Chapter 4
Limit Theorems

4.1 Tools

4.1.1 Introduction

Most statistical procedures in time series analysis (and in fact statistical inference in general) are based on asymptotic results. Limit theorems are therefore a fundamental part of statistical inference. Here we first review very briefly a few of the basic principles and results needed for deriving limit theorems in the context of long-memory and related processes.

4.1.2 How to Derive Limit Theorems?

To prove the convergence of an appropriately normalized process $S_n(\cdot)$, one has to verify the convergence of finite-dimensional distributions and tightness. With respect to the first issue, we usually prove just one-dimensional convergence because in most situations extensions to the multivariate case are straightforward. The tools we describe here are applicable to many statistics, not only partial sums. On the other hand, most of the statistics we will consider are just partial sums.

4.1.2.1 How to Verify Finite-Dimensional Convergence?

Suppose that X_t $(t \in \mathbb{N})$ is a stationary process. One of the common methods for deriving limit theorems is to evaluate its characteristic function. This is however rarely successful in a long-memory setting. An alternative method for partial sums of long-memory sequences is to study the asymptotic behaviour of cumulants. Recall that for a given random variable X, its cumulants are the coefficients in the power series

J. Beran et al., *Long-Memory Processes*,
DOI 10.1007/978-3-642-35512-7_4,

expansion of $\kappa_X(z) = \log E(e^{zX})$, i.e. $\kappa_j = \kappa_j(X)$ in

$$\kappa_X(z) = \sum_{j=0}^{\infty} \frac{z^j}{j!} \kappa_j.$$

In particular, $\kappa_1 = \mu_X = E(X)$, $\kappa_2 = \sigma_X^2 = \mathrm{var}(X)$. If $E(X) = 0$, then $\kappa_4 = E(X^4) - 3E^2(X^2)$. One of the useful properties of cumulants is that for a normal random variable X, we have $\kappa_j = 0$ for all $j \geq 3$, and this is only the case for the normal distribution. Moreover, a normal distribution is uniquely determined by its moments.

The justification for the approach based on cumulants is the following well-known result (see e.g. Rao 1965):

Theorem 4.1 *Let S_n $(n \in \mathbb{N})$ be a sequence of random variables such that $E[|S_n|^j] < \infty$ for all j, and let Y be a random variable whose distribution is uniquely determined by its moments $\mu_j = E(Y^j)$ $(j \in \mathbb{N})$. Then the convergence of all cumulants $\kappa_j(S_n)$ of S_n $(j \in \mathbb{N})$ to the cumulants $\kappa_j(Y)$ of Y implies that S_n converges to Y in distribution.*

Cumulants are useful if all moments exist. An approach that does not require finiteness of higher-order moments is referred to as a K-dependent approximation method and is adapted from Billingsley (1968, Theorem 4.2).

Proposition 4.1 *Let X_t $(t \in \mathbb{N})$ be a stationary sequence, c_n a sequence of constants, and $X_{t,K}$ $(t \in \mathbb{N})$ a sequence of K-dependent random variables. Define $S_n = \sum_{t=1}^{n} X_t$ and $S_{n,K} = \sum_{t=1}^{n} X_{t,K}$, and suppose that the following holds*:

(a) $c_n^{-1} S_{n,K} \overset{d}{\to} S_K$ *as* $n \to \infty$;

(b) $S_K \overset{P}{\to} S$ *as* $K \to \infty$;

(c)

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\left(c_n^{-1} |S_{n,K} - S_n| > \gamma\right) = 0$$

for each $\gamma > 0$.

Then, as $n \to \infty$,

$$c_n^{-1} S_n \overset{d}{\to} S.$$

To apply this theorem, we mention that if $v_K^2 \to v^2$ as $K \to \infty$, then $N(0, v_K^2) \overset{d}{\to} N(0, v^2)$. Furthermore, this approach requires the following result for K-dependent sequences.

Lemma 4.1 *Let $X_{t,K}$ $(t \in \mathbb{N})$ be a stationary sequence of K-dependent random variables with* $\mathrm{var}(X_{0,K}) < \infty$, *and define $S_{n,K} = \sum_{t=1}^{n} X_{t,K}$. Then*

$$n^{-\frac{1}{2}} S_{n,K} \xrightarrow{d} \sigma_K N(0,1),$$

where $\sigma_K^2 = \mathrm{var}(X_{0,K}) + 2\sum_{j=1}^{K} cov(X_{0,K}, X_{j,K})$.

Another useful result is the following martingale central limit theorem.

Lemma 4.2 *Let $(X_{t,n}, \mathcal{F}_t)$ $(t \in \mathbb{N}, n \geq 1)$ be a martingale difference array, and define $\tilde{X}_{t,n} = X_{t,n} - E(X_{t,n}|\mathcal{F}_{t-1})$. Furthermore, assume that the following conditions hold*:

(a) *for each $\delta > 0$,*

$$\sum_{t=1}^{n} E\big(\tilde{X}_{t,n}^2 1\big\{|\tilde{X}_{t,n}| > \delta\big\}\big) \to 0,$$

(b)

$$\sum_{t=1}^{n} E\big(\tilde{X}_{t,n}^2 \big|\mathcal{F}_{t-1}\big) \xrightarrow{p} 1.$$

Then

$$\sum_{t=1}^{n} X_{t,n} \xrightarrow{d} N(0,1).$$

4.1.2.2 How to Verify Tightness?

There are several ways to prove tightness. A particularly useful result given in Theorem 15.6 of Billingsley (1968) provides sufficient conditions for tightness in D (the space of right-continuous functions with left limits):

Lemma 4.3 *A stochastic process $Y_n(u)$ $(u \in [0,1])$ is tight if there exist $\eta > 1$, $a > 0$ and a nondecreasing function g such that for all $v_1 < u < v_2 \in [0,1]$,*

$$E\big[\big|Y_n(v_2) - Y_n(u)\big|^a \big|Y_n(u) - Y_n(v_1)\big|^a\big] \leq \big(g(v_2) - g(v_1)\big)^\eta.$$

In particular, assume that X_t $(t \in \mathbb{N})$ is a stationary sequence of random variables and G is a function such that $E[G(X_t)] = 0$. Consider the partial sum process

$$S_n(u) = \sum_{t=1}^{[nu]} G(X_t) \quad \big(u \in [0,1]\big). \tag{4.1}$$

Applying Lemma 4.3 to the partial sum process $Y_n(u) = d_n^{-1} S_n(u)$ yields the following result (see Theorem 2.1 in Taqqu 1975).

Lemma 4.4 *Assume that*

(a) $E[G(X_1)] = 0$ *and* $E[G^2(X_1)] < \infty$.
(b) $d_n^2 \sim n^{2d+1} L_S(n)$ *with* $-\frac{1}{2} \le d < \frac{1}{2}$ *and a slowly varying function* L_S.
(c) $E[S_n^2(1)] = O(d_n^2)$.
(d) *There exists* $a > (2d+1)^{-1}$ *such that* $E(|S_n(1)|^{2a}) = O((E[S_n^2(1)])^a)$.

Then $d_n^{-1} S_n(\cdot)$ *is tight.*

Proof Assume for simplicity that $L_S \equiv 1$. We note that the process $S_n(u)$, $u \in [0, 1]$, has stationary increments. In particular, for $0 \le u \le v \le 1$, $S_n(v) - S_n(u) \overset{d}{=} S_n(v-u)$. Thus, applying the Cauchy–Schwarz inequality and stationarity of increments, we have for $v_1 < u < v_2$, and a suitable constant $0 < C < \infty$,

$$\begin{aligned}
&d_n^{-2a} E\big[\big|S_n(v_2) - S_n(u)\big|^a \big|S_n(u) - S_n(v_1)\big|^a\big] \\
&\quad \le d_n^{-2a} \big(E\big[\big|S_n(v_2-u)\big|^{2a}\big]\big)^{1/2} \big(E\big[\big|S_n(u-v_1)\big|^{2a}\big]\big)^{1/2} \\
&\quad \le d_n^{-2a} d_n^{2a} \big\{(v_2-u)^{2d+1}(u-v_1)^{2d+1}\big\}^{a/2} C \le \big\{(v_2-u)(u-v_1)\big\}^{(d+\frac{1}{2})a} C \\
&\quad \le (v_2-v_1)^{(2d+1)a} C.
\end{aligned}$$

Since $(2d+1)a > 1$, Billingsley's criterium is fulfilled, and the process is tight. □

If we restrict ourselves to $d > 0$, then Lemma 4.3 leads to a particularly useful criterion in the long-memory case because it amounts to finding a bound on $E[(Y_n(v_2) - Y_n(v_1))^2]$ only.

Lemma 4.5 *Assume that* $Y_n(u)$ $(u \in [0, 1])$ *is a stochastic process with stationary increments. If*

$$E\big[\big|Y_n(v_2) - Y_n(v_1)\big|^2\big] \le (v_2 - v_1)^{2d+1}, \tag{4.2}$$

$d > 0$, *then the process is tight.*

Indeed, if we consider again $Y_n(u) = d_n^{-1} S_n(u)$, then

$$\begin{aligned}
&d_n^{-2} E\big[\big|S_n(v_2) - S_n(u)\big|\big|S_n(u) - S_n(v_1)\big|\big] \\
&\quad \le d_n^{-2} \big(E\big[\big|S_n(v_2-u)\big|^2\big]\big)^{1/2} \big(E\big[\big|S_n(u-v_1)\big|^2\big]\big)^{1/2} \\
&\quad \le d_n^{-2} d_n^2 \big\{(v_2-u)^{2d+1}(u-v_1)^{2d+1}\big\}^{1/2} C \le \big\{(v_2-u)(u-v_1)\big\}^{(d+\frac{1}{2})} C \\
&\quad \le (v_2-v_1)^{(2d+1)} C,
\end{aligned}$$

and the exponent exceeds one since $d > 0$. We note that this approach does not work when $d \le 0$. Hence, in a sense, showing tightness in a long-memory case is easier than in a weakly dependent and antipersistent situation. We note further that condition (4.2) is almost the same as a moment condition for tightness of processes in C; see Theorem 12.3 in Billingsley (1968).

4.1.2.3 Functional Central Limit Theorem for Processes

The following result is used to establish a functional limit theorem for a sum of independent stochastic processes; see e.g. p. 226 of Whitt (2002).

Lemma 4.6 *Let $X_t(u)$ ($u \in [0, \infty)$, $t \in \mathbb{N}$) be an i.i.d. sequence of processes viewed as random elements in $D[0, \infty)$. If $E(X_1(u)) = 0$, $E(X_1^2(u)) < \infty$ for each $u \in [0, \infty)$ and there exist continuous nondecreasing functions f, g and numbers $a > 1/2$, $b > 1$ such that*

$$E\left[\left(X_1(v) - X_1(u)\right)^2\right] \le \left(g(v) - g(u)\right)^a,$$
$$E\left[\left(X_1(v_2) - X_1(u)\right)^2\left(X_1(u) - X_1(v_1)\right)^2\right] \le \left(g(v_2) - g(v_1)\right)^b,$$

for all $0 \le u < v \le \infty$, $0 \le v_1 < u < v_2 < \infty$, then

$$n^{-1/2}\sum_{t=1}^{n} X_t(u) \Rightarrow G(u),$$

where G is a zero-mean Gaussian process with continuous sample paths, $cov(G(0), G(u)) = cov(X_1(0), X_1(u))$, and $\Rightarrow$ denotes weak convergence in $D[0, \infty)$.

4.1.2.4 Functional Central Limit Theorem for Inverses

The following result, known as Vervaat's lemma (see Vervaat 1972 or De Haan and Ferreira 2006), plays a crucial role in deriving limit theorems for appropriately scaled and normalized quantile processes (as inverses of empirical processes; see Sect. 4.8.2), or counting processes (as inverses of partial sum processes; see Sect. 4.9).

Lemma 4.7 (FCLT for Inverse Functions) *Denote by $D_0([0, \infty))$ the subset of $D[0, \infty)$ consisting of non-decreasing, non-negative, unbounded functions. Let $y_n(\cdot)$ ($n \ge 1$) be a sequence of elements of $D_0([0, \infty))$. Moreover, let $y(\cdot)$ be a continuous function on $[0, \infty)$, and c_n ($n \ge 1$) a sequence of positive numbers such that $c_n \to 0$. If*

$$\frac{y_n(u) - u}{c_n} \to y(u)$$

uniformly on compact sets in $[0, \infty)$, then

$$\frac{y_n^{-1}(u) - u}{c_n} \to -y(u)$$

uniformly on compact sets in $[0, \infty)$, *where* $y_n^{-1}(u) := \inf\{v : y_n(v) > u\}$ *is the generalized inverse of* $y_n(\cdot)$.

It is important to mention that the continuity assumption on $y(\cdot)$ cannot be relaxed. If the limiting function has jumps, then the uniform convergence of the inverse processes does not follow necessarily. In particular, this theorem will be applicable to situations where we have weak convergence in $D[0, 1]$ equipped with the standard J_1-topology, to a continuous process, and from that we will conclude weak convergence in that topology for the inverse processes. If the limiting process has jumps, we may not be able to conclude weak convergence of the inverse processes in the same topology, even though we may have weak convergence of the original processes. Nevertheless, at least finite-dimensional convergence follows. We refer to Whitt (2002, Chap. 13) for more details.

It is also important to see that in this lemma we assume the identity function to be the correct quantity to subtract. Thus, for instance, when dealing with the empirical distribution function $F_n(x) = n^{-1} \sum_{t=1}^{n} 1\{X_t \le x\}$ (where $X \sim F_X$), the result actually refers to $\tilde{F}_n(x) = n^{-1} \sum_{t=1}^{n} 1\{F_X(X_t) \le x\}$ and the corresponding inverse. The reason is that $F_X(X)$ is uniformly distributed, so that we are in the situation described in Vervaat's lemma. The result for F_n (and F_n^{-1}) then follows by the continuous mapping theorem.

4.1.3 Spectral Representation of Stationary Sequences

In this section we collect several standard results on spectral theory for stationary processes. Some of these properties have been used in the preliminary discussion on long memory, see Chap. 1. We state these results without a reference since they can be found in standard textbooks on time series such as Brockwell and Davis (1991).

Recall that for a zero-mean second-order stationary process X_t $(t \in \mathbb{Z})$ with autocovariances $\gamma_X(k)$, there is a spectral distribution function F such that

$$\gamma_X(k) = \int_{-\pi}^{\pi} e^{ik\lambda} \, dF(\lambda).$$

Moreover, X_t has a spectral representation of the form

$$X_t(\omega) = \int_{-\pi}^{\pi} e^{it\lambda} \, dM(\lambda; \omega),$$

where $M(\cdot; \omega)$ is a spectral measure (for simplicity, we will often write $M(\lambda)$ instead of $M(\lambda; \omega)$). The spectral measure is a complex-valued zero mean stochastic process on $[-\pi, \pi]$ with (a.s.) right-continuous sample paths and *uncorrelated* (but not necessarily independent) increments with a variance that is directly related to F. More specifically, we have

$$\begin{aligned} cov\big(dM(\lambda), dM(\nu)\big) &= E\big[dM(\lambda)\overline{dM(\nu)}\big] = 0 \quad (\lambda \neq \nu), \\ \mathrm{var}\big(dM(\lambda)\big) &= E\big[|dM(\lambda)|^2\big] = dF(\lambda). \end{aligned}$$

In particular, if the spectral density exists, then we may write the infinitesimal equation $\mathrm{var}(dM(\lambda)) = E[|dM(\lambda)|^2] = f(\lambda)\,d\lambda$.

