,

$$\lambda\big([x, x+2] \cap [x+y, x+y+2]\big) = 2 - y \geq 1 > w(x) + w(x+y)$$
$$\geq \Big\{t \in [x, x+2] : |g(t) - g(x)| > \delta/2\Big\} \cup \Big\{t \in [x+y, x+y+2] : |g(t) - g(x+y)| > \delta/2\Big\}.$$

Therefore, the difference of the sets

$$\Big([x, x+2] \cap [x+y, x+y+2]\Big)$$

and

$$\Big\{t \in [x, x+2] : |g(t)-g(x)| > \delta/2\Big\} \cup \Big\{t \in [x+y, x+y+2] : |g(t)-g(x+y)| > \delta/2\Big\}$$

has positive Lebesgue measure, hence contains a point, say z. By definition, both $|g(z) - g(x)| \leq \delta/2$ and $|g(z) - g(x+y)| \leq \delta/2$, so that $|g(x) - g(x+y)| \leq \delta$. This proves (10.26). □

According to Corollary 10.5.4, a regularly varying function can be viewed as a power function "contaminated" by a slowly varying function. The next theorem says that regularly varying functions behave like power functions under integration.

Theorem 10.5.6. *Let f be a positive function regularly varying at infinity with exponent $\beta \geq -1$. Assume that f is locally integrable, i.e., $\int_0^a f(x)\, dx < \infty$ for every $0 < a < \infty$. Then the function $F(x) = \int_0^x f(t)\, dt$, $x \geq 0$, is regularly varying with exponent $\beta + 1$ and satisfies*

$$\lim_{x \to \infty} \frac{F(x)}{x f(x)} = \frac{1}{\beta + 1}, \tag{10.27}$$

with $1/0$ defined as $+\infty$.

Proof. Clearly, for $\beta > -1$, the regular variation of F follows from (10.27). We nonetheless begin by proving this regular variation. It will be used later in the proof of (10.27).

For $n = 1, 2, \ldots$, we set

$$a_n = \int_{2^{n-1}}^{2^n} f(t)\, dt\,.$$

Note that by the regular variation of f,

$$\frac{a_{n+1}}{a_n} = \frac{2\int_{2^{n-1}}^{2^n} f(2t)\, dt}{\int_{2^{n-1}}^{2^n} f(t)\, dt} \to 2^{\beta+1} \quad \text{as } n \to \infty.$$

Therefore, in the case $\beta > -1$, $\int_0^\infty f(t)\, dt = \infty$, while in the case $\beta = -1$, the integral may be either infinite or finite. In the latter case, if the integral is finite, $F(x)$ converges, as $x \to \infty$, to a finite positive constant and hence is trivially slowly varying. Suppose now that the integral is infinite (either with $\beta > -1$ or $\beta = -1$). Fix $b > 1$. For a given $0 < \varepsilon < 1$, there is $t_\varepsilon \in (0, \infty)$ such that for all $t > t_\varepsilon$, $f(bt)/f(t) \in [b^{\beta-\varepsilon}, b^{\beta+\varepsilon}]$. Therefore, for every $x > t_\varepsilon$,

$$\frac{\int_{bt_\varepsilon}^{bx} f(t)\, dt}{\int_{t_\varepsilon}^{x} f(t)\, dt} = \frac{b\int_{t_\varepsilon}^{x} f(bt)\, dt}{\int_{t_\varepsilon}^{x} f(t)\, dt} \in [b^{1+\beta-\varepsilon}, b^{1+\beta+\varepsilon}]\,,$$

implying that

$$b^{1+\beta-\varepsilon} \le \liminf_{x\to\infty} \frac{\int_{bt_\varepsilon}^{bx} f(t)\, dt}{\int_{t_\varepsilon}^{x} f(t)\, dt} \le \limsup_{x\to\infty} \frac{\int_{bt_\varepsilon}^{bx} f(t)\, dt}{\int_{t_\varepsilon}^{x} f(t)\, dt} \le b^{1+\beta+\varepsilon}\,.$$

Since $\int_0^\infty f(t)\, dt = \infty$, we conclude that

$$b^{1+\beta-\varepsilon} \le \liminf_{x\to\infty} \frac{F(bx)}{F(x)} \le \limsup_{x\to\infty} \frac{F(bx)}{F(x)} \le b^{1+\beta+\varepsilon}.$$

Since $\varepsilon > 0$ can be taken arbitrarily close to zero, we conclude that $F(bx)/F(x) \to b^{1+\beta}$ for every $b > 1$, which implies the regular variation of F.

In particular, the function

$$l(x) = \frac{xf(x)}{F(x)},\ x > 0\,,$$

is, by Lemma 10.5.3, slowly varying. By Fatou's lemma,

$$\limsup_{x\to\infty} l(x) \le \frac{1}{\int_0^1 \liminf_{x\to\infty}(f(tx)/f(x))\, dt} \le \frac{1}{\int_0^1 t^\beta\, dt} = 1 + \beta \tag{10.28}$$

(which actually already proves (10.27)). Note that the function F is absolutely continuous, with derivative f. Therefore, so is the function $h(x) = \log F(x)$, $x > 1$, with derivative

$$h'(x) = \frac{f(x)}{F(x)} = \frac{l(x)}{x}\,, \tag{10.29}$$

in the sense of absolute continuity. Hence

$$h(x) = h(1) + \int_1^x \frac{l(t)}{t}\,dt$$

for $x > 1$, and in the same range,

$$F(x) = F(1) \exp\left\{ \int_1^x \frac{l(t)}{t}\,dt \right\}\,. \tag{10.30}$$

By the regular variation of F,

$$2^{\beta+1} = \lim_{x\to\infty} \frac{F(2x)}{F(x)} = \exp\left\{ \lim_{x\to\infty} \int_x^{2x} \frac{l(t)}{t}\,dt \right\}\,.$$

Consequently,

$$\begin{aligned}(\beta+1)\log 2 &= \lim_{x\to\infty} \int_x^{2x} \frac{l(t)}{t}\,dt = \lim_{x\to\infty} \int_1^2 \frac{l(xt)}{t}\,dt \\ &= \lim_{x\to\infty} \left(l(x)\log 2 + \int_1^2 \frac{l(xt)-l(x)}{t}\,dt \right)\,.\end{aligned}$$

However, by (10.28), $(l(xt) - l(x))/t$ is uniformly bounded over x large enough and $1 \le t \le 2$. Furthermore, for every fixed $1 \le t \le 2$,

$$\frac{l(xt)-l(x)}{t} = \frac{l(xt)-l(x)}{l(x)} \frac{l(x)}{t} \to 0$$

as $x \to \infty$ by the slow variation of l and its eventual boundedness, as guaranteed by (10.28). By the bounded convergence theorem, we conclude that $l(x) \to \beta + 1$ as $x \to \infty$, which establishes (10.27). □

It turns out that while proving Theorem 10.5.6, we have (almost) established another very useful result, the Karamata representation of slowly varying functions.

Theorem 10.5.7 (Karamata representation). *For every slowly varying function L, there exist a measurable function $a : [0,\infty) \to \mathbb{R}$ such that $a(x) \to a \in \mathbb{R} \setminus \{0\}$ as $x \to \infty$ and a bounded measurable function $\varepsilon : [0,\infty) \to \mathbb{R}$ such that $\varepsilon(x) \to 0$ as $x \to \infty$ such that*

$$L(x) = a(x) \exp\left\{ \int_1^x \frac{\varepsilon(t)}{t}\, dt \right\}, \quad x \geq 1. \tag{10.31}$$

Proof. It is enough to prove the claim in the case that L is eventually positive, and by modifying the function a on a finite interval, we see that it is enough to prove the theorem in the case that L is strictly positive on $[0, \infty)$. By Proposition 10.5.5, the function L is bounded on every interval of the kind $[x, 2x]$ for $x \geq x_0$, for some $x_0 > 0$, and hence it is bounded on every compact interval with left endpoint x_0. By modifying, if necessary, the function L on $[0, x_0]$, we may assume that L is locally bounded, hence integrable.

We are therefore in the situation described in Theorem 10.5.6, with $\beta = 0$. Combining (10.29) with (10.30) (and writing $L(x)$ instead of $f(x)$) gives us

$$L(x) = \frac{l(x)}{x} F(1) \exp\left\{ \int_1^x \frac{l(t)}{t}\, dt \right\}.$$

Since $l(x) \to 0 + 1 = 1$ as $x \to \infty$, we obtain the Karamata representation (10.31) with $a(x) = l(x)F(1)$ and $\varepsilon(x) = l(x) - 1$, $x \geq 1$. □

An immediate corollary of the Karamata representation is the following result on the ratio of two values of a regularly varying function. It is known as the *Potter bounds*.

Corollary 10.5.8. *Let f be regularly varying at infinity with exponent β. For every $0 < \epsilon < 1$, there is $x_0 \in (0, \infty)$ such that for every $x \geq x_0$ and $b \geq 1$,*

$$(1 - \epsilon) b^{\beta - \epsilon} \leq \frac{f(bx)}{f(x)} \leq (1 + \epsilon) b^{\beta + \epsilon}. \tag{10.32}$$

Proof. By Corollary 10.5.4, it is enough to establish the bounds in the slowly varying case $\beta = 0$. The claim (10.32) follows from the Karamata representation (10.31) by choosing x_0 so large that for all $x \geq x_0$, we have both

$$\frac{1}{(1 + \epsilon)^{1/2}} \leq \frac{a(x)}{a} \leq (1 + \epsilon)^{1/2} \text{ and } |\varepsilon(x)| \leq \epsilon.$$

□

We can now add new angles to our previous comments that a regularly varying function is similar to a power function. First of all, it follows from the Potter bounds that an eventually positive regularly varying function with exponent β satisfies

$$\lim_{x \to \infty} \frac{f(x)}{x^{\beta - \epsilon}} = \infty \text{ and } \lim_{x \to \infty} \frac{f(x)}{x^{\beta + \epsilon}} = 0 \tag{10.33}$$

for every $\epsilon > 0$. Furthermore, the following counterpart of Theorem 10.5.6 shows that regularly varying functions integrate like power functions near infinity as well.