It is important to distinguish between the role of the spectral distribution F and the spectral measure M. The spectral distribution determines the autocovariance structure, i.e. linear dependence, of the process only. In contrast, the spectral measure fully specifies the process (in the sense of the probability distribution of sample paths). In the special case where $M = M_\varepsilon$ with $E[|dM_\varepsilon(\lambda)|^2] = \sigma_\varepsilon^2/(2\pi) \cdot d\lambda$ we obtain a white noise process with variance σ_ε^2 where "white noise" stands for uncorrelated observations. This follows directly from the spectral representation

$$\varepsilon_t = \int_{-\pi}^{\pi} e^{it\lambda}\,dM_\varepsilon(\lambda) \quad (t \in \mathbb{Z}) \tag{4.3}$$

since

$$\begin{aligned} E[\varepsilon_t \varepsilon_s] &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(t\lambda - s\nu)} E\big[dM_\varepsilon(\lambda)\,\overline{dM_\varepsilon(\nu)}\,\big] \\ &= \frac{\sigma_\varepsilon^2}{2\pi}\int_{-\pi}^{\pi} e^{i(t-s)\lambda}\,d\lambda = \sigma_\varepsilon^2 \delta_{ts}. \end{aligned}$$

The spectral density of ε_t is $f_\varepsilon(\lambda) = \sigma_\varepsilon^2/(2\pi)$. One should bear in mind that, in general, this does not imply the independence of ε_t ($t \in \mathbb{Z}$). Such a direct conclusion can only be made if $M(\lambda;\omega)$ is a Gaussian process.

A zero mean, purely nondeterministic second-order stationary process always has a Wold decomposition

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = A(B)\varepsilon_t \quad (t \in \mathbb{Z})$$

with uncorrelated (i.e. "white noise") innovations ε_t and $A(z) = \sum a_j z^j$ such that $\sum_{j=0}^{\infty} a_j^2 < \infty$. Therefore, the spectral measure and spectral distribution have a simple form, namely (with equality in the $L^2(\Omega)$ sense)

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda}\,dM_X(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda} A\big(e^{-i\lambda}\big)\,dM_\varepsilon(\lambda) \quad (t \in \mathbb{Z}). \tag{4.4}$$

In other words,

$$dM_X(\lambda) = \left(\sum_{j=0}^{\infty} a_j e^{-ij\lambda}\right) dM_\varepsilon(\lambda) = A\big(e^{-i\lambda}\big)\,dM_\varepsilon(\lambda).$$

The spectral density

$$f_X(\lambda) = \frac{1}{2\pi}\sum_{k=\infty}^{\infty} \gamma_X(k)\exp(-i\lambda k)$$

is then given by

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\left|\sum_{j=0}^{\infty} a_j e^{-ij\lambda}\right|^2 = \frac{\sigma_\varepsilon^2}{2\pi}\left|A\left(e^{-i\lambda}\right)\right|^2.$$

These formulas are valid generally. More specifically, if we consider linear processes only, the ε_ts in the Wold representation are not only uncorrelated but even *independent*. This means that the increments of M_ε are independent (instead of being just uncorrelated). Even more specifically, a *Gaussian* process is a linear process that has normally distributed ε_ts, namely $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. This means that we are in the following situation. The measure M_ε is a Gaussian spectral measure such that for all sets A, $E[M_\varepsilon(A)] = 0$, $E_\varepsilon[M(A \cap B)] = 0$ for all disjoint sets A and B, and $E[M_\varepsilon(A)\overline{M_\varepsilon(A)}] = \sigma_\varepsilon^2|A|/(2\pi)$, where $|\cdot|$ denotes the Lebesgue measure. Moreover, for all $\lambda_1 \le \lambda_2 < \lambda_3 \le \lambda_4$, the increments $M_\varepsilon(\lambda_4) - M_\varepsilon(\lambda_3)$ and $M_\varepsilon(\lambda_2) - M_\varepsilon(\lambda_1)$ are independent. (For simplicity of notation, we will mostly assume that $\sigma_\varepsilon^2 = 1$, which means that $M_\varepsilon(\cdot)$ is a spectral measure of an i.i.d. $N(0, 1)$ sequence.) The Gaussian process X_t is then given by

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \int_{-\pi}^{\pi} e^{it\lambda}\, dM_X(\lambda) \quad (t \in \mathbb{N}), \tag{4.5}$$

where M_X is the Gaussian spectral measure defined by

$$dM_X(\lambda) = \left(\sum_{j=0}^{\infty} a_j e^{-ij\lambda}\right) dM_\varepsilon(\lambda) = A\left(e^{-i\lambda}\right) dM_\varepsilon(\lambda) =: \sqrt{2\pi}\, a(\lambda)\, dM_\varepsilon(\lambda).$$

Note that in the notation with $a(\lambda)$, the spectral density can be written as

$$f_X(\lambda) = \sigma_\varepsilon^2 \left|a(\lambda)\right|^2.$$

Thus, for $\sigma_\varepsilon^2 = 1$, we have the identity $f_X(\lambda) = |a(\lambda)|^2$.

Another result that is very useful in many situations, such as prediction or (Gaussian) maximum likelihood estimation, is the following factorization of the spectral density. Let us write $\log f_X$ as a Fourier series

$$\log f_X(\lambda) = \sum_{j=-\infty}^{\infty} \alpha_j e^{-ij\lambda}$$

with coefficients

$$\alpha_j = \alpha_{-j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ij\lambda} \log f_X(\lambda)\, d\lambda. \tag{4.6}$$

Then we obtain the factorization

$$f_X(\lambda) = \exp(\alpha_0)\left|A\left(e^{-i\lambda}\right)\right|^2 = \frac{\sigma_\varepsilon^2}{2\pi}\left|A\left(e^{-i\lambda}\right)\right|^2 =: \frac{\sigma_\varepsilon^2}{2\pi} h_X(\lambda), \tag{4.7}$$

where

$$A(z) = \sum_{j=0}^{\infty} a_j z^j = \exp\left(\sum_{j=1}^{\infty} \alpha_j z^j\right)$$

and

$$\frac{\sigma_\varepsilon^2}{2\pi} = \exp(\alpha_0).$$

The last equation, together with (4.6), implies

$$\alpha_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_X(\lambda)\, d\lambda = \log \sigma_\varepsilon^2 - \log 2\pi.$$

For the function $h_X(\cdot)$ defined in (4.7), we therefore obtain

$$\int_{-\pi}^{\pi} \log h_X(\lambda)\, d\lambda = 0. \tag{4.8}$$

This property is particularly useful for the asymptotic theory of (Gaussian) quasi-maximum likelihood estimation.

Finally, the following lemma is useful in spectral analysis of stationary sequences (see Lemma 2 in Moulines et al. 2007a). Consider the spectral radius $Sp(A)$ of an $n \times n$ matrix A, defined as the maximal absolute eigenvalue, or

$$Sp(A) = \sup_{\mathbf{x}\in\mathbb{R}^n:\|\mathbf{x}\|\le 1} \mathbf{x}^T A\mathbf{x}.$$

Now let $A = \Sigma_n = [\gamma_X(i-j)]_{i,j=1,\ldots,n}$ be the covariance matrix of $X = (X_1,\ldots,X_n)^T$, where X_t is a zero-mean stationary process with spectral density f_X. Then

$$\begin{aligned}
\mathbf{x}^T A\mathbf{x} &= \sum_{j,l=1}^{n} \gamma_X(j-l) x_j x_l \\
&= \int_{-\pi}^{\pi} f_X(\lambda) \left|\sum_{j=1}^{n} x_j \exp(-ij\lambda)\right|^2 d\lambda \\
&\le \sup_{\lambda\in[-\pi,\pi]} \left|f_X(\lambda)\right| \int_{-\pi}^{\pi} \left|\sum_{j=1}^{n} x_j \exp(-ij\lambda)\right|^2 d\lambda = 2\pi|\mathbf{x}|^2 \sup_{\lambda\in[-\pi,\pi]} \left|f_X(\lambda)\right|,
\end{aligned}$$

where the last expression follows from the Parseval identity. Hence, we have the following result.

Lemma 4.8 *Assume that X_t $(t \in \mathbb{Z})$ is a stationary process with the spectral density f_X. Assume that Σ_n is the covariance matrix of $X_1, \ldots, X_n$. Then*

$$Sp(\Sigma_n) \le 2\pi \sup_{\lambda \in [-\pi,\pi]} \left| f_X(\lambda) \right|.$$

4.2 Limit Theorems for Sums with Finite Moments

4.2.1 Introduction

Let X_t $(t \in \mathbb{N})$ be a stationary process. The asymptotic behaviour of partial sums

$$S_n(u) = S_{n,G}(u) = \sum_{t=1}^{[nu]} G(X_t) \tag{4.9}$$

is at the core of probability theory. In this section we present limit theorems for partial sums associated with long-memory or antipersistent processes. Two types of distinctions have to be made. One is between linear and nonlinear processes. The other is between processes with finite and infinite variance. The case of infinite variance is studied in Sect. 4.3. Depending on which of these cases is considered, different results and mathematical techniques are required.

In this section we discuss finite-variance processes only. We will begin our exposition by assuming that X_t $(t \in \mathbb{N})$ is a Gaussian process, since computations and proofs are technically less challenging than for instance for general Appell polynomials. The limiting phenomena related to partial sums of subordinated Gaussian sequences were observed first by Rosenblatt (1961) and then developed independently by Taqqu (1975, 1977, 1979), Dobrushin (1980) and Dobrushin and Major (1979). Further developments can be found in Breuer and Major (1983), Giraitis and Surgailis (1985), Ho and Sun (1987, 1990), Dehling and Taqqu (1989a, 1989b) and Arcones (1994). Although the original technique in Taqqu (1975) to show convergence to the so-called Hermite–Rosenblatt distribution was based on characteristic functions, the common method to obtain a non-central limit theorem is based on (multiple) Wiener–Itô integrals, together with the diagram formula. For long-memory linear processes, the first result was obtained in Davydov (1970a, 1970b); see also Gorodetskii (1977), Lang and Soulier (2000), Wang et al. (2003).

As for subordinated linear processes, there are two common approaches: Appell polynomials (Surgailis 1981, 1982; Giraitis 1985; Giraitis and Surgailis 1986, 1989; Avram and Taqqu 1987; Surgailis and Vaičiulis 1999; Surgailis 2000; also see Surgailis 2003 for a review) and a martingale decomposition (Ho and Hsing 1996, 1997; Giraitis and Surgailis 1999; Wu 2003; see also Hsing 2000 for a review).

The theory for nonlinear models with long memory is less well developed. EGARCH-type models were considered in Surgailis and Viano (2002), whereas results for LARCH(∞) processes can be found for instance in Giraitis et al. (2000c), Giraitis and Surgailis (2002), Berkes and Horváth (2003), Beran (2006).

4.2.2 Normalizing Constants for Stationary Processes

Before getting into the details of limiting distributions, a first question can be answered relatively easily, namely which normalizing sequences should be used to obtain nondegenerate limits. Let $S_n = \sum_{t=1}^{n} X_t$, where X_t $(t \in \mathbb{N})$ is a stationary sequence with appropriate moment conditions. We consider the asymptotic behaviour of $\mathrm{var}(S_n)$ in three cases: long memory, short memory and antipersistence.

Lemma 4.9 (Long Memory) *Let X_t $(t \in \mathbb{N})$ be a stationary sequence with $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$ $(k \to \infty)$ for some $0 < d < \frac{1}{2}$, where L_γ is slowly varying at infinity. Then, as $n \to \infty$,*

$$\mathrm{var}(S_n) \sim L_S(n) n^{2d+1} \tag{4.10}$$

with

$$L_S(n) = L_1(n) = C_1 L_\gamma(n) = \frac{1}{d\,(2d+1)} L_\gamma(n). \tag{4.11}$$

Proof We have

$$\begin{aligned} \mathrm{var}(S_n) &= n \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma_X(k) \\ &\sim n \sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} L_\gamma(k) |k|^{2d-1} - \sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} L_\gamma(k) |k|^{2d}. \end{aligned}$$

The last expression can be written as

$$\begin{aligned} & L_\gamma(n) n^{2d+1} \left[\sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} \frac{L_\gamma(k)}{L_\gamma(n)} \left(\frac{|k|}{p}\right)^{2d-1} n^{-1} \right. \\ & \quad \left. - \sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} \frac{L_\gamma(k)}{L_\gamma(n)} \left(\frac{|k|}{n}\right)^{2d} n^{-1} \right] \\ & \sim 2 L_\gamma(n) n^{2d+1} \left[\int_0^1 u^{2d-1}\, du - \int_0^1 u^{2d}\, du \right] \\ & = 2 L_\gamma(n) n^{2d+1} \left(\frac{1}{2d} - \frac{1}{2d+1} \right) = \frac{L_\gamma(n)}{d(2d+1)} n^{2d+1}. \end{aligned}$$

□

Lemma 4.10 (Short Memory) *Let X_t $(t \in \mathbb{N})$ be a stationary sequence with $\sum_{k=-\infty}^{\infty} \gamma_X(k) > 0$ and $\sum_{k=-\infty}^{\infty} |\gamma_X(k)| < \infty$. Then, as $n \to \infty$,*

$$\mathrm{var}(S_n) \sim c_S n \tag{4.12}$$

with

$$c_S = \sum_{k=-\infty}^{\infty} \gamma_X(k). \tag{4.13}$$

Proof Cesaro summability implies

$$\sum_{k=-(n-1)}^{n-1} \frac{k}{n}\gamma_X(k) \to 0,$$

so that

$$\mathrm{var}(S_n) \sim n \sum_{k=-(n-1)}^{n-1} \gamma_X(k) \sim c_S n.$$

□

Lemma 4.11 (Antipersistence) *Let* X_t $(t \in \mathbb{N})$ *be a stationary sequence with* $\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$ $(k \to \infty)$ *for some* $-\frac{1}{2} < d < 0$, *where* L_γ *is slowly varying at infinity, and*

$$\sum_{k=-\infty}^{\infty} \gamma_X(k) = 0.$$

Then, as $n \to \infty$,

$$\mathrm{var}(S_n) \sim L_S(n)n^{2d+1} \tag{4.14}$$

with

$$L_S(n) = \frac{1}{d(2d+1)}L_\gamma(n). \tag{4.15}$$

Proof

$$\sum_{k=-(n-1)}^{n-1} \gamma_X(k) = -2\sum_{k=n}^{\infty}\gamma_X(k) \sim -2L_\gamma(n)\sum_{k=n}^{\infty}k^{2d-1}$$
$$\sim -2L_\gamma(n)n^{2d}\int_1^{\infty} u^{2d-1}\,du = \frac{2L_\gamma(n)}{2d}n^{2d}.$$

Then the result follows by the same arguments as in the long-memory case. □

Note that in the proof of Lemma 4.11, the Riemann approximation could not be applied to $\sum_{k=-(n-1)}^{n-1}\gamma_X(k)$ directly because u^{2d-1} is not integrable at the origin for $d < 0$. Note also that in the antipersistent case, $L_\gamma(k) < 0$ for k large enough. However, since $L_\gamma(k)$ is multiplied by d^{-1}, the slowly varying function $L_S(n)$ is positive asymptotically.