Theorem 10.5.9. *Let f be a positive function regularly varying at infinity with exponent $\beta \leq -1$. Then if $\beta < -1$, the function $G(x) = \int_x^\infty f(t)\,dt$ is eventually finite. If $\beta = -1$, assume that this function is eventually finite. Then G is regularly varying with exponent $\beta + 1$ and satisfies*

$$\lim_{x\to\infty} \frac{G(x)}{xf(x)} = \frac{-1}{\beta+1}\,, \tag{10.34}$$

with $-1/0$ defined as $+\infty$.

Proof. The eventual finiteness of the function G in the case $\beta < -1$ is, of course, clear from (10.33). The fact that G is regularly varying in the case $\beta < -1$ will follow from (10.34). Regardless of whether $\beta < -1$ or $\beta = -1$, as in Theorem 10.5.6, we can choose, for $b > 1$ and $0 < \varepsilon < 1$, a $t_\varepsilon \in (0,\infty)$ such that for all $t > t_\varepsilon, f(bt)/f(t) \in [b^{\beta-\varepsilon}, b^{\beta+\varepsilon}]$. Then for all $x > t_\varepsilon$,

$$\frac{G(bx)}{G(x)} = \frac{b\int_x^\infty f(bt)\,dt}{\int_x^\infty f(t)\,dt} \in [b^{1+\beta-\varepsilon}, b^{1+\beta+\varepsilon}]\,,$$

which shows that G is regularly varying with exponent $\beta + 1$.

In order to prove (10.34), consider first the case $\beta < -1$. As usual, we may assume that the function $x^{-\beta}f(x)$, $x > 0$, is locally integrable, so we can define a function $F_\beta(x) = \int_0^x t^{-\beta}f(t)\,dt$, $x \geq 0$. By Theorem 10.5.6,

$$F_\beta(x) \sim x^{1-\beta}f(x) \ \text{ as } x \to \infty. \tag{10.35}$$

Integrating by parts and taking (10.33) into account, we see that

$$G(x) = -x^\beta F_\beta(x) - \beta\int_x^\infty F_\beta(t)t^{\beta-1}\,dt\,. \tag{10.36}$$

According to (10.35),

$$\int_x^\infty F_\beta(t)t^{\beta-1}\,dt \sim G(x) \ \text{ as } x \to \infty,$$

so that by (10.36) and (10.35),

$$G(x) \sim \frac{-x^\beta F_\beta(x)}{1+\beta} \sim \frac{-xf(x)}{1+\beta}\,.$$

Finally, in the case $\beta = -1$, the statement (10.34) follows from Fatou's lemma:

$$\liminf_{x\to\infty} \frac{G(x)}{xf(x)} = \liminf_{x\to\infty} \int_1^\infty \frac{f(tx)}{f(x)}\, dt$$
$$\geq \int_1^\infty \liminf_{x\to\infty} \frac{f(tx)}{f(x)}\, dt = \int_1^\infty t^{-1}\, dt = \infty\,.$$

□

The following theorem complements the results in Theorems 10.5.6 and 10.5.9. It deals with two regularly varying functions in a convolution-like situation.

Theorem 10.5.10. *Let f and g be positive functions regularly varying at infinity with exponents α and β. Assume that f and g are locally bounded.*

(a) *Suppose that $\alpha > -1$ and $\beta > -1$. Then*

$$\lim_{x\to\infty} \frac{\int_0^x f(x-y)g(y)\, dy}{xf(x)g(x)} = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}\,. \tag{10.37}$$

(b) *Suppose that $\beta > -1$ and $\alpha + \beta < -1$. Then*

$$\lim_{x\to\infty} \frac{\int_0^\infty f(x+y)g(y)\, dy}{xf(x)g(x)} = \frac{\Gamma(-(\alpha+\beta+1))\Gamma(\beta+1)}{\Gamma(-\alpha)}\,. \tag{10.38}$$

Proof. For part (a), let $0 < \varepsilon < 1/2$, and write

$$\int_0^x f(x-y)g(y)\, dy = \int_0^{\varepsilon x} f(x-y)g(y)\, dy \tag{10.39}$$
$$+ \int_{\varepsilon x}^{(1-\varepsilon)x} f(x-y)g(y)\, dy + \int_{(1-\varepsilon)x}^{x} f(x-y)g(y)\, dy\,.$$

By the uniform convergence in Proposition 10.5.5,

$$\int_{\varepsilon x}^{(1-\varepsilon)x} f(x-y)g(y)\, dy \sim f(x)g(x) \int_{\varepsilon x}^{(1-\varepsilon)x} \Big(\frac{x-y}{x}\Big)^\alpha \Big(\frac{y}{x}\Big)^\beta\, dy$$
$$= xf(x)g(x) \int_\varepsilon^{1-\varepsilon} (1-z)^\alpha z^\beta\, dz$$

as $x \to \infty$. Letting $\varepsilon \to 0$ and noticing that

$$\int_0^1 (1-z)^\alpha z^\beta\, dz = B(\alpha+1, \beta+1) = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}\,,$$

we see that it remains to prove that the remaining two terms on the right-hand side of (10.39) are small. We use the Potter bounds of Corollary 10.5.8. Let $\delta > 0$ be so small that $\beta - \delta > -1$. With x_0 as in Corollary 10.5.8, for all x large enough,

$$\begin{aligned}
&\int_0^{\varepsilon x} f(x-y)g(y)\,dy \sim \int_{x_0}^{\varepsilon x} f(x-y)g(y)\,dy \\
&\le (1+\delta)^2 f(x)g(x) \int_{x_0}^{\varepsilon x} \left(\frac{x-y}{x}\right)^{\alpha-\delta} \left(\frac{y}{x}\right)^{\beta-\delta} dy \\
&\le (1+\delta)^2 x f(x)g(x) \int_0^{\varepsilon} (1-z)^{\alpha-\delta} z^{\beta-\delta}\,dz\,.
\end{aligned}$$

We conclude that

$$\lim_{\varepsilon\to 0} \limsup_{x\to\infty} \frac{\int_0^{\varepsilon x} f(x-y)g(y)\,dy}{xf(x)g(x)} = 0\,.$$

In a similar way, one can show that

$$\lim_{\varepsilon\to 0} \limsup_{x\to\infty} \frac{\int_{(1-\varepsilon)x}^{x} f(x-y)g(y)\,dy}{xf(x)g(x)} = 0\,.$$

For part (b), we proceed similarly. Let $0 < \varepsilon < 1$, and write

$$\begin{aligned}
\int_0^{\infty} f(x+y)g(y)\,dy &= \int_0^{\varepsilon x} f(x+y)g(y)\,dy \\
&+ \int_{\varepsilon x}^{x/\varepsilon} f(x+y)g(y)\,dy + \int_{x/\varepsilon}^{\infty} f(x+y)g(y)\,dy\,.
\end{aligned} \tag{10.40}$$

Now we use the fact that by Proposition 10.5.5, as $x \to \infty$,

$$\int_{\varepsilon x}^{x/\varepsilon} f(x+y)g(y)\,dy \sim xf(x)g(x) \int_{\varepsilon}^{1/\varepsilon} (1+z)^{\alpha} z^{\beta}\,dz\,,$$

and as $\varepsilon \to 0$,

$$\begin{aligned}
\int_{\varepsilon}^{1/\varepsilon} (1+z)^{\alpha} z^{\beta}\,dz &\to \int_0^{\infty} (1+z)^{\alpha} z^{\beta}\,dz \\
&= B\big(-(\alpha+\beta+1), \beta+1\big) = \frac{\Gamma(-(\alpha+\beta+1))\Gamma(\beta+1)}{\Gamma(-\alpha)}\,.
\end{aligned}$$

Furthermore, we can show in the same way as in part (i) that the remaining two terms on the right-hand side of (10.40) are appropriately small. □

Sometimes it is useful to talk about certain subclasses of slowly varying functions.

Definition 10.5.11. A measurable function $g : \mathbb{R}_+ \to \mathbb{R}_+$ that is eventually nonvanishing is said to be of the Zygmund class if for every $\delta > 0$, the function $\big(x^{\delta} g(x),\ x > 0\big)$ is eventually nondecreasing and the function $\big(x^{-\delta} g(x),\ x > 0\big)$ is eventually nonincreasing.

Definition 10.5.11 refers to nonnegative functions. We extend it automatically to nonpositive functions by saying that g is of the Zygmund class if $-g$ is of that class. Functions of the Zygmund class are automatically slowly varying; however, not every slowly varying function is of the Zygmund class; see Exercise 10.9.10. In fact, the following proposition shows that slowly varying functions of the Zygmund class have a special type of the Karamata representation.