Taking into account Theorem 1.3, a unified formula including (4.10), (4.12) and (4.14) can be written in terms of the spectral density. Using the notation

$$L_f(\lambda) = L_\gamma\left(\lambda^{-1}\right)\pi^{-1}\Gamma(2d)\sin\left(\frac{\pi}{2} - \pi d\right)$$

and

$$\begin{aligned} \nu(d) &= \frac{2\sin\pi d}{d(2d+1)}\Gamma(1-2d) \quad (d \neq 0), \\ \nu(0) &= \lim_{d\to 0}\nu(d) = 2\pi, \end{aligned} \tag{4.16}$$

we have

$$\operatorname{var}(S_n) \sim \nu(d)L_f\left(n^{-1}\right)n^{2d+1} \sim \nu(d)f_X\left(n^{-1}\right)n.$$

4.2.3 Subordinated Gaussian Processes

We begin our exposition by assuming that X_t $(t \in \mathbb{N})$ are normal random variables because computations and proofs are technically less challenging than in the case of Appell polynomials, for instance. The limiting phenomena related to partial sums of subordinated Gaussian sequences were first observed by Rosenblatt (1961) and then developed independently by Taqqu (1975, 1977, 1979) Dobrushin (1980) and Dobrushin and Major (1979). Further developments can be found in Breuer and Major (1983), Giraitis and Surgailis (1985), Ho and Sun (1987, 1990) and Arcones (1994). Although the original technique in Taqqu (1975) to show convergence to the so-called Hermite–Rosenblatt distribution was based on characteristic functions, the common method to obtain non-central limit theorems is based on (multiple) Wiener–Itô integrals, together with the diagram formula.

4.2.3.1 Moment Bounds and Normalizing Constants

Recall from Sect. 3.1.2 that each function $G(\cdot)$ in $L^2(\mathbb{R},\phi)$ with $\phi(x) = (2\pi)^{-1/2} \times \exp(-x^2/2)$ can be expanded as

$$G(X) = E\left[G(X)\right] + \sum_{l=1}^{\infty}\frac{J(l)}{l!}H_l(X) = E\left[G(X)\right] + \sum_{l=m}^{\infty}\frac{J(l)}{l!}H_l(X),$$

where $J(l) = E[G(X)H_l(X)]$, X is a standard Gaussian random variable, and m is the Hermite rank of G (i.e. the smallest $m \geq 1$ such that $J(m) \neq 0$). Moreover, recall the formula (3.16) for $H_m(\sum_{j=1}^{l} a_j x_j)$,

$$H_m\left(\sum_{j=1}^{l} a_j x_j\right) = \sum_{m_1+\cdots+m_k=m}\frac{m!}{m_1!\ldots m_k!}\prod_{j=1}^{l} a_j^{m_j}H_{m_j}(x_j). \tag{4.17}$$

This was used for deriving the formula for covariances of Hermite polynomials given in Lemma 3.5. For convenience, we repeat the result here:

Lemma 4.12 *Let* X_1, X_2 *be a pair of jointly standard normal random variables with covariance* $\gamma = cov(X_1, X_2)$. *Then*

$$cov\big(H_l(X_1), H_l(X_2)\big) = l!\gamma^l, \tag{4.18}$$

whereas for $j \neq l$,

$$cov\big(H_j(X_1), H_l(X_2)\big) = 0. \tag{4.19}$$

In particular, assume now that

$$\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$$

with $d \in (0, 1/2)$, and consider the sum of $H_m(X_t)$. From Lemma 4.12 we see that if $d > 1 - \frac{1}{2}m^{-1}$, the autocovariance $\gamma_{H_m}(k) = cov(H_m(X_t), H_m(X_{t+k}))$ of the transformed process $H_m(X_t)$ is not summable because it is (up to the slowly varying function) of the order $k^{m(2d-1)}$ with $m(2d-1) > -1$. Using the same argument as in the proof of Lemma 4.9, we then obtain

$$\mathrm{var}\left(\sum_{t=1}^{n} H_m(X_t)\right) = m!\sum_{k=1}^{n}\sum_{j=1}^{n}\gamma_X^m(j-k) \sim L_m(n)n^{(2d-1)m+2}, \tag{4.20}$$

where

$$L_m(n) = m!C_m L_\gamma^m(n) \tag{4.21}$$

and

$$C_m = \frac{2}{[(2d-1)m+1][(2d-1)m+2]}. \tag{4.22}$$

Furthermore, if G has the Hermite rank m, then the variance of $G(X)$ can be decomposed into (orthogonal) contributions of the Hermite coefficients,

$$\mathrm{var}\big(G(X)\big) = \sum_{l=1}^{\infty}\left(\frac{J(l)}{l!}\right)^2 l! = \sum_{l=m}^{\infty}\frac{J^2(l)}{l!}. \tag{4.23}$$

Similarly, if X_1 and X_2 are as in Lemma 4.12,

$$cov\big(G(X_1), G(X_2)\big) = \sum_{l=m}^{\infty}\frac{J^2(l)}{l!}\gamma^l. \tag{4.24}$$

Consequently, applying this to the stationary Gaussian sequence X_t ($t \in \mathbb{N}$), we obtain

$$\gamma_G(k) = cov\big(G(X_t), G(X_{t+k})\big) = \sum_{l=m}^{\infty}\frac{J^2(l)}{l!}\gamma_X^l(k). \tag{4.25}$$

Thus, as $k \to \infty$, the asymptotic behaviour of $cov(G(X_t), G(X_{t+k}))$ is determined by the leading term $(J^2(m)/m!)\gamma_X^m(k)$. From (4.25) we therefore conclude that for a function G with the Hermite rank m, the asymptotic behaviour of the autocovariance is given by

$$\gamma_G(k) \sim \frac{J^2(m)}{m!} L_\gamma^m(k) k^{m(2d-1)} \quad (k \to \infty).$$

Therefore, if $m(1-2d) < 1$, then by the same argument as in (4.20),

$$\operatorname{var}\left(\sum_{t=1}^{n} G(X_t)\right) \sim \frac{J^2(m)}{m!} C_m L_\gamma^m(n) n^{(2d-1)m+2} = \left(\frac{J(m)}{m!}\right)^2 L_m(n) n^{(2d-1)m+2}, \tag{4.26}$$

where C_m is the constant in (4.22), and $L_m(\cdot)$ is the slowly varying function defined in (4.21). Otherwise, if $m(1-2d) > 1$, then

$$\sum_{k=1}^{\infty} \left| cov\big(G(X_t), G(X_{t+k})\big)\right| < \infty.$$

Therefore, one can expect two different types of convergence: either a long-memory type where the normalization for partial sums is

$$n^{-((d-\frac{1}{2})m+1)} L_m^{-\frac{1}{2}}(n) = n^{-\frac{1}{2}-((m-1)/2-d)} L_m^{-\frac{1}{2}}(n) \tag{4.27}$$

or a weakly-dependent type with the usual normalization $n^{-1/2}$.

We conclude the discussion of normalizing constants by mentioning two useful bounds derived by Arcones (1994):

- If $m(1-2d) < 1$, then there is a constant C such that for any function G with Hermite rank m,

$$\operatorname{var}\left(n^{-1}\sum_{t=1}^{n} G(X_t)\right) \le C\gamma_X^m(n) \operatorname{var}\big(G(X_1)\big).$$

- If $m(1-2d) > 1$, then there is a constant C such that for any function G with Hermite rank m,

$$\operatorname{var}\left(n^{-1}\sum_{t=1}^{n} G(X_t)\right) \le Cn^{-1} \operatorname{var}\big(G(X_1)\big).$$

The first inequality looks very similar to (4.26). However, the important difference is that the constant C depends on the Gaussian process X_t only and not on the function G.

4.2.3.2 Limiting Distribution

The Hermite rank of $G(x) = x$ is one. Furthermore, $\sum_{t=1}^{[nu]} X_t$ is normally distributed for all n and $u \in [0,1]$. Therefore, in view of (4.27), the following result is obvious. Note that it is valid for all values of $d \in (-\frac{1}{2}, \frac{1}{2})$, i.e. for long memory ($d \in (0, \frac{1}{2})$), short memory ($d = \frac{1}{2}$) and antipersistence ($d \in (-\frac{1}{2}, 0)$). The limiting process is Gaussian. The dependence structure of the increments depends on d.

Theorem 4.2 *Assume that X_t ($t \in \mathbb{N}$) is a stationary sequence of standard normal random variables such that $f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d}$ with $d \in (-1/2, 1/2)$ and the assumptions of Lemma 4.9 (for $d > 0$, Lemma 4.10) (for $d = 0$) or Lemma 4.11 (for $d < 0$) hold respectively. Let $S_n(u) = \sum_{t=1}^{[nu]} X_t$. Then*

$$n^{-(d+\frac{1}{2})} L_1^{-\frac{1}{2}}(n) S_n(u) \Rightarrow B_H(u) \quad \bigl(u \in [0,1]\bigr),$$

where $B_H(\cdot)$ is a standard fractional Brownian motion with Hurst parameter $H = d + \frac{1}{2}$, "$\Rightarrow$" denotes weak convergence in $D[0,1]$, and $L_1(n) = L_f(n^{-1})\nu(d)$ with $\nu(d)$ defined in (4.16).

Proof As mentioned in the introduction to this chapter, we prove finite-dimensional convergence just in the one-dimensional case. Clearly, $S_n(u)$ is normal, and $r_n^2 = \operatorname{var}(S_n(1))/(n^{2d+1} L_1(n)) \to 1$. Thus, with $d_n^2 = n^{2d+1} L_1(n)$,

$$E\bigl(e^{i\theta d_n^{-1} S_n(1)}\bigr) = \exp\biggl(-\frac{1}{2}\theta^2 r_n^2\biggr) \to \exp\bigl(-\theta^2/2\bigr).$$

Thus, one-dimensional distributions of $S_n(u)$ converge to the standard normal distribution.

For tightness, note that $S_n(1)$ is normal, so that $E[S_n^{2l}(1)]$ ($l \in \mathbb{N}$) is proportional to $(E[S_n^2(1)])^l$. Therefore, the conditions of Lemma 4.4 are fulfilled, and tightness follows. □

We will now present another proof of this theorem. The reason is that it will be easily extendable to more complicated cases of general Hermite polynomials and non-normal random variables. Recall some notions on the spectral representation of stationary time series from Sect. 4.1.3. Let ε_t ($t \in \mathbb{Z}$) be a centred, finite-variance i.i.d. sequence. Then ε_t can be represented in terms of a Gaussian spectral measure with uncorrelated increments,

$$\varepsilon_t = \int_{-\pi}^{\pi} e^{it\lambda}\, dM_\varepsilon(\lambda) \quad (t \in \mathbb{Z}).$$

Recall also that

$$E\bigl[\bigl|dM_\varepsilon(\lambda)\bigr|^2\bigr] = \frac{\sigma_\varepsilon^2}{2\pi}\, d\lambda = f_\varepsilon(\lambda)\, d\lambda,$$

where $\sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_t)$. Without loss of generality, we will assume that $\sigma_\varepsilon^2 = 1$ in the following. Moreover it will be convenient to use instead of M_ε the spectral measure

$$M_0(A) = \sqrt{2\pi} M_\varepsilon(A),$$

so that

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{it\lambda} \, dM_0(\lambda)$$

and $E[|dM_0(\lambda)|^2] = d\lambda$. For a linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ $(t \in \mathbb{Z})$ with $\sum_{j=0}^{\infty} a_j^2 < \infty$ (and $\sigma_\varepsilon^2 = 1$), one then has the spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \, dM_X(\lambda) \quad (t \in \mathbb{Z}) \tag{4.28}$$

with

$$\begin{aligned} dM_X(\lambda) &= \left(\sum_{j=0}^{\infty} a_j e^{-ij\lambda} \right) dM_\varepsilon(\lambda) = A\left(e^{-i\lambda}\right) dM_\varepsilon(\lambda) \\ &= \frac{1}{\sqrt{2\pi}} A\left(e^{-i\lambda}\right) dM_0(\lambda) =: a(\lambda) \, dM_0(\lambda). \end{aligned}$$

The spectral density of X_t is

$$f_X(\lambda) = \frac{1}{2\pi} \left| A\left(e^{-i\lambda}\right) \right|^2 = \left| a(\lambda) \right|^2.$$

Assume that $f_X(\lambda) = L_f(\lambda)|\lambda|^{-2d}$ as $\lambda \to 0$ or $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$ as $k \to \infty$. Recall that, under suitable conditions, these assumptions are equivalent to

$$L_f(\lambda) = L_\gamma\left(\lambda^{-1}\right) \pi^{-1} \Gamma(2d) \sin\left(\frac{\pi}{2} - \pi d \right)$$

and

$$L_\gamma(k) = 2 L_f\left(k^{-1}\right) \Gamma(1 - 2d) \sin(\pi d). \tag{4.29}$$

Then $|a(\lambda)| = L_f^{1/2}(\lambda) |\lambda|^{-d}$. Now, we are ready to present an alternative proof of Theorem 4.2. This type of approach was initiated in Dobrushin (1980), Dobrushin and Major (1979); also see Arcones (1994) and Lang and Soulier (2000). We will use a representation of a fractional Brownian motion that appears in Sect. 3.7.1.

Alternative proof of Theorem 4.2 Let $S_n = S_n(1) = \sum_{t=0}^{n-1} X_t$ (note that we take summation from $t = 0$ to $n - 1$) and write the spectral representation

$$
\begin{aligned}
S_n &= \sum_{t=0}^{n-1} \int_{-\pi}^{\pi} e^{it\lambda} dM_X(\lambda) \\
&= \sum_{t=0}^{n-1} \int_{-\pi}^{\pi} e^{it\lambda} a(\lambda)\, dM_0(\lambda) = \int_{-\pi}^{\pi} \left(\sum_{t=0}^{n-1} e^{it\lambda} \right) a(\lambda)\, dM_0(\lambda) \\
&= \int_{-\pi}^{\pi} \frac{e^{i\lambda n} - 1}{e^{i\lambda} - 1} a(\lambda)\, dM_0(\lambda) \\
&= n^{1/2} \int_{-n\pi}^{n\pi} D_n(\lambda/n) a\left(\frac{\lambda}{n}\right) n^{1/2}\, dM_0\big(n^{-1}\lambda\big),
\end{aligned}
$$

where

$$
D_n(\lambda) = \frac{e^{i\lambda n} - 1}{n(e^{i\lambda} - 1)} 1\big\{|\lambda| \le \pi n\big\}. \tag{4.30}
$$

Since $\lim_{u\to 0}(e^{\lambda u} - 1)/u = \lambda$, we conclude that

$$
\lim_{n\to\infty} D_n(\lambda/n) \to \frac{e^{i\lambda} - 1}{i\lambda} =: D(\lambda). \tag{4.31}
$$

Now, $E(|dM_0(n^{-1}\lambda)|^2) = n^{-1} d\lambda$. Hence, $n^{1/2} M_0(n^{-1}A)$ and $M_0(A)$ have the same distribution (as stochastic processes indexed by A), and we can write

$$
S_n \overset{\mathrm{d}}{=} n^{1/2} \int_{-n\pi}^{n\pi} D_n(\lambda/n) a\left(\frac{\lambda}{n}\right) dM_0(\lambda) \approx n^{1/2} \int_{-\infty}^{\infty} D_n(\lambda/n) a\left(\frac{\lambda}{n}\right) dM_0(\lambda).
$$

Consequently, we have two possible scenarios:

- $\lim_{\lambda\to 0} a(\lambda) = a(0) = \sqrt{f_X(0)} \ne 0$. Then we expect

$$
n^{-1/2} S_n \overset{\mathrm{d}}{\to} a(0) \int_{-\infty}^{\infty} \frac{e^{i\lambda} - 1}{i\lambda}\, dM_0(\lambda).
$$

- $a(\lambda) = L_f^{1/2}(\lambda)|\lambda|^{-d}$, $d \in (-1/2, 0) \cup (0, 1/2)$. Then we expect

$$
n^{-(1/2+d)} L_f^{-1/2}\big(n^{-1}\big) S_n \overset{\mathrm{d}}{\to} \int_{-\infty}^{\infty} D(\lambda) \frac{1}{|\lambda|^d}\, dM_0(\lambda). \tag{4.32}
$$

In the latter case, applying (4.21) and (4.22) with $m = 1$ and (4.29), we obtain

$$
L_1(n) = \frac{2\Gamma(1-2d)\sin\pi d}{d(2d+1)} L_f\big(n^{-1}\big) =: K_1^{-2}(1,d) L_f\big(n^{-1}\big).
$$

Thus,

$$
n^{-(1/2+d)} L_1^{-1/2}(n) S_n = K_1(1,d) \int_{-\infty}^{\infty} |\lambda|^{-d} \frac{e^{i\lambda} - 1}{i\lambda}\, dM_0(\lambda).
$$

Recall Proposition 3.1. We can verify that $K_1(1,d)$ agrees with $K_1(1,H)$ there by setting $H = d + \frac{1}{2}$, so that the limiting random variable is $B_H(1)$.

To make the argument (4.32) precise, we note that for $|\lambda| < \pi n$,

$$\left|D_n(\lambda/n) - D(\lambda)\right| = \left|\frac{e^{i\lambda} - 1}{n(e^{i\lambda/n} - 1)} - \frac{e^{i\lambda} - 1}{i\lambda}\right| = O\left(n^{-1}\right)$$

uniformly w.r.t. λ (the bound does not depend on λ). Thus,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \left|D_n(\lambda/n) - D(\lambda)\right|^2 d\lambda \\
&= \int_{-n\pi}^{n\pi} \left|D_n(\lambda/n) - D(\lambda)\right|^2 d\lambda \\
&\quad + \int_{|\lambda|>n\pi} \left|D(\lambda)\right|^2 d\lambda \le O\left(n^{-1}\right) + 2\int_{|\lambda|>n\pi} \frac{1}{|\lambda|^2}\, d\lambda = O\left(n^{-1}\right).
\end{aligned}$$

We conclude that $D_n(\lambda/n)$ converges to $D(\lambda)$ in $L^2(\mathbb{R}, d\lambda)$ (here "$d\lambda$" stands for the Lebesgue measure). Also,

$$n^{-d} L_f^{-1/2}\left(n^{-1}\right) D_n(\lambda/n) a\left(\frac{\lambda}{n}\right)$$

converges in $L^2(\mathbb{R}, d\lambda)$ to $D(\lambda)|\lambda|^{-d}$. Since

$$\begin{aligned}
&E\left[\left(\int_{-\infty}^{\infty}\left(n^{-d} L_f^{-1/2}\left(n^{-1}\right) D_n(\lambda/n) a\left(\frac{\lambda}{n}\right) - D(\lambda)|\lambda|^{-d}\right) dM_0(\lambda)\right)^2\right] \\
&= \int_{-\infty}^{\infty}\left(n^{-d} L_f^{-1/2}\left(n^{-1}\right) D_n(\lambda/n) a\left(\frac{\lambda}{n}\right) - D(\lambda)|\lambda|^{-d}\right)^2 d\lambda \to 0,
\end{aligned}$$

we conclude the convergence in L^2. Thus, the result of Proposition 4.2 follows. □

The limiting distribution in formula (4.32) can be also written as

$$n^{-(1/2+d)} L_f^{-1/2}\left(n^{-1}\right) S_n(1) \xrightarrow{\mathrm{d}} \int_{-\infty}^{\infty} D(\lambda)\, dW_X(\lambda), \tag{4.33}$$

where

$$dW_X(\lambda) = \frac{1}{|\lambda|^d}\, dM_0(\lambda). \tag{4.34}$$

The measure W_X is called the limiting spectral measure that depends (via the parameter d) on the sequence X_t. This representation will be essential in Sect. 4.4.

The longish version of the proof of Theorem 4.2 will allow us to obtain the limiting behaviour of subordinated Gaussian sequences. First, we extend the theorem to partial sum processes $S_{n,H_m}(u) := \sum_{t=1}^{[nu]} H_m(X_t)$, where H_m is the mth Hermite polynomial. Remarkably, the limit is no longer an fBm process, provided that long memory is strong enough and $m \geq 2$. This was first observed in Rosenblatt (1961), also see Taqqu (1975). Note that their method of proof is based on characteristic functions and is different from the one used in the alternative proof of Theorem 4.2.

Theorem 4.3 *Assume that X_t $(t \in \mathbb{N})$ is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$ with $d \in (0, 1/2)$. Let $S_{n,H_m}(u) = \sum_{t=1}^{[nu]} H_m(X_t)$. If $m(1-2d) < 1$, then*

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) S_{n,H_m}(u) \Rightarrow Z_{m,H}(u) \quad \bigl(u \in [0, 1]\bigr),$$

where $Z_{m,H}(\cdot)$ is a Hermite–Rosenblatt process with $H = d + \frac{1}{2}$, $\Rightarrow$ denotes weak convergence in $D[0, 1]$, and $L_m(n) = m!C_m L_\gamma^m(n)$, see (4.21) and (4.22).

Note that this type of convergence requires long memory to be strong enough. In particular, if $m = 2$, we require $d \in (1/4, 1/2)$. If this is not the case, then the partial sum process has weak dependence properties.

Example 4.1 Assume that $m = 2$. If $d \in (1/4, 1/2)$, then

$$n^{-2d} L_2^{-1/2}(n) \sum_{t=1}^{[nu]} \bigl(X_t^2 - 1\bigr) \Rightarrow Z_{2,H}(u),$$

where

$$L_2(n) = 2C_2 L_\gamma^2(n),$$

$$C_2 = \frac{1}{(2(2d-1)+1)(2d+1)}.$$

For each fixed $u \in [0, 1]$, the limit is non-normal. This will be illustrated by simulations in computer Example 4.3 later in this section.

Proof of Theorem 4.3 The proof is almost a copy of the alternative proof of Theorem 4.2. We replace (4.28) by

$$H_m(X_t) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} e^{it(\lambda_1+\cdots+\lambda_m)}\, dM_X(\lambda_1) \cdots dM_X(\lambda_m)$$

(we refer to Sect. 3.7.1.3 for the formula and the meaning of this integral). Recalling

$$dM_X(\lambda) = \sqrt{2\pi}\, a(\lambda)\, dM_\varepsilon(\lambda) = a(\lambda)\, dM_0(\lambda),$$

we have

$$S_{n,H_m}(1) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \frac{e^{in(\lambda_1+\cdots+\lambda_m)}-1}{e^{i(\lambda_1+\cdots+\lambda_m)}-1} \prod_{r=1}^{m} a(\lambda_r)\, dM_0(\lambda_1)\cdots dM_0(\lambda_m)$$
$$= \frac{n}{n^{m/2}} \int \cdots \int D_n\left(\frac{\lambda_1+\cdots+\lambda_m}{n}\right)$$
$$\times \prod_{r=1}^{m} a\left(\frac{\lambda_r}{n}\right) n^{1/2}\, dM_0\left(n^{-1}\lambda_1\right)\cdots n^{1/2}\, dM_0\left(n^{-1}\lambda_m\right),$$

where the integration is over $[-n\pi, n\pi]^m$. Therefore, if $a(\lambda) = L_f^{1/2}(\lambda)|\lambda|^{-d}$, $d \in (0, 1/2)$, then we expect

$$n^{-(1-m(\frac{1}{2}-d))} L_f^{-m/2}\left(n^{-1}\right) S_{n,H_m}(1)$$
$$\overset{\mathrm{d}}{\to} \int_{\mathbb{R}^m} D(\lambda_1+\cdots+\lambda_m) \prod_{r=1}^{m} \frac{1}{|\lambda_r|^d}\, dM_0(\lambda_1)\cdots dM_0(\lambda_m), \tag{4.35}$$

cf. (4.31). Again, we identify

$$L_m(n) = m! C_m \left(2\Gamma(1-2d)\sin\pi d\right)^m L_f^m\left(n^{-1}\right) = K_1^{-2}(m,d) L_f^m\left(n^{-1}\right),$$

and from Proposition 3.1 we recognize the representation of the Hermite–Rosenblatt process.

A precise argument for (4.35) is the same as in the case $m = 1$; see the proof of Proposition 4.2. Furthermore, we do not verify tightness here since it will be done in the next theorem. □

Finally, convergence of partial sums $S_{n,G}(u) = \sum_{t=1}^{[nu]} G(X_t)$ is just a consequence of Theorem 4.3, using the so-called reduction principle, proven originally in Taqqu (1975).

Theorem 4.4 *Assume that X_t $(t \in \mathbb{N})$ is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$ $(d \in (0, 1/2))$. Let $S_{n,G}(u) = \sum_{t=1}^{[nu]} G(X_t)$, where G is a function such that $E[G(X_1)] = 0$, $E[G^2(X_1)] < \infty$. If m is the Hermite rank of G and $m(1-2d) < 1$, then*

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) S_{n,G}(u) \Rightarrow \frac{J(m)}{m!} Z_{m,H}(u) \quad \left(u \in [0,1]\right),$$

where $Z_{m,H}(\cdot)$ is a Hermite–Rosenblatt process, $H = d + \frac{1}{2}$, $\Rightarrow$ denotes weak convergence in $D[0, 1]$, *and L_m is given in* (4.21):

$$L_m(n) = m! C_m L_\gamma^m(n).$$

Proof Decompose

$$G(x) = \frac{J(m)}{m!} H_m(x) + \sum_{l=m+1}^{\infty} \frac{J(l)}{l!} H_l(x) =: \frac{J(m)}{m!} H_m(x) + G^*(x).$$

Using (4.18) and (4.25), we have

$$cov\left[\frac{J(m)}{m!} H_m(X_0), \frac{J(m)}{m!} H_m(X_k)\right] = \frac{J^2(m)}{m!} \gamma_X^m(k)$$

and

$$cov\left[G^*(X_0), G^*(X_k)\right] = \sum_{l=m+1}^{\infty} \frac{J^2(l)}{l!} \gamma_X^l(k).$$

Furthermore, for any t, s, the random variables $G^*(X_t)$ and $H_m(X_s)$ are uncorrelated. Therefore,

$$\begin{aligned} \mathrm{var}\left(\sum_{t=1}^{n} G(X_t)\right) &= \sum_{t=1}^{n}\sum_{s=1}^{n} E\left[G^*(X_t)G^*(X_s)\right] + \frac{J^2(q)}{m!} \sum_{t=1}^{n}\sum_{s=1}^{n} \gamma_X^m\left(|t-s|\right) \\ &= \sum_{t=1}^{n}\sum_{s=1}^{n} E\left[G^*(X_t)G^*(X_s)\right] + \left(\frac{J(m)}{m!}\right)^2 \mathrm{var}\left(\sum_{t=1}^{n} H_m(X_t)\right). \end{aligned} \tag{4.36}$$

The Hermite rank of the function G^* is at least $m + 1$. Consequently, we have two scenarios. Either $\sum_k \gamma_X^m(k) < \infty$, and then both terms in (4.36) are of the order $O(n)$, or $\sum_k \gamma_X^m(k) = +\infty$, and then the second term dominates the first one. The latter happens if $m(1 - 2d) < 1$, and in this case the asymptotic behaviour of $\sum_{t=1}^{n} G(X_t)$ is the same as that of $(J(m)/m!) \sum_{t=1}^{n} H_m(X_t)$.

A proof of tightness is immediate. If we set

$$S'_{n,G}(u) := n^{-(m(d-1/2)+1)} L_m^{-m/2}(n) S_n(u),$$

we have

$$E\left[\left(S'_{n,G}(u) - S'_{n,G}(v)\right)^2\right] \sim |u - v|^{m(2d-1)+2}.$$

Since $m(1 - 2d) < 1$, the exponent is greater than one, and tightness follows from Lemma 4.3. □

In contrast, if the Hermite rank is large enough such that $m(1 - 2d) > 1$, then we have a weakly dependent-type behaviour of partial sums. The statement and proof of this result is postponed to the section on limit theorems for Appell polynomials.