Proposition 10.5.12. *Let g be a slowly varying function of the Zygmund class. Then there exist $a \in \mathbb{R} \setminus \{0\}$ and a bounded measurable function $\varepsilon : [0, \infty) \to \mathbb{R}$ such that $\varepsilon(x) \to 0$ as $x \to \infty$ such that for some $x_0 \in (0, \infty)$,*

$$g(x) = a \exp\left\{ \int_1^x \frac{\varepsilon(t)}{t}\, dt \right\}, \quad x \geq x_0. \tag{10.41}$$

Conversely, every function with a representation (10.41) is of the Zygmund class.

Proof. It is enough to prove that a function of the Zygmund class has a representation (10.41) when g is nonnegative, so we will consider that case. The function g has, as a slowly varying function, the general Karamata representation (10.31); let $\tilde{a}$ and $\tilde{\varepsilon}$ be the two functions appearing in that representation (we use the tilde notation because we are going to modify these functions to obtain the representation (10.41)). Then $\tilde{a}(x) \to a > 0$ as $x \to \infty$. Let $x_0 > 0$ be such that on $[x_0, \infty)$, $\tilde{a}(x) > 0$, the function $xg(x)$ is nondecreasing, and the function $x^{-1} g(x)$ is nonincreasing. Let $b(x) = \log\big(\tilde{a}(x)/a\big)$, $x \geq x_0$.

We begin by checking that the function b is absolutely continuous. Indeed, by monotonicity, for $x_2 > x_1 \geq x_0$,

$$\begin{aligned} b(x_2) - b(x_1) &= \log\big(x_2 \tilde{a}(x_2)\big) - \log\big(x_1 \tilde{a}(x_1)\big) + \big(\log x_1 - \log x_2\big) \\ &\geq \big(\log x_1 - \log x_2\big), \end{aligned}$$

and similarly,

$$\begin{aligned} b(x_2) - b(x_1) &= \log\big(x_2^{-1} \tilde{a}(x_2)\big) - \log\big(x_1^{-1} \tilde{a}(x_1)\big) + \big(\log x_1 - \log x_2\big) \\ &\leq \big(\log x_2 - \log x_1\big). \end{aligned}$$

This implies the absolute continuity of b. If b' is the derivative of b in the sense of absolute continuity, we already know that $|b'(x)| \le 1/x$ a.e. on $[x_0, \infty)$. Define

$$\varepsilon(t) = \begin{cases} \tilde{\varepsilon}(t) & \text{if } 1 \le t <\le x_0, \\ \tilde{\varepsilon}(t) + tb'(t) & \text{if } t \ge x_0. \end{cases}$$

Then the representation (10.41) is automatically satisfied, so we need to show only that we can modify the function ε on a set of measure zero to ensure the property $\varepsilon(x) \to 0$ as $x \to \infty$.

For each $0 < \delta < 1$ we can choose $x_\delta \ge x_0$ such that on $[x_\delta, \infty)$, the function $x^\delta g(x)$ is nondecreasing and the function $x^{-\delta} g(x)$ is nonincreasing. Repeating the above calculation, we see that for $x_2 > x_2 \ge x_\delta$, we have

$$\left|b(x_2) - b(x_1)\right| \le \delta\left(\log x_2 - \log x_1\right).$$

Therefore, $|b'(x)| \le \delta/x$ a.e. on $[x_\delta, \infty)$. This means that we can choose a version of the derivative b' such that for every $n \ge 1$, $|b'(x)| \le n^{-1}/x$ for all x large enough. Obviously, using this version of the derivative ensures the property $\varepsilon(x) \to 0$ as $x \to \infty$.

The fact that every function with a representation (10.41) is of the Zygmund class is easy, and is left as an exercise. □

Remark 10.5.13. Slowly varying functions for which there exists a Karamata representation of the type (10.41) are also called *normalized slowly varying*; see Bingham et al. (1987).

One can also talk about regularly varying sequences.

Definition 10.5.14. An eventually positive sequence (a_n) is called regularly varying with exponent $\beta \in \mathbb{R}$ if the function $f(x) = a_{\lceil x \rceil}$, $x > 0, f(0) = 0$, is regularly varying at infinity with exponent β.

Versions of the following lemma are used in the proof of a number of limit theorems in probability.

Lemma 10.5.15. *Let $h, g : (0, \infty) \to [0, \infty)$ be two nonincreasing functions, with g strictly positive in a neighborhood of the origin and $g(x) \to 0$ as $x \to \infty$. Assume that there is a sequence $a_n \uparrow \infty$ such that $nh(a_n x) \to g(x)$ as $n \to \infty$ for all continuity points $x > 0$ of g. Then there is $\beta > 0$ such that the sequence (a_n) is regularly varying with exponent β.*

Proof. Let $c > 0$. We begin by showing that the limit

$$\psi(c) := \lim_{n \to \infty} \frac{a_{\lceil nc \rceil}}{a_n} \tag{10.42}$$

exists. Indeed, suppose that to the contrary, the limit does not exist. Then there are two sequences $n_k \to \infty$ and $m_k \to \infty$ and numbers $0 < r_1 < r_2 < \infty$ such that

$$\frac{a_{\lceil n_k c \rceil}}{a_{n_k}} \le r_1 < r_2 \le \frac{a_{\lceil m_k c \rceil}}{a_{m_k}}$$

for all k. Let I be a neighborhood of the origin over which g is strictly positive, and let C be the set of discontinuity points of g. For every $x \notin C \cup C/r_1 \cup C/r_2$, we have

$$\begin{aligned} g(x) &= \lim_{k\to\infty} cn_k h\big(a_{\lceil n_k c \rceil} x\big) = \lim_{k\to\infty} cn_k h\Big(\big(a_{\lceil n_k c \rceil}/a_{n_k}\big) a_{n_k} x\Big) \\ &\ge c \limsup_{k\to\infty} n_k h\big(a_{n_k} r_1 x\big) = c g(r_1 x)\,. \end{aligned}$$

A similar argument with the sequence (m_k) gives us $g(x) \le c g(r_2 x)$, so that we have, by the monotonicity of g, $g(r_1 x) = g(r_2 x)$ for all $x \notin C \cup C/r_1 \cup C/r_2$. Note that the set

$$\bigcup_{j=1}^{\infty} \Big((r_2/r_1)^j C \cup r_1 (r_2/r_1)^j C \Big)$$

is at most countable, so there is $x \in I$ not in this set. For such x, we have

$$g\big((r_2/r_1)^j x\big) = g(x),\ j = 1, 2, \ldots\,.$$

Letting $j \to \infty$, we obtain $g(x) = 0$, which contradicts the fact that g takes positive values on I.

Therefore, the limit in (10.42) exists. Note that the function ψ is nondecreasing, $\psi(c) \ge 1$ if $c \ge 1$, and $\psi(c) \le 1$ if $c \le 1$. We claim that $0 < \psi(c) < \infty$ for all $c > 0$. Indeed, suppose that $\psi(c) = \infty$ for some $c > 1$. Then for every $M > 0$, we have $a_{\lceil nc \rceil}/a_n \ge M$ for all n large enough. For $x \notin C \cup C/M$, as above,

$$g(x) = \lim_{n\to\infty} cnh\big(a_{\lceil nc \rceil} x\big) \le c \liminf_{n\to\infty} nh\big(a_n M x\big) = c g(Mx)\,.$$

Choosing $x \in I$, $x \notin \cup_{n\in\mathbb{N}} C/n$, and letting $M \to \infty$ over the integers, we conclude $g(x) = 0$, which is, once again, a contradiction. Therefore, $\psi(c) < \infty$ for all $c > 0$, and a similar argument shows that $\psi(c) > 0$ for all $c > 0$. A similar argument also shows that ψ is continuous at $c = 1$. Indeed, suppose, for example, that $\psi(1+) = b > 1$, and let $b_1 = (1 + b)/2 > 1$. Let $x \notin C \cup C/b_1$. Then for all $c > 1$,

$$g(x) = \lim_{n\to\infty} cnh\big(a_{\lceil nc \rceil} x\big) \le c \liminf_{n\to\infty} nh\big(a_n b_1 x\big) = c g(b_1 x)\,.$$

Letting $c \downarrow 1$ and using the monotonicity of g, we see that $g(x) = g(b_1 x)$. Repeating the argument used to prove existence of the limit in (10.42) provides now the

necessary contradiction. Similarly impossible is the assumption $\psi(1-) < 1$. Hence we have continuity at $c = 1$.

Next, let $c > 0$. For $x \notin C \cup C/c$, we have

$$g(x) = \lim_{n\to\infty} cnh\big(a_{\lceil nc\rceil}x\big) \tag{10.43}$$

$$= c \lim_{n\to\infty} cnh\Big(\big(a_{\lceil n\rceil}/a_n\big)a_n x\Big) = cg\big(\psi(c)x\big)\,,$$

since $\psi(c)x$ is a continuity point of g. If now $c_1, c_2 > 0$, then we can use (10.43) with $c = c_1$, $c = c_2$, and $c = c_1c_2$ to obtain, for every $x \notin C \cup C/c_1 \cup C/c_2 \cup C/(c_1c_2)$,

$$c_1c_2g\big(\psi(c_1c_2)x\big) = g(x) = c_1g\big(\psi(c_1)x\big) = c_1c_2g\big(\psi(c_1)\psi(c_2)x\big)\,.$$

We have already seen that this implies that

$$\psi(c_1c_2) = \psi(c_1)\psi(c_2), \quad c_1 > 0,\ c_2 > 0\,. \tag{10.44}$$

Let $\beta = \log\big(\psi(e)\big)$. Then a repeated application of (10.44) shows first that the relation

$$\psi(c) = c^\beta \tag{10.45}$$

holds for c of the type $c = e^k$, k an integer, then for c of the type $c = e^q$, q rational, and finally, by monotonicity, for all $c > 0$. It follows now from (10.42) that

$$\lim_{x\to\infty} \frac{a_{\lceil xc\rceil}}{a_{\lceil x\rceil}} = c^\beta$$

for all $c > 0$, so the sequence (a_n) is regularly varying with exponent β.