Example 4.2 We illustrate the theoretical findings by a simulation example. First, we generate $n = 1000$ i.i.d. standard normal random variables X_t and plot the partial

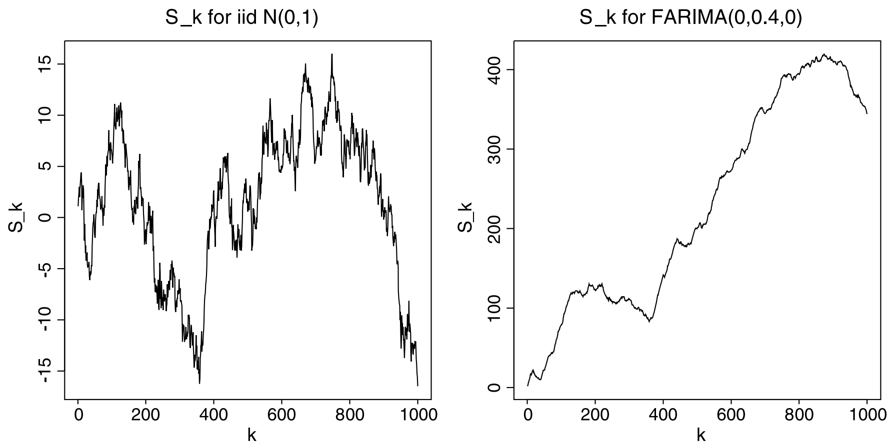

Fig. 4.1 Partial sum sequence $S_k = \sum_{t=1}^{k} X_t$ $(k = 1, \ldots, n)$ with X_t i.i.d. $N(0,1)$ (*left*) and X_t generated by a FARIMA(0, 0.4, 0) process (*right*)

sum sequence $S_k = \sum_{t=1}^{k} X_t$, $k = 1, \ldots, n$. This procedure is repeated for a Gaussian fractional ARIMA$(0, d, 0)$ process with parameter $d = 0.4$. The corresponding partial sum processes are plotted in Fig. 4.1. They can be considered approximations of a Brownian motion and a fractional Brownian motion with $H = 0.9$ respectively. Note that the path of the fractional Brownian motion is much smoother than the one of Brownian motion. This is due to long memory, which acts like a smoothing filter.

Example 4.3 In this example we generate $n = 1000$ random variables X_t from a Gaussian fractional ARIMA$(0, d, 0)$ process with parameter $d = 0.4$ and compute their sum. This procedure is repeated $N = 1000$ times. A normal probability plot of the $N = 1000$ sums $\sum_{t=1}^{n} X_t$ is displayed in the left panel of Fig. 4.2. The right panel shows a normal probability plot for the sums $\sum_{t=1}^{n} X_t^2$. The non-normal behaviour is clearly visible.

4.2.4 Linear Processes

In this section we consider a causal linear process

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad (t \in \mathbb{N}), \tag{4.37}$$

where, without loss of generality, $\sum_{j=0}^{\infty} a_j^2 = 1$, and ε_t $(t \in \mathbb{Z})$ are i.i.d. zero mean random variables with $\operatorname{var}(\varepsilon_1) = \sigma_\varepsilon^2 < \infty$. Thus, $\operatorname{var}(X_1) = \sigma_X^2 = \sigma_\varepsilon^2$. Note that Gaussian processes are included in this definition, but the class is much more general. Three different assumptions on the coefficients will be considered as $j \to \infty$ and with L_a denoting a slowly varying function at infinity:

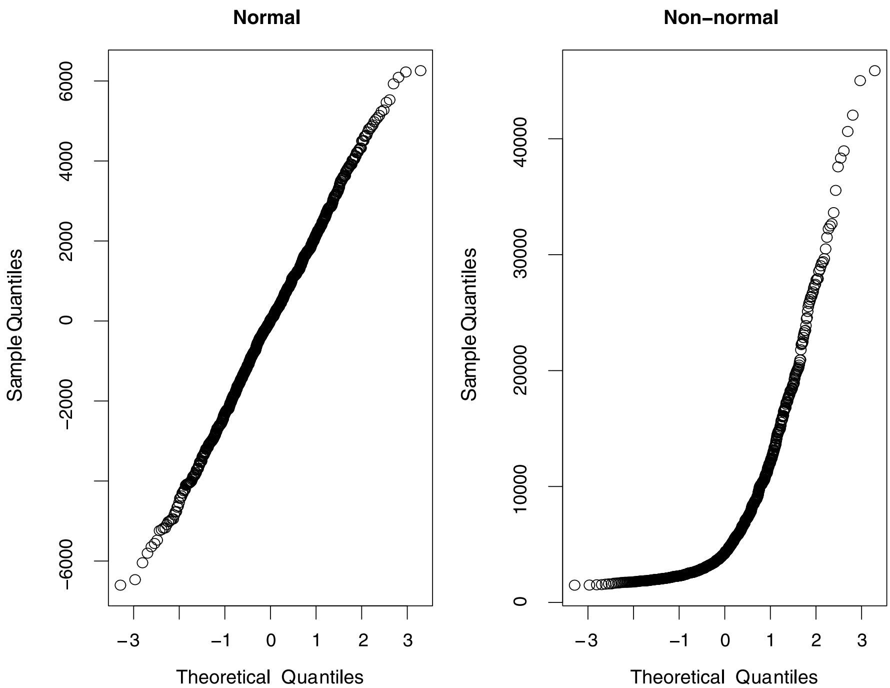

Fig. 4.2 Illustration of Theorem 4.3: normal probability plots of partial sums $\sum_{t=1}^{k} X_t$ (*left*) and $\sum_{t=1}^{k} X_t^2$, where X_t is generated by a FARIMA(0, 0.4, 0) process

- (B1) long memory:

$$a_j \sim L_a(j) j^{d-1} \quad \left(0 < d < \frac{1}{2}\right);$$

- (B2) short memory:

$$\sum_{j=0}^{\infty} |a_j| < \infty, \qquad \sum_{j=0}^{\infty} a_j \neq 0.$$

- (B3) antipersistence:

$$a_j \sim L_a(j) j^{d-1}$$

with $-\frac{1}{2} < d < 0$, and

$$\sum_{j=0}^{\infty} a_j = 0.$$

Under the short-memory assumption (B2), limiting behaviour is classical (see Theorem 4.5); see Brockwell and Davis (1991). Under long memory (B1), the first result was obtained in Davydov (1970a, 1970b); see also Gorodetskii (1977), Lang and Soulier (2000), Wang et al. (2003).

4.2.4.1 Asymptotic Covariances and Normalizing Constants

The behaviour of the autocovariance function γ_X and the spectral density f_X for the three cases can be characterized as follows. Combining Lemmas 4.13–4.15 with Lemmas 4.9–4.11, respectively, yields the asymptotic behaviour of $\operatorname{var}(S_n)$ (where $S_n(u)=\sum_{t=1}^{[nu]} X_t$, $S_n=S_n(1)$).

Lemma 4.13 *Under assumption* (B1), *we have, as* $\lambda\to 0$ *and* $k\to\infty$ *respectively,*

$$\begin{aligned} f_X(\lambda) &\sim L_f(\lambda)|\lambda|^{-2d}, \\ \gamma_X(k) &\sim L_\gamma(k)k^{2d-1}, \end{aligned} \tag{4.38}$$

where

$$L_\gamma(k)=L_a^2(k)\cdot\sigma_\varepsilon^2\int_0^\infty v^{d-1}(1+v)^{d-1}\,dv=\sigma_\varepsilon^2 L_a^2(k)B(1-2d,d), \tag{4.39}$$

$B(x,y)$ *denotes the Beta function, and* L_f *is obtained from* L_γ *by* (*cf.* (1.1))

$$L_f(\lambda)=L_\gamma\bigl(\lambda^{-1}\bigr)\pi^{-1}\Gamma(2d)\sin\biggl(\frac{\pi}{2}-\pi d\biggr). \tag{4.40}$$

Hence, via Lemma 4.9,

$$\operatorname{var}(S_n)\sim L_S(n)n^{2d+1}=\frac{1}{d(2d+1)}L_\gamma(n)n^{2d+1}. \tag{4.41}$$

Proof We have

$$\gamma_X(k)\sim\sigma_\varepsilon^2\sum_{j=1}^{\infty}L_a(j)L_a(j+k)j^{d-1}(j+k)^{d-1}=\sigma_\varepsilon^2 S_{\infty,k}\cdot k^{2d-1},$$

where

$$S_{\infty,k}=\lim_{n\to\infty}S_{n,k}$$

and

$$\begin{aligned} S_{n,k}&=\sum_{j=1}^{nk}L_a(j)L_a(j+k)\biggl(\frac{j}{k}\biggr)^{d-1}\biggl(\frac{j}{k}+1\biggr)^{d-1}n^{-1} \\ &=L_a^2(k)\sum_{j=1}^{nk}\frac{L_a(j)}{L_a(k)}\frac{L_a(j+k)}{L_a(k)}\biggl(\frac{j}{k}\biggr)^{d-1}\biggl(\frac{j}{k}+1\biggr)^{d-1}n^{-1} \\ &\underset{k\to\infty}{\sim}L_a^2(k)\int_0^n v^{d-1}(v+1)^{d-1}\,dv, \end{aligned}$$

where the last approximation is uniform in n. The approximation formula for f_X follows from Theorem 1.3. □

Example 4.4 (ARFIMA Model) Consider an ARFIMA$(0, d, 0)$ model, $d \in (0, 1/2)$. This process has the linear representation $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where

$$a_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \sim \frac{1}{\Gamma(d)} j^{d-1} \quad (j \to \infty).$$

Thus, $L_a \sim 1/\Gamma(d)$, so that

$$\gamma_X(k) \sim c_\gamma k^{2d-1}$$

with

$$\begin{aligned} c_\gamma &= \sigma_\varepsilon^2 \Gamma^{-2}(d) \int_0^\infty v^{d-1}(1+v)^{d-1}\, dv \\ &= \sigma_\varepsilon^2 \Gamma^{-2}(d) B(1-2d, d) = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)\Gamma(d)}{\Gamma^2(d)\Gamma(1-d)} \\ &= \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} = \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1-2d)\sin(\pi d). \end{aligned}$$

The last equality follows from $\Gamma(d)\Gamma(1-d) = \pi/\sin \pi d$. Moreover,

$$\begin{aligned} L_f(\lambda) &= \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1-2d)\sin(\pi d)\pi^{-1}\Gamma(2d)\sin\left(\frac{\pi}{2} - \pi d\right) \\ &= \frac{\sigma_\varepsilon^2}{\pi} \frac{\sin(\pi d)\sin(\frac{\pi}{2} - \pi d)}{\sin(2\pi d)} = \frac{\sigma_\varepsilon^2}{\pi} \frac{\sin(\pi d)\cos(\pi d)}{\sin(2\pi d)} \\ &= \frac{\sigma_\varepsilon^2}{\pi} \frac{\sin(\pi d)\cos(\pi d)}{2\sin(\pi d)\cos(\pi d)} = \frac{\sigma_\varepsilon^2}{2\pi}, \end{aligned}$$

so that

$$f_X(\lambda) \sim \frac{\sigma_\varepsilon^2}{2\pi} |\lambda|^{-2d}.$$

Lemma 4.14 *Under assumption* (B2), *we have*

$$\sum_{k=-\infty}^{\infty} \left|\gamma_X(k)\right| < \infty, \qquad \sum_{k=-\infty}^{\infty} \gamma_X(k) > 0.$$

If, in addition, $\sum_{j=0}^{\infty} j|a_j| < \infty$, *then* $f_X(\lambda)$ *is continuous on* $[-\pi, \pi]$.

Proof We have

$$\sum_{k=-\infty}^{\infty} \left|\gamma_X(k)\right| = \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \left|\sum_{j=0}^{\infty} a_j a_{j+|k|}\right| \le 2\sigma_\varepsilon^2 \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} |a_j||a_{j+|k|}|$$
$$= 2\sigma_\varepsilon^2 \left(\sum_{j=0}^{\infty} |a_j|\right)^2 < \infty.$$

Furthermore,

$$\sum_{k=-\infty}^{\infty} \gamma_X(k) = 2\pi f_X(0) = 2\pi \frac{\sigma_\varepsilon^2}{2\pi} \left|\sum_{j=0}^{\infty} a_j\right|^2 > 0.$$

To show that f_X is continuous, consider

$$\tilde{a}(\lambda) = \sum_{j=0}^{\infty} a_j e^{-ij\lambda}.$$

Since, as $x \to 0$, $\sin x \sim x$ and $\cos x - 1 \sim x^2/2$, we obtain for $\varepsilon < 1$,

$$\left|\tilde{a}(\lambda+\varepsilon) - \tilde{a}(\lambda)\right| \le \sum_{j=0}^{\infty} |a_j| \left|e^{-ij\lambda}\left(e^{-ij\varepsilon} - 1\right)\right|$$
$$\le 2\varepsilon \sum_{j=0}^{\infty} j|a_j|,$$

so that $\tilde{a}(\cdot)$ is continuous, and hence so is $f_X(\lambda) = \sigma_\varepsilon^2/(2\pi)\,|\tilde{a}(\lambda)|^2$. □

Lemma 4.15 *Under assumption* (B3), *we have, as* $\lambda \to 0$ *and* $k \to \infty$ *respectively,*

$$f_X(\lambda) \sim L_f(\lambda)|\lambda|^{-2d}, \tag{4.42}$$

$$\gamma_X(k) \sim L_\gamma(k) k^{2d-1}, \qquad \sum_{k=-\infty}^{\infty} \gamma_X(k) = 0, \tag{4.43}$$

where

$$L_\gamma(k) = L_a^2(k) \cdot \sigma_\varepsilon^2 \int_0^\infty v^{d-1}\left[1 - (v+1)^{d-1}\right] du$$
$$= \sigma_\varepsilon^2 L_a^2(k) B(1-2d, d),$$

and L_f *is obtained from* L_γ *by* (4.40).

Proof Similarly to the proof of Lemma 4.13,

$$\gamma_X(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j a_{j+k} = \sigma_\varepsilon^2 S_{\infty,k} \cdot k^{2d-1}$$

with $S_{\infty,k} = \lim_{n\to\infty} S_{n,k}$,

$$S_{n,k} = k^{1-2d} \sum_{j=0}^{nk} a_j a_{j+k} = S_{n,k}(1) + S_{n,k}(2)$$

and

$$S_{n,k}(1) = k^{1-2d} \sum_{j=0}^{nk} a_j (a_{j+k} - a_k) \sim L_a^2(n) \int_0^n v^{d-1}\big[(v+1)^d - 1\big]\, du,$$

$$S_{n,k}(2) = k^{1-2d} a_k \sum_{j=0}^{nk} a_j = -k^{1-2d} a_k \sum_{j=nk+1}^{\infty} a_j \sim L_a^2(n) \int_n^\infty v^{d-1}\, dv = o(n),$$

where the approximations are uniform in n. Moreover,

$$\sum_{k=-\infty}^{\infty} \gamma_X(k) = 2\pi f_X(0) = 2\pi \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{j=0}^{\infty} a_j \right|^2 = 0.$$

The approximation of f_X for $\lambda \to 0$ follows from Theorem 1.3. □

4.2.4.2 Asymptotic Distribution

Proofs of the next results illustrate different techniques that are applicable in various situations:

- Under short memory (B2), we apply the K-dependent approximation method, i.e. a combination of Proposition 4.1 and Lemma 4.1. This is easier than the cumulant method and does not require restrictive moment assumptions. It is particularly suited for linear processes (see Brockwell and Davis 1991).
- Under long memory (B1), we apply the method based on random spectral measures, as outlined in the alternative proof of Theorem 4.2; see Lang and Soulier (2000).