Clearly, $\beta \geq 0$, and $\beta = 0$ is impossible, since in this case, $\psi \equiv 1$, contradicting (10.43). This completes the proof. □

We finish this section by mentioning several useful formulas that are frequently used in Fourier analysis of regularly varying functions. The first two can be found, for example, in Section 4.3 of Bingham et al. (1987):

$$\int_0^\infty \frac{\sin x}{x^\beta}\,dx = \frac{\pi/2}{\Gamma(\beta)\sin(\pi\beta/2)} \quad \text{for } 0 < \beta < 2 \tag{10.46}$$

(the integral is only conditionally convergent if $0 < \beta \leq 1$) and

$$\int_0^\infty \frac{\cos x}{x^\beta}\,dx = \frac{\pi/2}{\Gamma(\beta)\cos(\pi\beta/2)} \quad \text{for } 0 < \beta < 1, \tag{10.47}$$

with the integral, once again, converging only conditionally.

Integrating by parts, using (10.46), as well as the following property of the gamma function,

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)},\ z \in \mathbb{R} \setminus \mathbb{Z}, \tag{10.48}$$

we obtain the following formula:

$$\int_0^\infty \frac{1-\cos x}{x^\beta}\,dx = \begin{cases} \frac{\Gamma(3-\beta)}{(2-\beta)(\beta-1)}\sin(\pi\beta/2) & \text{for } 1<\beta<3,\ \beta\neq 2, \\ \pi/2 & \text{for } \beta = 2. \end{cases} \tag{10.49}$$

10.6 Multiple Integrals with Respect to Gaussian and SαS Measures

In this section, we present certain basic facts on multiple stochastic integrals. For details on multiple integration in the Gaussian case, see Nualart (1995); for the SαS case, see Rosiński et al. (1991).

We begin with the Gaussian case. Let M be a centered Gaussian random measure on $(S, \mathcal{S})$ with a σ-finite control measure m; see Example 3.2.4. Assume that the control measure m is atomless. Let $k \geq 1$. The k-tuple integral

$$I_k(f) = \int_S \dots \int_S f(s_1, \dots, s_k)\, M(ds_1) \dots M(ds_k)\,, \tag{10.50}$$

can be defined for each measurable function $f : S^k \to \mathbb{R}$ such that

$$\|f\|^2_{L^2(m^k)} = \int_S \dots \int_S f(s_1, \dots, s_k)^2\, m(ds_1) \dots m(ds_k) < \infty\,,$$

i.e., for $f \in L^2(S^k, m^k)$. If $k = 1$, the integral coincides with the usual integral with respect to a Gaussian random measure, as in Example 3.3.6, and $I_1(f)$ is a zero-mean Gaussian random variable. For a general $k \geq 1$, the basic properties of the integral are described in the following proposition.

Proposition 10.6.1. (i) *The integral is invariant under a permutation of the arguments of the function f: for every permutation $\sigma = (\sigma_1, \dots, \sigma_k)$ of $\{1, \dots, k\}$ and every $f \in L^2(S^k, m^k)$, one has $I_k(f_\sigma) = I_k(f)$, where $f_\sigma(t_1, \dots, t_k) = f(t_{\sigma_1}, \dots, t_{\sigma_k})$.*

(ii) *For every $f \in L^2(S^k, m^k)$, $I_k(f)$ is a random variable with zero mean and finite variance. If $f \in L^2(S^k, m^k)$ and $g \in L^2(S^l, m^l)$, then*

$$E(I_k(f) I_l(g)) = \mathbf{1}(k = l)\, k!(\tilde{f}, \tilde{g})_{L^2(m^k)}\,.$$

Here $\tilde{f}, \tilde{g}$ are the symmetrizations of the functions f, g. That is,

$$\tilde{f} = \frac{1}{k!} \sum_{\sigma} f_\sigma \,,$$

with the sum taken over all permutations σ of $\{1, \ldots, k\}$.

(iii) *The integral is linear: for every $f, g \in L^2(S^k, m^k)$ and real numbers a, b, one has $I_k(af + bg) = aI_k(f) + bI_k(g)$ a.s.*

(iv) *Let f be a simple function of the form*

$$f(t_1, \ldots, t_k) = a_{i_1,\ldots,i_k} \quad \text{if } t_1 \in A_{i_1}, \ldots, t_k \in A_{i_k}$$

for $i_1, \ldots, i_k = 1, \ldots, d$, $d = 1, 2, \ldots$, where $A_1, \ldots, A_d$ are disjoint sets of finite measure m, and $(a_{i_1,\ldots,i_k})$ is an array of real numbers that vanishes on the diagonals. That is, $a_{i_1,\ldots,i_k} = 0$ if any two of the numbers $i_1, \ldots, i_k$ coincide. The function f vanishes if its argument is not in the above range. Then

$$I_k(f) = \sum_{i_1,\ldots,i_k=1}^{d} a_{i_1,\ldots,i_k} M(A_{i_1}) \ldots M(A_{i_k}) \,.$$

Furthermore, simple functions of this type are dense in $L^2(S^k, m^k)$.

Note that it follows from the proposition that for $k \geq 2$, the integral $I_k(f)$ is no longer a Gaussian random variable. In fact, it can be viewed as a homogeneous polynomial of order k in the Gaussian random measure M.

The situation is similar in the SαS case, $0 < \alpha < 2$. Let now M be an SαS random measure on $(S, \mathcal{S})$ with a σ-finite modified control measure m; see Example 3.2.6. Once again, assume that the control measure m is atomless. For $k \geq 1$, we still use the notation (10.50) for the k-tuple integral of f with respect to M. The integral is now defined for every measurable function $f : S^k \to \mathbb{R}$ with the following property: there is a measurable function $\psi : S \to (0, \infty)$ with

$$\int_S \psi(s)^\alpha \, m(ds) < \infty$$

such that $N_\psi(f) < \infty$, where

$$N_\psi(f) = \int_S \ldots \int_S |f(s_1, \ldots, s_k)|^\alpha \left(1 + \log_+ \frac{|f(s_1, \ldots, s_k)|}{\psi(s_1) \ldots \psi(s_k)}\right)^{k-1} m(ds_1) \ldots m(ds_k) \,.$$

Here $\log_+ a = \max(0, \log a)$ for $a \geq 0$. We call the class of such functions f, $L^\alpha \log^{k-1} L_*$. It is easily seen to be a linear space. The subclass of $L^\alpha \log^{k-1} L_*$ for which $N_\psi(f) < \infty$ for a fixed ψ is also a linear space, which we call $L^\alpha \log^{k-1} L_\psi$.

One can introduce a notion of convergence in $L^\alpha \log^{k-1} L_\psi$ by saying that $f_m \to f$ if $N_\psi(f - f_n) \to 0$. Similar to Proposition 10.6.1, we have the following result.

Proposition 10.6.2. *The integral is invariant under a permutation of the arguments of the function f and is linear. Furthermore, for every simple function of the form of part (iv) of Proposition 10.6.1, we have*

$$I_k(f) = \sum_{i_1,\dots,i_k=1}^{d} a_{i_1,\dots,i_k} M(A_{i_1}) \dots M(A_{i_k}) .$$

Furthermore, simple functions of this type are dense in $L^\alpha \log^{k-1} L_\psi$ for a fixed ψ. If $N_\psi(f - f_n) \to 0$, then $I_k(f_n) \to I_k(f)$ in probability.

10.7 Inequalities, Random Series, and Sample Continuity

This last section is a collection of general inequalities and results on series of random variables and on sample continuity of stochastic processes used throughout the book. These results are available in many other texts. They are collected in one place here for easier reference.

The first inequality is a useful combination of Hölder's inequality and Fubini's theorem.

Theorem 10.7.1. *Let $(E_i, \mathcal{F}_i, m_i)$ be two σ-finite measure spaces, and let $f : E_1 \times E_2 \to [0, \infty)$ be a measurable function. Then for every $0 < q \le p < \infty$,*

$$\left(\int_{E_2} \left(\int_{E_1} f(x_1, x_2)^q \, m_1(dx_1) \right)^{p/q} m_2(dx_2) \right)^{1/p}$$
$$\le \left(\int_{E_1} \left(\int_{E_2} f(x_1, x_2)^p \, m_2(dx_2) \right)^{q/p} m_1(dx_1) \right)^{1/q} .$$

Proof. See Lemma 3.3.1 in Kwapień and Woyczyński (1992). □

The following two results deal with sums of independent random variables.

Theorem 10.7.2 (Marcinkiewicz-Zygmund inequalities). *For each $p \ge 1$, there exists a finite constant $B_p \ge 1$ such that for every $n = 1, 2, \dots$ and independent random variables $X_1, \dots, X_n$ such that $E|X_j|^p < \infty$ and $EX_j = 0$ for each $j = 1, \dots, n$,*

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

Proof. See Theorem 8.1 in Gut (2005). □

Theorem 10.7.3 (Lévy–Ottaviani inequalities). *Let $X_1, \ldots, X_n$ be independent random variables. Then for every $s, t \geq 0$,*

$$P\left(\max_{i=1,\ldots,n}\left|\sum_{j=1}^{i} X_j\right| > t + s\right) \leq \frac{P\left(\left|\sum_{j=1}^{n} X_j\right| > t\right)}{1 - \max_{i=1,\ldots,n} P\left(\left|\sum_{j=i}^{n} X_j\right| > s\right)}.$$

Proof. See Proposition 1.1.1 in Kwapień and Woyczyński (1992). □

A related inequality is the following maximal inequality for discrete-time martingales. Let $X_1, X_2, \ldots$ be a sequence of square integrable martingale differences with respect to some filtration $(\mathcal{F}_n,\ n = 0, 1, 2 \ldots)$; that is, X_n is $\mathcal{F}_n$-measurable and $E(X_n|\mathcal{F}_{n-1}) = 0$ a.s. for every $n = 1, 2, \ldots$.

Theorem 10.7.4. *For every $\lambda > 0$ and $n = 1, 2, \ldots$,*

$$\lambda P\left(\sup_{k=1,\ldots,n}\left|\sum_{j=1}^{k} X_j\right| > \lambda\right) \leq 2E\left(\sum_{j=1}^{n} X_j^2\right)^{1/2}.$$

Proof. See Theorem 5.6.1 in Kwapień and Woyczyński (1992). □

The result of the next theorem is sometimes referred to as a contraction principle for probabilities.

Theorem 10.7.5. *Let E be a normed space, and $X_1, \ldots, X_n$ independent symmetric random variables with values in E. Then for all real numbers $a_1, \ldots, a_n \in [-1, 1]$,*

$$P\left(\left\|\sum_{i=1}^{n} a_i X_i\right\| > t\right) \leq 2P\left(\left\|\sum_{i=1}^{n} X_i\right\| > t\right)$$

for every $t > 0$.

Proof. See Corollary 1.2.1 in Kwapień and Woyczyński (1992). □

Necessary and sufficient conditions for convergence of series of independent random variables are given in the following theorem, often called the *three series theorem*. The equivalence between a.s. and weak convergence in this case is known as the *Itô–Nisio theorem*.

Theorem 10.7.6. *Let (X_n) be a sequence of independent random variables. Then the series $\sum_{n=1}^{\infty} X_n$ converges a.s. (equivalently, in distribution) if and only if the following three series converge for some (equivalently every) $c > 0$:*

$$\sum_{n=1}^{\infty} P(|X_n| > c), \quad \sum_{n=1}^{\infty} E\big(X_n \mathbf{1}(|X_n| \leq c)\big), \quad \sum_{n=1}^{\infty} \mathrm{Var}\big(X_n \mathbf{1}(|X_n| \leq c)\big).$$

Proof. The equivalence between the modes of convergence (the Itô–Nisio part) can be found in Theorem 2.1.1 in Kwapień and Woyczyński (1992). The necessity and sufficiency of the three series criterion is in Theorem 22.8 in Billingsley (1995). □

The next theorem is usually referred to as the Kolmogorov continuity criterion.

Theorem 10.7.7. *Let* $\mathbf{X} = \big(X(\mathbf{t}),\, \mathbf{t} \in \mathbb{R}^d\big)$ *be a stochastic process. Suppose that for some* $a, b, C > 0$,

$$E|X(\mathbf{t}) - X(\mathbf{s})|^a \le C\|\mathbf{t} - \mathbf{s}\|^{d+b} \text{ for all } \mathbf{t}, \mathbf{s} \in \mathbb{R}^d. \tag{10.51}$$

Then the process $\mathbf{X}$ *has a continuous version, which is in addition Hölder continuous with Hölder exponent* γ *for every* $\gamma < b/a$.

Proof. See Theorem 3.23 in Kallenberg (2002).

We finish this section with two bounds from the Gaussian world. The next theorem says that the tail of the supremum of a bounded Gaussian process is not much heavier that the tail a single normal random variable whose variance equals the maximal variance of the Gaussian process. It is sometimes referred to as the Borell–TIS inequality.

Theorem 10.7.8. *Let* $\big(X(t),\, t \in T\big)$ *be a zero-mean Gaussian process on a countable set* T. *Assume that* $X^* := \sup_{t\in T} \|X(t)\| < \infty$ *a.s. Then* $\sigma^* := \sup_{t\in T}\big(\mathrm{Var}\big(X(t)\big)\big)^{1/2} < \infty$, *and for every* $x > 0$,

$$P\big(\big|X^* - m\big| > x\big) \le 2\Psi\big(x/\sigma^*\big)\,, \tag{10.52}$$

where m *is a median of* X^*, *and*

$$\Psi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,\ x \in \mathbb{R}$$

is the tail of the standard normal random variable. In particular, $EX^* < \infty$. *Furthermore,* (10.52) *remains true with the median* m *replaced by* EX^*.

Proof. The finiteness of σ^* is obvious. The version of (10.52) with the median is Theorem 3.1 in Borell (1975). The version with the expectation is in Theorem 2.1.1 in Adler and Taylor (2007). □

The following theorem, sometimes called the Slepian lemma, allows one to compare the distribution functions of two centered Gaussian random vectors with the same variances when one covariance matrix dominates the second covariance matrix pointwise.

Theorem 10.7.9. *Let* $(X_1, \ldots, X_n)$ *and* $(Y_1, \ldots, Y_n)$ *be centered Gaussian vectors. Assume that* $EX_i^2 = EY_i^2$ *for all* $i = 1, \ldots, n$ *and that* $E(X_iX_j) \le E(Y_iY_j)$ *for all*

$i, j = 1, \ldots, n$. *Then for all real numbers* $u_1, \ldots, u_n$,

$$P\big(X_1 \le u_1, \ldots, X_n \le u_n\big) \le P\big(Y_1 \le u_1, \ldots, Y_n \le u_n\big)\,.$$

Proof. See Slepian (1962). □

10.8 Comments on Chapter 10

Comments on Section 10.2

The classical text on weak convergence is Billingsley (1999). A very informative presentation of vague convergence and weak convergence in the vague topology is in Resnick (1987). Much of the second part of Section 10.2 is based on the latter text. In particular, Theorem 10.2.13 is Proposition 3.19 there.

Comments on Section 10.4

In the one-dimensional case, the statement of Lemma 10.4.4 is in Problem 11, p. 147, in Chung (1968).

The theory of local times was originally developed for Markov processes, beginning with Lévy (1939). Extending the idea of local times to non-Markov processes is due to Berman (1969), which considers mostly Gaussian processes. The existence of "nice" local times requires certain roughness of the sample paths of a stochastic process, and the powerful idea of local nondeterminism introduced in Berman (1973) can be viewed as exploiting this observation in the case of Gaussian processes. This approach was extended in Pitt (1978) to Gaussian random fields (with values in finite-dimensional Euclidian spaces), and to stable processes in Nolan (1982). Many details on local times of stochastic processes and random fields can be found in Geman and Horowitz (1980) and Kahane (1985).

Estimates similar to those in Proposition 10.4.7 (but with a slightly worse power of the logarithm) were also obtained in Csörgo et al. (1995).

Comments on Section 10.5

The modern theory of regular variation began with the paper Karamata (1930).

An encyclopedic treatment of regularly varying functions is in Bingham et al. (1987), following the earlier monograph Seneta (1976). A very readable exposition is in Section 0.4 in Resnick (1987).

The notion of functions of the Zygmund class was introduced in Bingham et al. (1987) with a reference to Zygmund (1968).

Comments on Section 10.6

Multiple integrals with respect to Gaussian measures were introduced in Wiener (1938). The definition used today is due to Itô (1951).

Multiple integrals with respect to SαS random measures have been introduced and studied in a series of papers, including Rosiński and Woyczyński (1986),

McConnell and Taqqu (1984), Kwapień and Woyczyński (1987), Krakowiak and Szulga (1988).

Comments on Section 10.7

The name "Borell–TIS" of the inequality in Theorem 10.7.8 is due to the fact that the version of (10.52) using the median of the supremum was proved at about the same time in Borell (1975) and Tsirelson et al. (1976).

10.9 Exercises to Chapter 10

Exercise 10.9.1. (i) *Let A be a subset of a metric space S. Show that the boundary ∂A is always a closed, hence Borel measurable, set.*

(ii) *Let S and S_1 be metric spaces, and $h : S \to S_1$ a map. Let*

$$D_h = \Big\{x \in S : \ h \text{ is not continuous at } x\Big\}$$

be the set of discontinuities of h. Show that D_h can be written as a countable union of closed sets and hence is Borel measurable.

Exercise 10.9.2. *Let (P_n), P be* probability measures *on a complete separable locally compact metric space S. Prove that if P_n converges vaguely to P, then P_n also converges to P weakly.*

Hint: you may find it useful to appeal to Prokhorov's theorem (Theorem 10.2.3) as well as to the following lemma, known as Urysohn's lemma (see Kallenberg (1983)).

Lemma 10.9.3. *Let S be a complete separable locally compact metric space.*

(i) *Let $K \subset S$ be a compact set. Then there exist a sequence of compact sets $K_n \downarrow K$ and a nonincreasing sequence f_n of continuous functions with compact support such that for every n,*

$$\mathbf{1}_K(x) \le f_n(x) \le \mathbf{1}_{K_n}(x), \ \ x \in S\,.$$

(ii) *Let $G \subset S$ be an open set relatively compact in S. Then there exist a sequence of open relatively compact sets $G_n \uparrow G$ and a nondecreasing sequence g_n of continuous functions with compact support such that for every n,*

$$\mathbf{1}_G(x) \ge g_n(x) \ge \mathbf{1}_{G_n}(x), \ \ x \in S\,.$$

Exercise 10.9.4. *Let M be a Poisson random measure on a locally compact complete separable metric space with Radon mean measure m. Prove that the Laplace functional of M is given by*

$$\Psi_M(f) = \exp\left\{-\int_S \big(1 - e^{-f(s)}\big)\, m(ds)\right\}\,,$$

f a nonnegative continuous function $S \to \mathbb{R}$ with compact support.