Theorem 4.5 *Assume that* X_t $(t \in \mathbb{N})$ *is a stationary linear process* (4.37) *such that* (B2) *holds. Then*

$$n^{-1/2} S_n = n^{-1/2} \sum_{t=1}^{n} X_t \to N\big(0, \nu^2\big),$$

where the variance $\nu^2 = \sigma_X^2 + 2\sum_{k=1}^{\infty} \gamma_X(k)$.

This theorem can be formulated in terms of functional convergence to Brownian motion.

Proof Let $X_{t,K} = \sum_{j=0}^{K} a_j \varepsilon_{t-j}$. Since the sequence $X_{t,K}$ ($t \in \mathbb{N}$) is K-dependent, an application of Lemma 4.1 yields

$$n^{-1/2} S_{n,K} = n^{-1/2} \sum_{t=1}^{n} X_{t,K} \xrightarrow{d} N\left(0, \nu_K^2\right)$$

with $\nu_K^2 = \mathrm{var}(X_{0,K}) + 2\sum_{k=0}^{K} \gamma_{X_K}(k)$, where

$$\gamma_{X_K}(k) = E[X_{t,K} X_{t+k,K}] = \sigma_\varepsilon^2 \sum_{j=0}^{K} a_j a_{j+k}.$$

Since $\nu_K \to \nu$ as $K \to \infty$, we conclude $N(0, \nu_K^2) \xrightarrow{d} N(0, \nu^2)$. It suffices to prove that for all $\delta > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\left(n^{-1/2}|S_n - S_{n,K}| > \delta\right) = 0.$$

The result of our theorem will then follow by Proposition 4.1. By Markov's inequality, it is sufficient to verify that

$$\lim_{K\to\infty} \lim_{n\to\infty} n^{-1} \mathrm{var}(S_n - S_{n,K}) = 0.$$

Let $\bar{X}_{t,K} = X_t - X_{t,K}$. Then

$$\lim_{n\to\infty} n^{-1} \mathrm{var}(S_n - S_{n,K}) = \lim_{n\to\infty} \sigma_\varepsilon^2 \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \sum_{j=K+1}^{\infty} a_j a_{j+k}$$
$$= \sigma_\varepsilon^2 \sum_{k=-\infty}^{\infty} \sum_{j=K+1}^{\infty} a_j a_{j+k} = \sigma_\varepsilon^2 \sum_{j=K+1}^{\infty} a_j \sum_{k=-\infty}^{\infty} a_{j+k}.$$

The $\lim_{n\to\infty}$ behaviour above is obtained by applying the dominated convergence theorem. For this, we need $\sum_k \sum_j |a_j a_{j+k}| < \infty$. This is true under the summability condition $\sum_{j=0}^{\infty} |a_j| < \infty$. Under this condition, we can also exchange the summations $\sum_k$ and $\sum_j$. Finally,

$$\lim_{K\to\infty} \lim_{n\to\infty} n^{-1} \mathrm{var}(S_n - S_{n,K}) \le \sum_{k=-\infty}^{\infty} |a_k| \lim_{m\to\infty} \sum_{j=K+1}^{\infty} |a_j| = 0.$$ □

Under (B1), the asymptotic behaviour of partial sums changes. This result was proven first in Davydov (1970a, 1970b). The method below is adapted from Lang and Soulier (2000), where the reader is referred to for details.

Theorem 4.6 *Assume that* X_t $(t \in \mathbb{N})$ *is a stationary linear process* (4.37) *such that the long-memory condition* (B1) *holds, i.e.* $a_j \sim L_a(j) j^{d-1}$, $d \in (0, \frac{1}{2})$. *Then*

$$n^{-(d+\frac{1}{2})} L_S^{-1/2}(n) S_n(u) = n^{-(d+\frac{1}{2})} L_S^{-1/2}(n) \sum_{t=1}^{[nu]} X_t \Rightarrow B_H(u) \quad \bigl(u \in [0,1]\bigr),$$

where $B_H(u)$ *is a standard fractional Brownian motion,* $H = d + \frac{1}{2}$, $\Rightarrow$ *denotes weak convergence in* $D[0,1]$, *and*

$$L_S(n) = \frac{1}{d(2d+1)} L_\gamma(n)$$

with L_γ *defined in* (4.39):

$$L_\gamma(k) = L_a^2(k)\sigma_\varepsilon^2 \int_0^\infty v^{d-1}(v+1)^{d-1}\, dv$$
$$= L_a^2(k)\sigma_\varepsilon^2 B(1-2d, d).$$

Proof We use the spectral method, as in the alternative proof of Theorem 4.2. Recall that any stationary sequence with finite variance can be written as

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{it\lambda} M_0(d\lambda), \quad t \in \mathbb{Z}.$$

The only difference between the spectral measure M_0 here and M_0 in the proof of Theorem 4.2 is that the measure here is not necessarily Gaussian. In particular, there is no guarantee that $n^{1/2} M_0(n^{-1}\cdot)$ and $M_0(\cdot)$ have the same distribution. Nevertheless, the same argument can be applied (see Lang and Soulier 2000). □

Example 4.5 (ARFIMA) Assume that X_t $(t \in \mathbb{N})$ is a FARIMA$(0, d, 0)$ model as in Example 4.4. Then

$$\gamma_X(k) \sim c_\gamma k^{2d-1},$$
$$c_\gamma = \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1-2d)\sin(\pi d).$$

Hence,

$$n^{-(d+\frac{1}{2})} L_S^{-1/2}(n) \sum_{t=1}^{[nu]} X_t \Rightarrow B_H(u)$$

and

$$L_S(n) = c_\gamma \frac{1}{d(2d+1)}.$$

Note that the innovations ε_t do not need to be Gaussian.

4.2.5 Subordinated Linear Processes

Next we consider the case where instead of the linear process X_t $(t \in \mathbb{N})$ a subordinated process, i.e. a transformation $Y_t = G(X_t)$ $(t \in \mathbb{N})$, is observed. Recall that in the Gaussian case asymptotic properties of partial sums of X_t and $H_m(X_t)$ (and, via the reduction principle of Theorem 4.4, of general functionals) can be studied using the spectral method. For linear processes, we applied again the spectral method in Theorem 4.6. However, this extension is not feasible for subordinated linear processes. In this setup, there are two common approaches: Appell polynomials (Surgailis 1982; Giraitis 1985; Giraitis and Surgailis 1986, 1989; Avram and Taqqu 1987; Surgailis and Vaičiulis 1999; Surgailis 2000; see also Surgailis 2003, for overview) and a martingale decomposition (Ho and Hsing 1996, 1997; Wu 2003; see also Hsing 2000 for an overview).

4.2.5.1 Normalizing Constants: Simple Example

Before we develop a general formula, let us consider the simple case of $G(X_t) = X_t^2$.

Example 4.6 Let X_t $(t \in \mathbb{N})$ be a linear process defined by (4.37). Assume that $E[\varepsilon_1^4] < \infty$ and that the long-memory condition (B1) holds. Using formula (4.38) for the covariance of X_t $(t \in \mathbb{N})$, we have

$$\gamma_X^2(k) \sim L_\gamma^2(k) k^{2(2d-1)}.$$

On the other hand,

$$\begin{aligned}\gamma_X^2(k) = cov^2(X_t, X_{t+k}) &= \left(\sum_{j=0}^{\infty} a_j a_{j+k}\right)^2 \\ &= \sum_{j=0}^{\infty} a_j^2 a_{j+k}^2 + \sum_{j,l=0;\ j\neq l}^{\infty} a_j a_l a_{j+k} a_{l+k}.\end{aligned}$$

Note that under (B1) the limiting behaviour of $\gamma_X^2(k)$ is determined by the second term. Now,

$$X_0^2 = \sum_{j=0}^{\infty} a_j^2 \varepsilon_{0-j}^2 + \sum_{j,l=0;\ j\neq l}^{\infty} a_j a_l \varepsilon_{0-j} \varepsilon_{0-l} =: X_{0,1} + X_{0,2}.$$

Analogously, we define $X_k^2 := X_{k,1} + X_{k,2}$. Note that $X_{0,1}$ and $X_{k,2}$ are uncorrelated. The same holds for $X_{0,2}$ and $X_{k,1}$. Furthermore,

$$cov(X_{0,1}, X_{k,1}) = E\left[\varepsilon_1^4\right] \sum_{j=0}^{\infty} a_j^2 a_{j+k}^2$$

and

$$cov(X_{0,2}, X_{k,2}) = 2 \sum_{j,l=0;\ j\neq l}^{\infty} a_j a_l a_{j+k} a_{l+k}.$$

Recalling that the second covariance is of a larger order than the first one, we conclude

$$\gamma_{X^2}(k) \sim 2 \sum_{j,l=0;\ j\neq l}^{\infty} a_j a_l a_{j+k} a_{l+k} \sim 2\gamma_X^2(k) \sim 2L_\gamma^2(k) k^{2(2d-1)}.$$

4.2.5.2 Normalizing Constants: Appell Polynomials

Now, we turn our attention to general nonlinear functionals. For a general non-normal distribution, in view of Sect. 3.3, a natural approach is to start with the Wick product $Y_t = A_m(X_t) = {:}X_t, \ldots, X_t{:}$ where A_m is the mth Appell polynomial associated with the marginal distribution of X_t. Suppose that $\gamma_X(k)$ is known, either exactly or its asymptotic behaviour. Can we give a simple formula for $\gamma_Y(k)$? In principle, the diagram formulas given in Theorem 3.10 provide an answer because

$$\kappa(Y_t, Y_{t+k}) = \left[\frac{\partial^2}{\partial z_1 \partial z_2} \log E\big[\exp(z_1 Y_t + z_2 Y_{t+k})\big]\right]_{z=0} = \gamma_Y(k).$$

To apply the diagram formula, consider a table W with two rows W_1, W_2 of length m. The positions in W_1 are associated with X_t and those in W_2 with X_{t+k}, i.e. we may write $W_1 = \{\tilde{X}_{(1,1)}, \ldots, \tilde{X}_{(1,m)}\}$ with $\tilde{X}_{(1,t)} = X_t$ and $W_2 = \{\tilde{X}_{(2,1)}, \ldots, \tilde{X}_{(2,m)}\}$ with $\tilde{X}_{(2,j)} = X_{j+k}$. Using the same notation as in Theorem 3.10, we obtain from (3.81)

$$\gamma_Y(k) = \kappa\big({:}X^{W_1}{:}, {:}X^{W_2}{:}\big) = \sum_{\gamma\in\Gamma_W^{\neq,c}} \kappa\big(X'^{V_1}\big)\cdots\kappa\big(X'^{V_r}\big). \tag{4.44}$$

Unfortunately, this is a rather complicated expression because in general $\kappa(X'^V)$ may not be zero for any subset V. There is one exception where (4.44) simplifies considerably, namely if X_t ($t \in \mathbb{N}$) is a Gaussian process. In this case, all cumulants $\kappa(X'^V)$ are zero except for normal edges, i.e. $\kappa(X'^V) = 0$ if $|V| \neq 2$, so that the sum in (4.44) is over $\Gamma_W^{\neq,c,\mathcal{N}}$, and, up to a constant, we obtain a sum of correlations to the power m, see Corollary 3.5.

Although (4.44) is complicated, it is possible to give simple asymptotic formulas for $\gamma_Y(k)$ and, consequently, the variance of $S_{n,A_m} = \sum_{t=1}^n A_m(X_t)$. A first simplification can be obtained in the representation of Appell polynomials of linear processes:

Lemma 4.16 *Let X_t ($t \in \mathbb{N}$) be a linear process (4.37) such that the Appell polynomials of its marginal distribution A_m ($m \in \mathbb{N}$) exist. Then*

$$A_m(X_t) = \sum_{k_1,\ldots,k_m=0}^{\infty} a_{k_1} \cdots a_{k_m} (:\varepsilon_{t-k_1} \cdots \varepsilon_{t-k_m}:). \tag{4.45}$$

Proof The result follows from

$$A_m(X_t) = :\underbrace{X_t, \ldots, X_t}_{m}:$$

and multilinearity of the Wick product. □

A direct consequence of this result is a simplified expression for S_n:

Corollary 4.1 *Let X_t ($t \in \mathbb{N}$) be a linear process defined by (4.37) such that the Appell polynomials of its marginal distribution A_m ($m \in \mathbb{N}$) exist. Let*

$$S_{n,A_m} = \sum_{t=1}^{n} A_m(X_t).$$

Then

$$S_{n,A_m} = \sum_{k_1,\ldots,k_m=0}^{\infty} a_{k_1} \cdots a_{k_m} \sum_{t=1}^{n} (:\varepsilon_{t-k_1} \cdots \varepsilon_{t-k_m}:)$$

with $a_k = 0$ for $k < 0$.

Furthermore, the diagram formula can be used to obtain an expression for the asymptotic autocovariance function of the subordinated sequence Y_t ($t \in \mathbb{N}$) under long memory:

Corollary 4.2 *Let X_t ($t \in \mathbb{N}$) be a linear process defined by (4.37) such that the Appell polynomials of its marginal distribution A_m ($m \in \mathbb{N}$) exist and the long-memory assumption* (B1) *holds. Then $Y_t = A_m(X_t)$ has an autocovariance function $\gamma_Y(k)$ with*

$$\begin{aligned}\gamma_Y(k) &\sim m!\gamma_X^m(k) \\ &\sim m!\left(L_a^2(k)\sigma_\varepsilon^2 \int_0^\infty v^{d-1}(v+1)^{d-1}\,dv\right)^m \cdot k^{(2d-1)m} \\ &= m!L_\gamma^m(k)k^{(2d-1)m}\end{aligned} \tag{4.46}$$

as $k \to \infty$, cf. (4.39).