Exercise 10.9.5. *Let ν and μ be two signed measures on $(S, \mathcal{S})$. Construct the signed measure $\nu + \mu$ on $(S, \mathcal{S})$. Is it true that the positive and negative parts of the new measure are equal to the sums of the corresponding parts of the original measure?*

Exercise 10.9.6. *Prove that if ν and μ are two signed measures on $(S, \mathcal{S})$ and $f_i : S \to \mathbb{R}$, $i = 1, 2$, are measurable functions such that (10.6) holds with both $f = f_1$ and $f = f_2$, then $f_1 = f_2$ a.e. with respect to the total variation measure $\|\nu\|$.*

Exercise 10.9.7. *Let $(X(t),\ t \geq 0)$ be an H-self-similar SαS process with stationary increments, $0 < \alpha \leq 2$ (a fractional Brownian motion in the case $\alpha = 2$). Use Proposition 10.4.5 to show that over each compact interval the process has square integrable local time if $0 < H < 1$, and a bounded and uniformly continuous local time if $0 < H < 1/2$.*

Exercise 10.9.8. *Theorems 10.5.6 and 10.5.9 show that integrals of regularly varying functions are themselves regularly varying. This exercise explores to what extent the derivatives of regularly varying functions are themselves regularly varying.*

(i) Let $\beta \in \mathbb{R}$. Construct an example of a continuously differentiable function f that is regularly varying with exponent β such that the derivative f' is not regularly varying.

(ii) Let f be an absolutely continuous function with a derivative f' (in the sense of absolute continuity) that is eventually monotone. If f is regularly varying with exponent $\beta \neq 0$, show that f' is regularly varying with exponent $\beta - 1$.

Exercise 10.9.9. *Regularly varying functions can often be made nice! Let f be a positive function. Show that if f is regularly varying with exponent $\beta > 0$, we can find a monotone increasing absolutely continuous function g such that $f(x)/g(x) \to 1$ as $x \to \infty$. Show that the same can be achieved if f is regularly varying with exponent $\beta < 0$, except that now g is monotone decreasing.*

Exercise 10.9.10. *(i) Prove that every function of the Zygmund class is slowly varying.*

*(ii) Construct a slowly varying function that is not of the Zygmund class. **Hint:** Start with the Karamata representation of a slowly varying function and choose a function a that is not eventually continuous.*

Exercise 10.9.11. *Show by example that there are measurable functions g_1 and g_2 with $g_1(x)/g_2(x) \to 1$ as $x \to \infty$ such that g_1 is of the Zygmund class, while g_2 is not.*

Exercise 10.9.12. *Prove the converse part of Proposition 10.5.12.*

Bibliography

J. Aaronson, *An Introduction to Infinite Ergodic Theory*. Mathematical Surveys and Monographs, vol. 50 (American Mathematical Society, Providence, 1997)

R.J. Adler, J.E. Taylor, *Random Fields and Geometry* (Springer, New York, 2007)

L. Ahlfors, *Complex Analysis* (McGraw-Hill Book Co., New York, 1953)

A. Astrauskas, Limit theorems for quadratic forms of linear processes. Lith. Math. J. **23**, 355–361 (1983)

F. Avram, M.S. Taqqu, Noncentral limit theorems and Appell polynomials. Ann. Probab. **15**, 767–775 (1987)

F. Avram, M.S. Taqqu, Weak convergence of sums of moving averages in the α -stable domain of attraction. Ann. Probab. **20**, 483–503 (1992)

A. Ayache, F. Roueff, Y. Xiao, Joint continuity of the local times of linear fractional stable sheets. C. R. Math. Acad. Sci. Paris **344**, 635–640 (2008)

R. Baillie, T. Bollerslev, H. Mikkelsen, Fractionally integrated generalized autoregressive conditional heteroscedasticity. J. Econ. **74**, 3–30 (1996)

R. Balan, A. Jakubowski, S. Louhichi, Functional convergence of linear processes with heavy-tailed innovations. J. Theor. Probab. **27**, 1–36 (2014)

S. Banach, *Théorie des Opérations Linéaires* (PWN, Warsaw, 1955)

B. Basrak, R. Davis, T. Mikosch, A characterization of multivariate regular variation. Ann. Appl. Probab. **12**, 908–920 (2002)

J. Beran, Y. Feng, S. Ghosh, R. Kulik, *Long-Memory Processes. Probabilistic Properties and Statistical Methods* (Springer, Heidelberg, 2013)

S. Berman, Limit theorems for the maximum term in stationary sequences. Ann. Math. Stat. **35**, 502–516 (1964)

S. Berman, Local times and sample function properties of stationary Gaussian processes. Trans. Am. Math. Soc. **137**, 277–299 (1969)

S.M. Berman, Local nondeterminism and local times of Gaussian processes. Indiana Univ. Math. J. **23**, 69–94 (1973)

J. Bertoin, *Lévy Processes* (Cambridge University Press, Cambridge, 1996)

R. Bhattacharya, V. Gupta, E. Waymire, The Hurst effect under trends. J. Appl. Probab. **20**, 649–662 (1983)

P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968)

P. Billingsley, *Probability and Measure*, 3rd edn. (Wiley, New York, 1995)

P. Billingsley, *Convergence of Probability Measures*, 2nd edn. (Wiley, New York, 1999)

N. Bingham, Limit theorems for occupation times of Markov processes. Probab. Theory Relat. Fields **17**, 1–22 (1971)

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4

N. Bingham, C. Goldie, J.Teugels, *Regular Variation* (Cambridge University Press, Cambridge, 1987)
T. Bollerslev, Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**, 307–327 (1986)
C. Borell, The Brunn-Minkowski inequality in Gauss space. Invent. Math. **30**, 205–216 (1975)
R. Bradley, Equivalent mixing conditions for random fields. Ann. Probab. **21**, 1921–1926 (1993)
R. Bradley, Basic properties of strong mixing conditions. A survey and some open questions. Probab. Surv. **2**, 107–144 (2005)
P. Breuer, P. Major, Central limit theorems for non-linear functionals of Gaussian fields. J. Multivar. Anal. **13**, 425–441 (1983)
P. Brockwell, R. Davis, *Time Series: Theory and Methods* (Springer, New York, 1987)
P. Brockwell, R. Davis, *Time Series: Theory and Methods*, 2nd edn. (Springer, New York, 1991)
K. Chung, *A Course in Probability Theory* (Harcourt, Brace and World, Inc., New York, 1968)
D. Cline, Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. Thesis, Colorado State University (1983)
S. Cohen, G. Samorodnitsky, Random rewards, Fractional Brownian local times and stable self-similar processes. Ann. Appl. Probab. **16**, 1432–1461 (2006)
I. Cornfeld, S. Fomin, Y.G. Sinai, *Ergodic Theory* (Springer, New York, 1982)
N. Crato, P. Rothman, Fractional integration analysis of long-run behavior for US macroeconomic time series. Econ. Lett. **45**, 287–291 (1994)
M. Csörgo, Z.-Y. Lin, Q.-M. Shao, On moduli of continuity for local times of Gaussian processes. Stoch. Process. Appl. **58**, 1–21 (1995)
E. Damek, T. Mikosch, J. Rosiński, G. Samorodnitsky, General inverse problems for regular variation. J. Appl. Probab. **51A**, 229–248 (2014)
R. Davis, T. Hsing, Point processes for partial sum convergence for weakly dependent random variables with infinite variance. Ann. Probab. **23**, 879–917 (1995)
R. Davis, T. Mikosch, The sample autocorrelations of heavy–tailed stationary processes with applications to ARCH. Ann. Stat. **26**, 2049–2080 (1998)
R. Davis, S. Resnick, Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **13**, 179–195 (1985)
Y. Davydov, The invariance principle for stationary processes. Theor. Probab. Appl. **15**, 487–498 (1970)
L. deHaan, *On Regular Variation and its Application to the Weak Convergence of Sample Extremes* (Mathematisch Centrum, Amsterdam, 1970)
L. deHaan, A. Ferreira, *Extreme Value Theory: An Introduction* (Springer, New York, 2006)
L. Decreusefond, A. Üstünel, Stochastic analysis of the Fractional Brownian motion. Potential Anal. **10**, 177–214 (1999)
F. Diebold, A. Inoue, Long memory and regime switching. J. Econ. **105**, 131–159 (2001)
R. Dobrushin, P. Major, Non-central limit theorems for non-linear functions of Gaussian fields. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **50**, 27–52 (1979)
C. Dombry, N. Guillotin-Plantard, Discrete approximation of a stable self-similar stationary increments process. Bernoulli **15**, 195–222 (2009)
M. Donsker, An invariance principle for certain probability limit theorems. Mem. Amer. Math. Soc. **6** (1951)
R. Douc, F. Roueff, P. Soullier, On the existence of some ARCH(∞) processes. Stoch. Process. Appl. **118**, 755–761 (2008)
R. Dudley, *Real Analysis and Probability* (Wadsworth and Brook/Cole, Pacific Grove, 1989)
M. Dwass, Extremal processes. Ann. Math. Stat. **35**, 1718–1725 (1964)
P. Embrechts, C. Goldie, On closure and factorization properties of subexponential distributions. J. Aust. Math. Soc. A **29**, 243–256 (1980)
P. Embrechts, C. Goldie, N. Veraverbeke, Subexponentiality and infinite divisibility. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **49**, 335–347 (1979)
R. Engle, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica **50**, 987–1008 (1982)