Proof Here, only an outline of the extended proof in Giraitis and Surgailis (1989) and Surgailis and Vaičiulis (1999) is given. Lemma 4.16 and the multilinearity of cumulants imply

$$\begin{aligned} &cov\big(A_m(X_t), A_m(X_{t+k})\big)\\ &\quad = \kappa\big(A_m(X_t), A_m(X_{t+k})\big)\\ &\quad = \kappa\Bigg(\sum_{j_1,\ldots,j_m=0}^{\infty} a_{j_1}\cdots a_{j_m}(:\varepsilon_{t-j_1}\cdots\varepsilon_{t-j_m}:),\\ &\qquad\qquad \sum_{j_1,\ldots,j_m=0}^{\infty} a_{j_1}\cdots a_{j_m}(:\varepsilon_{t+k-j_1}\cdots\varepsilon_{t+k-j_m}:)\Bigg)\\ &\quad = \sum_{\substack{j_1,\ldots,j_m=0,\\ j_1',\ldots,j_m'=0}}^{\infty} a_{j_1}\cdots a_{j_m}a_{j_1'}\cdots a_{j_m'}\kappa(:\varepsilon_{t-j_1}\cdots\varepsilon_{t-j_m}:, :\varepsilon_{t+k-j_1'}\cdots\varepsilon_{t+k-j_m'}:). \end{aligned}$$

Now consider a table W with two rows $W_i = \{\varepsilon_{(i,1)},\ldots,\varepsilon_{(i,m)}\}$ $(i = 1, 2)$ with $\varepsilon_{(1,s)} = \varepsilon_{t_s}$ and $\varepsilon_{(2,s)} = \varepsilon_{t_s'}$. The diagram formula for cumulants of Wick products implies

$$\kappa(:\varepsilon_{t-j_1},\ldots,\varepsilon_{t-j_m}:, :\varepsilon_{t+k-j_1'},\ldots,\varepsilon_{t+k-j_m'}:) = \sum_{\gamma\in\Gamma_W^{\neq,c}} \kappa\big(\varepsilon'^{V_1}\big)\cdots\kappa\big(\varepsilon'^{V_r}\big).$$

Using this equation, we have

$$\kappa\big(A_m(X_t), A_m(X_{t+k})\big) = r_{\text{main}} + r_k,$$

where

$$r_{\text{main}} = \sum_{\gamma\in\Gamma_W^{\neq,c,\mathcal{N}}}\ \sum_{\substack{j_1,\ldots,j_m=0\\ j_1',\ldots,j_m'=0}} \left(\prod_{i=1}^m a_{j_i}a_{j_i'}\right)\kappa\big(\varepsilon'^{V_1}\big)\cdots\kappa\big(\varepsilon'^{V_r}\big)$$

and

$$r_k = \sum_{\gamma\in\Gamma_W^{\neq,c}\setminus\Gamma_W^{\neq,c,\mathcal{N}}}\ \sum_{\substack{j_1,\ldots,j_m=0\\ j_1',\ldots,j_m'=0}} \left(\prod_{i=1}^m a_{j_i}a_{j_i'}\right)\kappa\big(\varepsilon'^{V_1}\big)\cdots\kappa\big(\varepsilon'^{V_r}\big).$$

It can be shown that, as $k\to\infty$, $r_k = o(k^{(2d-1)m})$, so that only diagrams in $\Gamma_W^{\neq,c,\mathcal{N}}$ matter asymptotically. For instance, for $\gamma = \bigcup_{i=1}^{m-1} V_i$ with $V_i = \{(1,i),(2,i)\}$ $(i = 1,\ldots,m-2)$ and $V_{m-1} = \{(1,m-1),(2,m-1),(1,m),(2,m)\}$, we have, because of independence of the random variables ε_i,

$$\kappa\left(\varepsilon'^{V_1}\right)\cdots\kappa\left(\varepsilon'^{V_{m-1}}\right)=0,$$

unless $j_1'=j_1+k,\ldots,j_{m-1}'=j_{m-1}+k$ and $j_{m-1}=j_m$, $j_{m-1}'=j_m'=j_{m-1}+k$. Thus, the contribution of γ to r_m is

$$\sigma_\varepsilon^2\left(\sum_{j=0}^{\infty}a_ja_{j+k}\right)^{m-2}\sum_{j=0}^{\infty}a_j^2a_{j+k}^2\sim\gamma_X^{m-2}(k)L(k)k^{4d-3}=o\left(k^{(2d-1)m}\right).$$

For κ_{main}, the calculation simplifies considerably because each $\gamma\in\Gamma_W^{\neq,c,\mathcal{N}}$ consists of edges $V_j=\{(1,j),(1,\pi(j))\}$ $(j=1,2,\ldots,m)$ where π is a permutation of $\{1,2,\ldots,m\}$. Thus, the number of diagrams in $\Gamma_W^{\neq,c,\mathcal{N}}$ is $|\Gamma_W^{\neq,c,\mathcal{N}}|=m!$. Moreover, for each permutation π,

$$\sum_{\substack{j_1,\ldots,j_m=0\\ j_1',\ldots,j_m'=0}}\left(\prod_{i=1}^{m}a_{j_i}a_{j_i'}\right)\kappa\left(\varepsilon'^{V_1}\right)\cdots\kappa\left(\varepsilon'^{V_r}\right)=\sigma_\varepsilon^{2m}\left(\sum_{j=0}^{\infty}a_ja_{j+k}\right)^m=\gamma_X^m(k).$$

Thus, taking the sum over all $m!$ permutations, we have

$$r_{\text{main}}=m!\gamma_X^m(k).$$ □

Note that, if X_t $(t\in\mathbb{N})$ is a Gaussian process, then we have the exact relationship $\gamma_{A_m}(k)=m!\gamma_X^m(k)$ for any finite k because all cumulants above order 2 are zero, so that all contributions except those from $\Gamma_W^{\neq,c,\mathcal{N}}$ are zero. (cf. Sect. 4.2.3).

The combination of Lemma 4.9 and formula (4.38) yields an asymptotic formula for the variance of $S_{A_m,n}=\sum_{t=1}^n A_m(X_t)$ under the assumption of long memory (see Giraitis and Surgailis 1989; Surgailis and Vaičiulis 1999):

Theorem 4.7 *Let X_t $(t\in\mathbb{N})$ be a linear process defined by* (4.37) *such that the Appell polynomials A_m $(m\in\mathbb{N})$ of its marginal distribution exist and the long-memory assumption* (B1) *holds. Assume further that $m(1-2d)<1$. Then, as $n\to\infty$,*

$$\operatorname{var}(S_{n,A_m})=\operatorname{var}\left(\sum_{t=1}^{n}A_m(X_t)\right)\sim L_m(n)n^{(2d-1)m+2}$$

with

$$\begin{aligned}L_m(n)&=m!C_mL_\gamma^m(n),\\ C_m&=\frac{2}{((2d-1)m+1)((2d-1)m+2)}\end{aligned}\tag{4.47}$$

and L_γ given by (4.39). *On the other hand, if $m(1-2d)>1$, then*

$$\operatorname{var}(S_{n,A_m})=O(n).$$

We recognize the same formula as in the Gaussian case, see (4.20). Furthermore, note that, in general, antipersistence is not inherited because the condition that autocovariances add up to zero is destroyed much more easily than nonsummability.

4.2.5.3 Asymptotic Distributions: Appell Polynomials

In the previous sections we obtained asymptotic expressions for the autocovariance function $\gamma_{A_m}(k) = cov(A_m(X_t), A_m(X_{t+k}))$ and the variance $v_n^2 := \text{var}(S_{n,A_m})$. The remaining question is which processes one obtains as limits of $S_{n,A_m}(t)/v_n$. It turns out that, under suitable moment conditions, the only possible limiting processes are Hermite–Rosenblatt processes. In fact this question has been answered in the Gaussian case, see Theorem 4.4.

Theorem 4.8 *Let X_t $(t \in \mathbb{N})$ be a linear process defined by* (4.37) *such that the Appell polynomials A_m $(m \in \mathbb{N})$ of its marginal distribution exist and the long-memory assumption* (B1) *holds, i.e. $a_j \sim L_a(j)j^{d-1}$, $d \in (0, 1/2)$. Let*

$$S_{n,A_m}(u) = \sum_{t=1}^{[nu]} A_m(X_t) \quad \bigl(u \in [0, 1]\bigr)$$

and assume that $E(\varepsilon_1^{2j}) < \infty$ for all j. Then, if $m(1 - 2d) < 1$,

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) S_{n,A_m}(u) \Rightarrow Z_{m,H}(u) \quad \bigl(u \in [0, 1]\bigr), \tag{4.48}$$

where $Z_{m,H}(\cdot)$ is the Hermite–Rosenblatt process with $H = d + \frac{1}{2}$, $\Rightarrow$ denotes weak convergence in $D[0, 1]$, and L_m is given in (4.47):

$$L_m(n) = m! C_m L_\gamma^m(n),$$

$$C_m = \frac{2}{((2d-1)m+1)((2d-1)m+2)},$$

with L_γ given by (4.39):

$$L_\gamma(k) = L_a^2(k) \cdot \sigma_\varepsilon^2 \int_0^\infty v^{d-1}(v+1)^{d-1}\, dv.$$

On the other hand, if $m(1 - 2d) > 1$, then $\text{var}(S_{n,A_m}) \sim \sigma_S n$ for some $\sigma_S > 0$, and

$$n^{-\frac{1}{2}} S_{n,A_m}(u) \Rightarrow \sigma_S B(u) \quad \bigl(u \in [0, 1]\bigr), \tag{4.49}$$

where $B(\cdot)$ is a standard Brownian motion, and $\Rightarrow$ denotes weak convergence in $D[0, 1]$.

In other words, the asymptotic distribution is the same as in case of Hermite polynomials. Moreover, L_m agrees with L_m in Theorem 4.3.

Proof At first consider the case with $m(1-2d)>1$. The proof is rather long, so that only a sketch is given here (for details, see e.g. Surgailis 2003). To prove the convergence of finite-dimensional distributions, we use the cumulant method (cf. Theorem 4.1). Recall that for the normal distribution, all cumulants of order $j\ge 3$ equal zero, and there is no other distribution with this property. It is therefore sufficient to show that for $j\ge 3$,

$$\lim_{n\to\infty}\kappa_j\left(n^{-\frac{1}{2}}S_{n,A_m}(t)\right)=n^{-\frac{j}{2}}\lim_{n\to\infty}\kappa\big(\underbrace{S_{n,A_m}(t),\ldots,S_{n,A_m}(t)}_{j}\big)=0.$$

Without loss of generality, we may fix t at $t=1$, and we write $S_{n,A_m}=S_{n,A_m}(1)$. Now for $s_1,\ldots,s_j\in\mathbb{N}$, consider a table W with rows

$$W_r=\{X_{(r,1)}=X_{s_r},\ldots,X_{(r,j)}=X_{s_r}\}\quad(1\le r\le j).$$

Then, because of multilinearity of κ,

$$\begin{aligned}\kappa(S_{n,A_m},\ldots,S_{n,A_m})&=\sum_{s_1,\ldots,s_j=1}^{n}\kappa\left(A_m(X_{s_1}),\ldots,A_m(X_{s_j})\right)\\&=\sum_{s_1,\ldots,s_j=1}^{n}\kappa\left(:X^{W_1}:,\ldots,:X^{W_j}:\right).\end{aligned}$$

The diagram formula implies

$$\kappa\left(:X^{W_1}:,\ldots,:X^{W_j}:\right)=\sum_{\gamma\in\Gamma_W^{\neq,c}}\kappa\left(X'^{V_1}\right)\cdots\kappa\left(X'^{V_r}\right),$$

and hence,

$$\begin{aligned}\kappa_j\left(n^{-\frac{1}{2}}S_{n,A_m}(t)\right)&=\sum_{\gamma\in\Gamma_W^{\neq,c}}n^{-\frac{j}{2}}\sum_{s_1,\ldots,s_j=1}^{n}\kappa\left(X'^{V_1}\right)\cdots\kappa\left(X'^{V_r}\right)\\&=\sum_{\gamma\in\Gamma_W^{\neq,c}}n^{-\frac{j}{2}}J_{n,\gamma}.\end{aligned}$$

Since the number of diagrams in $\Gamma_W^{\neq,c}$ is finite and does not depend on n, it is sufficient to show that $n^{-\frac{j}{2}}J_{n,\gamma}$ converges to zero. Note first that, for any $s_1,\ldots,s_j$ and $V\subseteq W$,

$$\kappa\left(X'^{V}\right)=\kappa\big(\underbrace{X_{s_1},\ldots,X_{s_1}}_{|V\cap W_1|\text{-times}},\ldots,\underbrace{X_{s_j},\ldots,X_{s_j}}_{|V\cap W_j|\text{-times}}\big).$$

Since X_t $(t \in \mathbb{N})$ is a linear process with i.i.d. innovations ε_j $(t \in \mathbb{Z})$, this can be written as

$$\kappa\left(X'^{V}\right) = \text{const} \cdot B_{V,s_1,\ldots,s_j},$$

where

$$B_{V,s_1,\ldots,s_j} = \sum_{i=-\infty}^{\infty} a_{i+s_1}^{|V\cap W_1|} \cdots a_{i+s_j}^{|V\cap W_j|}.$$

Hence,

$$\kappa\left(X'^{V_1}\right)\cdots\kappa\left(X'^{V_r}\right) = \text{const} \cdot \prod_{u=1}^{r} B_{V_u,s_1,\ldots,s_j},$$

so that it is sufficient to show that each $n^{-\frac{j}{2}} B_{V_u,s_1,\ldots,s_j}$ converges to zero. This requires a rather laborious detailed argument. However, the essential idea used in Surgailis (2003, Lemma 6.1) is to show this first for a finite moving average process $X_{t,K} = \sum_{j=0}^{K} a_j \varepsilon_{t-j}$ (actually Surgailis allows for a two-sided moving average) and then give an upper bound for the difference between the approximation $J_{n,\gamma}^{K}$ and $J_{n,\gamma}$ that converges to zero as K tends to infinity. Note that a similar approximation argument was used to establish convergence of partial sums of weakly dependent linear processes, see Theorem 4.5.

Tightness is easier than fidi-convergence but is omitted here; we refer the reader to Giraitis (1985).

Next, consider the case $m(1-2d) < 1$. This case has been considered for instance in Surgailis (1981, 1982), Giraitis and Surgailis (1986, 1989) and Avram and Taqqu (1987); see also Surgailis (2003) for an overview.

Recall from Corollary 4.1 that

$$S_{n,A_m} = \sum_{t=1}^{n} \sum_{j_1,\ldots,j_m=0}^{\infty} a_{j_1}\cdots a_{j_m} (:\varepsilon_{t-j_1}\cdots\varepsilon_{t-j_m}:).$$

Consider

$$U_{n,m} := m! \sum_{t=1}^{n} \sum_{0=j_1<j_2<\cdots<j_m}^{\infty} a_{j_1}\cdots a_{j_m} (:\varepsilon_{t-j_1}\cdots\varepsilon_{t-j_m}:). \tag{4.50}$$

Since the random variables $\varepsilon_{j_1}\cdots\varepsilon_{j_m}$ in this expression are independent, we have

$$:\varepsilon_{j_1}\cdots\varepsilon_{j_m}: = A_1(\varepsilon_{j_1})\cdots A_1(\varepsilon_{j_m}) = \varepsilon_{j_1}\cdots\varepsilon_{j_m}.$$

Therefore, we may write

$$U_{n,m} = m! \sum_{t=1}^{n} \sum_{0=j_1<j_2<\cdots<j_m}^{\infty} \prod_{s=1}^{m} a_{j_s}\varepsilon_{t-j_s} =: m! \sum_{t=1}^{n} V_{t,m}. \tag{4.51}$$

If we recall now (cf. proof of Theorem 4.6) that

$$\varepsilon_t = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{it\lambda} M_0(d\lambda),$$

where M_0 is a spectral measure with independent increments, then combining argument from the proof of Theorem 4.3 with the proof of Theorem 4.6, we expect that

$$n^{-(1-m(\frac{1}{2}-d))} L_f^{-m/2}(n^{-1}) U_{n,m} \xrightarrow{d} m! \int_{\lambda_1 < \cdots < \lambda_m} D(\lambda_1 + \cdots + \lambda_m) dW_X(\lambda_1) \cdots dW_X(\lambda_m), \quad (4.52)$$

where $dW_X(\lambda) = |\lambda|^{-d} dM_0(\lambda)$ is the limiting spectral measure defined in (4.34). The spectral-domain function L_f is replaced by the time-domain slowly varying function L_m using the same argument as in the proof of Theorem 4.3:

$$L_m(n) = m! C_m \big(2\Gamma(1-2d)\sin(\pi d)\big)^m L_f^m(n^{-1}).$$

Then,

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) U_{n,m} \xrightarrow{\mathrm{d}} Z_{m,H}(1). \quad (4.53)$$

Finally,

$$S_{n,A_m} = U_{n,m} + r_{n,m},$$

where the remainder $r_{n,m}$ involves summation over $j_1, \ldots, j_m$ such that at least two indices agree. The remainder is of a smaller order (see Avram and Taqqu 1987 for details).