W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd edn. (Wiley, New York, 1968)
W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd edn. (Wiley, New York, 1971)
T. Ferguson, M. Klass, A representation theorem of independent increment processes without Gaussian component. Ann. Math. Stat. **43**, 1634–1643 (1972)
R.A. Fisher, L. Tippett, Limiting forms of the frequency distributions of the largest or smallest member of a sample. Proc. Camb. Philis. Soc. **24**, 180–190 (1928)
G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd edn. (Wiley, New York, 1999)
D. Geman, J. Horowitz, Occupation densities. Ann. Probab. **8**, 1–67 (1980)
I. Gikhman, S. Skorohod, *Introduction to the Theory of Random Processes* (W.B. Saunders, London, 1996)
L. Gil-Alana, A fractionally integrated model for the Spanish real GDP. Econ. Bull. **3**, 1–6 (2004)
L. Giraitis, H. Koul, D. Surgailis, *Large Sample Inference for Long Memory Processes* (World Scientific, Singapore, 2012)
B. Gnedenko, Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. **44**, 423–453 (1943)
C. Granger, R. Joyeux, An introduction to long-memory time series and fractional differencing. J. Time Series Anal. **1**, 15–30 (1980)
A. Gut, *Probability: A Graduate Course* (Springer, New York, 2005)
E.J. Hannan, The central limit theorem for time series regression. Stoch. Process. Appl. **9**, 281–289 (1979)
T. Harris, H. Robbins, Ergodic theory of Markov chains admitting an infinite invariant measure. Proc. Natl. Acad. Sci. **39**, 860–864 (1953)
M. Hauser, R. Kunst, Fractionally integrated models with ARCH errors: with an application to the Swiss 1-month Euromarket interest rate. Rev. Quant. Finan. Acc. **10**, 95–113 (1998)
E. Hewitt, K. Ross, *Abstract Harmonic Analysis I* (Springer, New York, 1979)
C. Heyde, Y. Yang, On defining long range dependence. J. Appl. Probab. **34**, 939–944 (1997)
J. Hosking, Fractional differencing. Biometrika **68**, 165–176 (1981)
H. Hult, F. Lindskog, On Kesten's counterexample to the Cramér-Wold device for regular variation. Bernoulli **12**, 133–142 (2006)
H. Hult, G. Samorodnitsky, Tail probabilities for infinite series of regularly varying random vectors. Bernoulli **14**, 838–864 (2008)
H. Hurst, Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Eng. **116**, 770–808 (1951)
H. Hurst, Methods of using long-term storage in reservoirs. Proc. Inst. Civil Eng. **5**, 519–543 (1956)
K. Itô, Multiple Wiener integral. J. Math. Soc. Jpn. **3**, 157–169 (1951)
K. Itô, N. Nisio, On the oscillation functions of Gaussian processes. Math. Scand. **22**, 209–223 (1968)
M. Jacobsen, T. Mikosch, J. Rosiński, G. Samorodnitsky, Inverse problems for regular variation of linear filters, a cancellation property for σ-finite measures, and identification of stable laws. Ann. Appl. Probab. **19**, 210–242 (2008)
L. Jones, U. Krengel, On transformations without finite invariant measure. Adv. Math. **12**, 275–295 (1974)
P. Jung, T. Owada, G. Samorodnitsky, Functional central limit theorem for a class of negatively dependent heavy-tailed stationary infinitely divisible processes generated by conservative flows. Ann. Probab. (2016), to appear
Z. Kabluchko, S. Stoev, Stochastic integral representations and classification of sum- and max-infinitely divisible processes. Bernoulli **22**, 107–142 (2016)
J.P. Kahane, *Some Random Series of Functions*, 2nd edn. (Cambridge University Press, Cambridge, 1985)

O. Kallenberg, Series of random processes without discontinuities of the second kind. Ann. Probab. **2**, 729–737 (1974)
O. Kallenberg, *Random Measures*, 3rd edn. (Akademie-Verlag, Berlin, 1983)
O. Kallenberg, *Foundations of Modern Probability* (Springer, New York, 1997)
O. Kallenberg, *Foundations of Modern Probability*, 2nd edn. (Springer, New York, 2002)
J. Karamata, Sur un mode de croissance régulière des fonctions. Mathematica (Cluj) **4**, 38–53 (1930)
A. Kolmogorov, Wienersche Spiralen und einige andere interessante kurven in Hilbertschen raum. C.R. (Doklady) Acad. Sci. USSR (N.S.) **26**, 115–118 (1940)
A. Kolmogorov, Y. Rozanov, On a strong mixing condition for a stationary Gaussian process. Teoria Veroyatnostei i Primeneniya **5**, 222–227 (1960)
W. Krakowiak, J. Szulga, A multiple stochastic integral with respect to a strictly p-stable measure. Ann. Probab. **16**, 764–777 (1988)
U. Krengel, Classification of states for operators, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. II (University of California Press, Berkeley, 1967), pp. 415–429
U. Krengel, *Ergodic Theorems* (De Gruyter, Berlin, New York, 1985)
H. Künsch, Discrimination between monotonic trends and long-range dependence. J. Appl. Probab. **23**, 1025–1030 (1986)
S. Kwapień, W. Woyczyński, Double stochastic integrals, random quadratic forms and random series in Orlicz spaces. Ann. Probab. **15**, 1072–1096 (1987)
S. Kwapień, W. Woyczyński, *Random Series and Stochastic Integrals: Single and Multiple* (Birkhäuser, Boston, 1992)
A. Kyprianou, *Introductory Lectures on Fluctuations of Lévy Processes with Applications* (Springer, Berlin, 2006)
R. Laha, V. Rohatgi, *Probability Theory* (Wiley, New York, 1979)
R. Laha, V. Rohatgi, Operator self-similar stochastic processes in R_d. Stoch. Process. Appl. **12**, 73–84 (1982)
J. Lamperti, Semi-stable stochastic processes. Trans. Am. Math. Soc. **104**, 62–78 (1962)
J. Lamperti, On extreme order statistics. Ann. Math. Stat. **35**, 1726–1737 (1964)
R. LePage, M. Woodroofe, J. Zinn, Convergence of a stable distribution via order statistics. Ann. Prob. **9**, 624–632 (1981)
J. Leslie, On the non-closure under convolution of the subexponential family. J. Appl. Probab. **26**, 58–66 (1989)
P. Lévy, Sur certains processus stochastiques homogénes. Compositio Mathematica **7**, 283–339 (1939)
S. Louhichi, E. Rio, Functional convergence to stable Lévy motions for iterated random Lipschitz mappings. Electron. J. Probab. **16**, paper 89 (2011)
M. Maejima, A self-similar process with nowhere bounded sample paths. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **65**, 115–119 (1983)
B. Mandelbrot, Une classe de processus stochastiques homothetiques a soi; application a loi climatologique de H.E. Hurst. C. R. Acad. Sci. Paris **240**, 3274–3277 (1965)
B. Mandelbrot, *The Fractal Geometry of Nature* (W.H. Freeman and Co., San Francisco, 1983)
B. Mandelbrot, M. Taqqu, Robust R/S analysis of long-run serial correlation, in *Proceedings of the 42nd Session of the International Statistical Institute*, vol.48, Book 2 (Bulletin of the I.S.I., Manila, 1979), pp. 69–104
B. Mandelbrot, J. Van Ness, Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10**, 422–437 (1968)
B. Mandelbrot, J. Wallis, Noah, Joseph and operational hydrology. Water Resour. Res. **4**, 909–918 (1968)
T. McConnell, M. Taqqu, Double integration with respect to symmetric stable processes. Technical Report 618, Cornell University (1984)
F. Merlevéde, M. Peligrad, On the weak invariance principles for stationary sequences under projective criteria. J. Theor. Probab. **19**, 647–689 (2006)

F. Merlevéde, M. Peligrad, S. Utev, Recent advances in invariance principles for stationary sequences. Probab. Surv. **3**, 1–36 (2006)
T. Mikosch, G. Samorodnitsky, The supremum of a negative drift random walk with dependent heavy–tailed steps. Ann. Appl. Probab. **10**, 1025–1064 (2000)
T. Mikosch, C. Stărică, Change of structure in financial time series, long range dependence and the GARCH model. IWI–preprint 99-5-06 of the University of Groningen (2016)
T. Mikosch, C. Stărică, Is it long memory or just changing variance we see in financial returns? Bull. EU & US Inflation Microecon. Anal. **68**, 46–68 (2000a)
T. Mikosch, C. Stărică, Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Ann. Stat. **28**, 1427–1451 (2000b)
T. Mori, H. Oodaira, The law of the iterated logarithm for self-similar processes represented by multiple Wiener integrals. Probab. Theory Relat. Fields **71**, 367–391 (1986)
K. Neeb, On a theorem of S. Banach. J. Lie Theory **8**, 293–300 (1997)
J. Nolan, Local times for stable processes. Ph.D. Thesis, University of Virginia (1982)
J. Nolan, Local nondeterminism and local times for stable processes. Probab. Theory Relat. Fields **82**, 387–410 (1989)
D. Nualart, *The Mallavin Calculus and Related Topics* (Springer, New York, 1995)
G. O'Brien, W. Vervaat, Marginal distributions of self-similar processes with stationary increments. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **64**, 129–138 (1983)
T. Owada, G. Samorodnitsky, Functional Central Limit Theorem for heavy tailed stationary infinitely divisible processes generated by conservative flows. Ann. Probab. **43**, 240–285 (2015a)
T. Owada, G. Samorodnitsky, Maxima of long memory stationary symmetric α-stable processes, and self-similar processes with stationary max-increments. Bernnoulli **21**, 1575–1599 (2015b)
A. Pakes, On the tails of waiting-time distribution. J. Appl. Probab. **12**, 555–564 (1975)
W. Palma, *Long-Memory: Theory and Methods* (Wiley, Hoboken, 2007)
M. Peligrad, Maximum of partial sums and an invariance principle for a class of weak dependent random variables. Proc. Am. Math. Soc. **126**, 1181–1189 (1998)
M. Peligrad, S. Sang, Asymptotic properties of self-normalized linear processes with long memory. Econ. Theory **28**, 548–569 (2012)
V. Pipiras, M.S. Taqqu, The limit of a renewal reward process with heavy-tailed rewards is not a linear fractional stable motion. *Bernoulli* **6**, 607–614 (2000)
V. Pipiras, M.S. Taqqu, Decomposition of self-similar stable mixing moving averages. Probab. Theory Relat. Fields **123**, 412–452 (2002a)
V. Pipiras, M.S. Taqqu, The structure of self-similar stable mixing moving averages. Ann. Probab. **30**, 898–932 (2002b)
V. Pipiras, M.S. Taqqu, Dilated Fractional stable motions. J. Theor. Probab. **17**, 51–84 (2004a)
V. Pipiras, M.S. Taqqu, Stable stationary processes related to cyclic flows. Ann. Probability **32**, 2222–2260 (2004b)
V. Pipiras, M.S. Taqqu, Integral representations of periodic and cyclic fractional stable motions. Electron. J. Probab. **12**, 181–206 (2007)
V. Pipiras, M.S. Taqqu, Identification of periodic and cyclic fractional stable motions. Ann. Inst. Henri Poincaré **44**, 612–637 (2008)
E. Pitman, Subexponential distribution functions. J. Aust. Math. Soc. A **29**, 337–347 (1980)
L. Pitt, Local times for Gaussian vector fields. Indiana Univ. Math. J. **27**, 309–330 (1978)
B. Rajput, J. Rosiński, Spectral representations of infinitely divisible processes. Probab. Theory Relat. Fields **82**, 451–488 (1989)
S. Resnick, Weak convergence to extremal processes. Ann. Probab. **3**, 951–960 (1975)
S. Resnick, Point processes, regular variation and weak convergence. Adv. Appl. Probab. **18**, 66–138 (1986)
S. Resnick, *Extreme Values, Regular Variation and Point Processes* (Springer, New York, 1987)
S. Resnick, *Adventures in Stochastic Processes* (Birkhäuser, Boston, 1992)
S. Resnick, *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling* (Springer, New York, 2007)

S. Resnick, M. Rubinovitch, The structure of extremal processes. Adv. Appl. Probab. **5**, 287–307 (1973)
S. Resnick, G. Samorodnitsky, Point processes associated with stationary stable processes. Stoch. Process. Appl. **114**, 191–210 (2004)
S. Resnick, C. Stărică, Consistency of Hill estimator for dependent data. J. Appl. Probab. **32**, 139–167 (1995)
M. Rosenblatt, A central limit theorem and a strong mixing condition. Proc. Natl. Acad. Sci. **42**, 43–47 (1956)
M. Rosenblatt, Independence and dependence, in *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1961), pp. 411–443
M. Rosenblatt, *Markov Processes, Structure and Asymptotic Behaviour* (Springer, Berlin, 1971)
M. Rosenblatt, Fractional integrals of stationary processes and the central limit theorem. J. Appl. Probab. **13**, 723–732 (1976)
M. Rosenblatt, Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **49**, 125–132 (1979)
M. Rosenblatt, *Stationary Sequences and Random Fields* (Birkhauser, Boston, 1985)
J. Rosiński, On series representation of infinitely divisible random vectors. Ann. Probab. **18**, 405–430 (1990)
J. Rosiński, On the structure of stationary stable processes. Ann. Probab. **23**, 1163–1187 (1995)
J. Rosiński, Lévy and Related Jump-type Infinitely Divisible Processes. Lecture Notes, Cornell University (2007)
J. Rosiński, G. Samorodnitsky, Distributions of subadditive functionals of sample paths of infinitely divisible processes. Ann. Probab. **21**, 996–1014 (1993)
J. Rosiński, G. Samorodnitsky, Infinitely Divisible Stochastic Processes (2016). In preparation
J. Rosiński, G. Samorodnitsky, M. Taqqu, Sample path properties of stochastic processes represented as multiple stable integrals. J. Multivar. Analysis **37**, 115–134 (1991)
J. Rosiński, W. Woyczyński, On Itô stochastic integration with respect to p-stable motion: inner clock, integrability of sample paths, double and multiple integrals. Ann. Probab. **14**, 271–286 (1986)
J. Rosiński, T. Żak, Simple conditions for mixing of infinitely divisible processes. Stoch. Process. Appl. **61**, 277–288 (1996)
J. Rosiński, T. Żak, The equivalence of ergodicity and weak mixing for infinitely divisible processes. J. Theor. Probab. **10**, 73–86 (1997)
E. Roy, Ergodic properties of Poissonian ID processes. Ann. Probab. **35**, 551–576 (2007)
H. Royden, *Real Analysis*, 2nd edn. (Macmillan, London, 1968)
W. Rudin, *Fourier Analysis on Groups* (Interscience Publishers, New York, 1962)
G. Samorodnitsky, Extreme value theory, ergodic theory, and the boundary between short memory and long memory for stationary stable processes. Ann. Probab. **32**, 1438–1468 (2004)
G. Samorodnitsky, Null flows, positive flows and the structure of stationary symmetric stable processes. Ann. Probab. **33**, 1782–1803 (2005)
G. Samorodnitsky, M. Taqqu, *Stable Non-Gaussian Random Processes* (Chapman and Hall, New York, 1994)
K. Sato, *Lévy Processes and Infinitely Divisible Distributions* (Cambridge University Press, Cambridge, 1999)
E. Seneta, *Regularly Varying Functions*. Lecture Notes in Mathematics, vol. 508 (Springer, Berlin, 1976)
A. Skorohod, Limit theorems for stochastic processes with independent increments. Theor. Probab. Appl. **2**, 137–171 (1957)
D. Slepian, The one-sided barrier problem for Gaussian noise. Bell Syst. Tech. J. **41**, 463–501 (1962)
J. Smit, Solution to problem130. Stat. Neerlandica **37**, 87 (1983)

D. Surgailis, Convergence of sums of nonlinear functions of moving averages to selfsimilar processes. Sov. Math. Dokl. **23**, 247–250 (1981a)
D. Surgailis, On L^2 and non-L^2 multiple stochastic integration, in *Stochastic Differential Systems*. Lecture Notes in Control and Information Sciences, vol. 36 (Springer, New York, 1981b), pp. 212–226
M. Taqqu, Weak convergence to fractional Brownian motion and to the Rosenblatt process. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **31**, 287–302 (1975)
M. Taqqu, A representation for self-similar processes. Stoch. Process. Appl. **7**, 55–64 (1978)
M. Taqqu, Convergence of integrated processes of arbitrary Hermite rank. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **50**, 53–83 (1979)
J. Teugels, The class of subexponential distributions. Ann. Probab. **3**, 1000–1011 (1975)
E. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, 3rd edn. (Chelsea, New York, 1986)
B. Tsirelson, I. Ibragimov, V. Sudakov, Norms of Gaussian sample functions, in *Proceedings of the 3d Japan-USSR Symposium on Probability Theory*. Lecture Notes in Mathematics, vol. 550 (Springer, Berlin, 1976), pp. 20–41
W. Vervaat, On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. Adv. Appl. Probab. **11**, 750–783 (1979)
W. Vervaat, Sample paths of self-similar processes with stationary increments. Ann. Probab. **13**, 1–27 (1985)
P. Walters *An Introduction to Ergodic Theory* (Springer, New York, 1982)
I. Weissman, On location and scale functions of a class of limiting processes with application to extreme value theory. Ann. Probab. **3**, 178–181 (1975)
W. Whitt, *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Applications to Queues* (Springer, New York, 2002)
N. Wiener, The homogeneous chaos. Am. J. Math. **60**, 897–936 (1938)
Y. Xiao, Hölder conditions for the local times and the Hausdorff measure of the level sets of Gaussian random fields. Probab. Theory Relat. Fields **109**, 129–157 (1997)
A. Yaglom, Correlation theory of processes with stationary random increments of order n. Mat. Sb. **37**, 141–196 (1955); English translation in Am. Math. Soc. Translations Ser. 2 **8**, 87–141 (1958)
K. Yosida, *Functional Analysis* (Springer, Berlin, 1965)
P. Zaffaroni, Stationarity and memory of ARCH(∞) models. Econ. Theory **20**, 147–160 (2004)
A. Zygmund, *Trigonometric Series*, vols. I, II (Cambridge University Press, Cambridge, 1968)

Index

© Springer International Publishing Switzerland 2016

G. Samorodnitsky, *Stochastic Processes and Long Range Dependence*,
Springer Series in Operations Research and Financial Engineering,
DOI 10.1007/978-3-319-45575-4