Tightness is very easy. We use the same argument as in the proof of Theorem 4.4, together with the variance estimates in Theorem 4.7. □

As noted in the proof, in the case with $m(1-2d) < 1$, the convergence of S_{n,A_m} is determined by the term $U_{n,m}$ defined in (4.51). In fact, the convergence equation (4.52) will play a crucial role in some of the results following below.

The assumptions of the theorem can be relaxed in various ways. For instance, in order to obtain the usual central limit theorem in (4.49), only $\sum |\gamma_X(k)|^m < \infty$ is required instead of the specific decay of γ_X (see Surgailis 2003). Moreover, the result can be extended to

$$S_{n,G}(u) = \sum_{t=1}^{[nu]} G(X_t)$$

with

$$G(x) = \sum_{j=m}^{\infty} \frac{a_{\mathrm{app},j}}{j!} A_j(x).$$

Assuming that $a_{\mathrm{app},m} \neq 0$ (i.e. G has Appell rank m), the contribution of $a_{\mathrm{app},m} \times A_m(X_t)/m!$ dominates, provided that $m(1-2d) < 1$. For example, Surgailis (2000) considers arbitrary polynomials G. Furthermore, Surgailis and Vaičiulis (1999) replace independent ε_t ($t \in \mathbb{Z}$) by martingale differences, and Surgailis (2000) considers $\tilde{X}_t = X_t + V_t$ where V_t ($t \in \mathbb{N}$) is a stationary short-memory process.

In view of the fact that for each distribution different Appell polynomials are obtained, and in general they are not orthogonal, it is quite remarkable that the same asymptotic limit is obtained as under Gaussian subordination and Hermite polynomials. Moreover, it is worth noting that, for fixed m, the condition $m(1-2d) < 1$ means that $d > \frac{1}{2}(1-m^{-1})$. Thus, a nonstandard limiting behaviour (which is also called noncentral limit theorem) is achieved for sufficiently strong long-range dependence. The higher the degree m of the Appell polynomial, the stronger dependence has to be to satisfy the condition. This is essentially due to (4.46). Since at the same time d does not exceed $\frac{1}{2}$, there is no such d for $m = 1$. In other words, for X_t ($t \in \mathbb{N}$), a noncentral limit theorem holds for all $0 < d < \frac{1}{2}$.

4.2.5.4 Asymptotic Distributions: Martingale Approach and Power Ranks

Recall now that the jth Appell coefficient can be obtained either by

$$a_{\mathrm{app},j} = E\left[G^{(j)}(X)\right] \tag{4.54}$$

if the jth derivative of G exists and its expected value is not zero (see (3.66)) or by

$$a_{\mathrm{app},j} = (-1)^j \int G(x) p_X^{(j)}(x)\,dx \tag{4.55}$$

(see (3.69)), where $p_X = F_X'$ is the density of X. Note that due to (4.54), a similar definition of Appell rank that has been proposed in the literature is the so-called power rank.

Definition 4.1 Let X be a random variable. The power rank of a function G (with respect to X) is the smallest integer $m \geq 1$ such that $G_\infty^{(m)}(x) \neq 0$, where $G_\infty(x) = E[G(X+x)]$.

Example 4.7 Let F_X be the distribution of a random variable X with $E(X) = 0$. If $G(x) = x^2 - E(X^2)$, then $G_\infty^{(1)}(0) = 2\int u\,dF_X(u) = 2E(X) = 0$. Furthermore, $G_\infty^{(2)}(0) = 2\int dF_X(u) = 2$. This implies that for a centred linear process $X_t = \sum a_j \varepsilon_{t-j}$, the power rank of the quadratic function is always 2, regardless of the distribution of ε_t (and the marginal distribution of X_t).

Using the power rank, Ho and Hsing (1996, 1997) developed a different approach to studying limit theorems for functionals of linear processes. To describe the idea, let us again consider

$$X_{t,K} = \sum_{j=0}^{K} a_j \varepsilon_{t-j},$$

$$\tilde{X}_{t,K} = X_t - X_{t,K} = \sum_{j=K+1}^{\infty} a_j \varepsilon_{t-j}$$

and

$$G_K(y) := E\big[G(X_{t,K} + y)\big] \quad (K \geq 0), \qquad G_\infty(y) = E\big[G(X_t + y)\big]. \tag{4.56}$$

We also use the convention $G_{-1} = G$ and $\tilde{X}_{0,-1} = X_0$. Note now, that if $\mathcal{F}$ is a sigma field, ξ_A is a random variable that is $\mathcal{F}$-measurable and ξ_B is a random variable that is independent of $\mathcal{F}$ and has distribution F_B, then

$$E\big[G(\xi_A + \xi_B + y)|\mathcal{F}\big] = \int G(\xi_A + v + y)\, dF_B(v) =: G_{B,*}(\xi_A + y) \tag{4.57}$$

and

$$G_*(y) := E\big[G(\xi_A + \xi_B + y)\big] = E\big[G_{B,*}(\xi_A + y)\big]. \tag{4.58}$$

Now let $\mathcal{F}_K = \sigma(\varepsilon_j, -\infty < j \leq K)$ $(K \in \mathbb{Z})$. We apply (4.57) and (4.58) with $(\xi_A, \xi_B, \mathcal{F}) = (\tilde{X}_{t,K-1}, X_{t,K-1}, \mathcal{F}_{t-K})$ and $(\xi_A, \xi_B, \mathcal{F}) = (\tilde{X}_{t,K}, X_{t,K}, \mathcal{F}_{t-(K+1)})$ respectively. We obtain

$$\begin{aligned}
&\sum_{t=1}^{n} \big\{G(X_t) - E\big[G(X_1)\big]\big\} \\
&\quad = \sum_{t=1}^{n} \sum_{K=0}^{\infty} \big\{E\big[G(X_t)|\mathcal{F}_{t-K}\big] - E\big[G(X_t)|\mathcal{F}_{t-(K+1)}\big]\big\} \\
&\quad = \sum_{t=1}^{n} \sum_{K=0}^{\infty} \big(G_{K-1}(\tilde{X}_{t,K-1}) - G_K(\tilde{X}_{t,K})\big) \\
&\quad \approx \sum_{t=1}^{n} \sum_{K=0}^{\infty} \big(G_K(\tilde{X}_{t,K-1}) - G_K(\tilde{X}_{t,K})\big) \\
&\quad \approx \sum_{t=1}^{n} \sum_{K=0}^{\infty} a_t \varepsilon_{t-K} G_K^{(1)}(\tilde{X}_{t,K}) && (4.59) \\
&\quad \approx G_\infty^{(1)}(0) \sum_{t=1}^{n} X_t + \sum_{t=1}^{n} \sum_{K=0}^{\infty} a_K \varepsilon_{t-K} \big(G_K^{(1)}(\tilde{X}_{t,K}) - G_\infty^{(1)}(0)\big). && (4.60)
\end{aligned}$$

The point of this approximation is that the first term in the last expression is just the partial sum of the linear sequence, multiplied by a constant. The first term is of a larger order than the second term. Consequently, using Theorem 4.6, we expect

$$n^{-(d+\frac{1}{2})}L_S^{-1/2}(n)\sum_{t=1}^{n}\big\{G(X_t)-E\big[G(X_1)\big]\big\}\overset{d}{\to}G_\infty^{(1)}(0)B_H(1).$$

This is useful, of course, only if $G_\infty^{(1)}(0)$, the first power rank of G, does not vanish. If $G_\infty^{(1)}(0)=0$, then the expansion is continued until we obtain a non-vanishing quantity $G_\infty^{(m)}(0)$. In that case we say that the power rank of G is m. If for example the power rank is 2, the expansion reads further

$$\begin{aligned}
&\sum_{j=1}^{n}\big\{G(X_t)-E\big[G(X)\big]\big\}\\
&\quad=\sum_{t=1}^{n}\sum_{K=0}^{\infty}\big\{E\big[G(X_t)|\mathcal{F}_{t-K}\big]-E\big[G(X_t)|\mathcal{F}_{t-(K+1)}\big]\big\}\\
&\quad\approx G_\infty^{(2)}(0)\sum_{t=1}^{n}\sum_{j_1=0}^{\infty}\sum_{j_2=j_1+1}^{\infty}a_{j_1}a_{j_2}\varepsilon_{t-j_1}\varepsilon_{t-j_2}\\
&\qquad+\sum_{t=1}^{n}\sum_{j_1=0}^{\infty}\sum_{j_2=j_1+1}^{\infty}a_{j_1}a_{j_2}\varepsilon_{t-j_1}\varepsilon_{t-j_2}\big(G_{j_2}^{(2)}(\tilde{X}_{t,j_2})-G_\infty^{(2)}(0)\big).
\end{aligned}$$

As before, the second term in the last expression is of a smaller order than the first one. We recognize the first term as $G_\infty^{(2)}(0)U_{n,2}/2!$ (cf. (4.51)). Therefore, using the convergence result (4.52), we have

$$n^{-2d}L_2^{-1/2}(n)\sum_{j=1}^{n}\big\{G(X_t)-E\big[G(X_1)\big]\big\}\Rightarrow G_\infty^{(2)}(0)Z_{2,H}(1)/2!.$$

This can be generalized to arbitrary power ranks. There are a lot of technical details missing in the heuristic explanation above. We make it more precise, using a modified version of Ho and Hsing's approach (see Wu 2003). In order to do this, let G be a function, and $p\in\mathbb{N}$. Define (cf. (4.51))

$$T_n(G;p)=\sum_{t=1}^{n}\left\{G(X_t)-E\big[G(X_1)\big]-\sum_{r=1}^{p}G_\infty^{(r)}(0)V_{t,r}\right\},$$

where

$$V_{t,r}=\sum_{0\le j_1<\cdots<j_r}\prod_{s=1}^{r}a_{j_s}\varepsilon_{t-j_s}.$$

In particular,

$$T_n(G;1)=\sum_{j=1}^{n}\left\{G(X_t)-E\left[G(X_1)\right]-G_\infty^{(1)}(0)X_t\right\}.$$

For any random variable Y, let $\|Y\|_r = E^{1/r}[Y^r]$. The following theorem establishes a reduction principle for $T_n(G;p)$ that can be viewed as a counterpart to the Gaussian case (see the proof of Theorem 4.4). We state the result assuming that the slowly varying function L_a in (B1) is constant. The statement can be modified appropriately to incorporate a general slowly varying function $L_a(j)$.

Theorem 4.9 *Let X_t $(t\in\mathbb{N})$ be a linear process defined by* (4.37) *with coefficients satisfying assumption* (B1) *with $L_a(j)\equiv 1$. Assume that $E[|\varepsilon|^{4+\gamma}]<\infty$ for some $\gamma>0$ and*

$$\max_{r=1,2,\ldots,p+1}\sup_y\left|G_\infty^{(r)}(y)\right|<\infty,\tag{4.61}$$

where G_∞ is defined in (4.56).

- *If $(p+1)(1-2d)>1$, then $\|T_n(G;p)\|_2^2=O(n)$.*
- *If $(p+1)(1-2d)<1$, then*

$$\left\|T_n(G;p)\right\|_2^2=O\left(n^{2-(p+1)(1-2d)}\right).\tag{4.62}$$

The proof of this result is postponed to the end of this section. At this moment, let us discuss its consequences and technical assumptions. Assumption (4.61) is in the spirit of Ho and Hsing (1997). Another assumption was considered in Wu (2003). Similarly to definition (4.56), one can argue that

$$G_K^{(r)}(y):=\frac{d}{dy^r}E\left[G(X_{0,K}+y)\right]=E\left[G^{(r)}(X_{0,K}+y)\right]\quad(K\geq 0),$$
$$G_\infty^{(r)}(y)=E\left[G^{(r)}(X+y)\right].$$

For example,

$$\begin{aligned}&\frac{E[G(X+y+\delta)]-G(X+y)}{\delta}-E\left[G^{(1)}(X+y)\right]\\&=\int\left\{\frac{G(x+y+\delta)-G(x+y)}{\delta}-G^{(1)}(x+y)\right\}p_X(x)\,dx\\&\leq\delta\sup_u\left|G^{(2)}(u)\right|\int p_X(x)\,dx.\end{aligned}$$

Hence, for instance, if G has uniformly bounded second-order derivatives, then the limit as $\delta\to 0$ exists. However, such a strong assumption is not needed in fact, and a condition like (4.61) suffices (see Ho and Hsing 1996, Lemma 6.2, Wu 2003). We may thus write $G_0^{(r)}(y) = E[G^{(r)}(a_0\varepsilon_0 + y)]$ and

$$G_1^{(r)}(y) = E\big[G^{(r)}(a_0\varepsilon_0 + a_1\varepsilon_{-1} + y)\big] = E\big\{E\big[G^{(r)}(a_0\varepsilon_0 + a_1\varepsilon_{-1} + y)|\varepsilon_{-1}\big]\big\}$$
$$= E\big[G_0^{(r)}(a_1\varepsilon_{-1} + y)\big].$$

Therefore, it is intuitively clear that properties of $G_0^{(r)}$ are transferred to $G_1^{(r)}$ and by induction to any of $G_K^{(r)}$, $K \ge 1$.

Example 4.8 Consider $G(u) = 1\{u \le x_0\}$ for a fixed x_0. Then $G_\infty(y) = E[1\{X + y \le x_0\}] = P(X \le x_0 - y)$, and

$$G_\infty^{(1)}(0) = \frac{d}{dy} P(X \le x_0 - y)|_{y=0} = -p_X(x_0 - y)|_{y=0} = -p_X(x_0),$$

where p_X is the density of X.

What is the consequence of the theorem above? Take $p = 1$. We obtain $\|T_n(G;1)\|_2^2 = O(\max\{n, n^{4d}\})$. Recall now Theorem 4.6 that describes convergence of partial sums $\sum_{t=1}^n X_t$. We conclude that the limiting behaviour of

$$n^{-(\frac{1}{2}+d)} L_1^{-1/2}(n) \sum_{t=1}^n \big\{G(X_t) - E\big(G(X_1)\big)\big\}$$

is the same as that of

$$n^{-(\frac{1}{2}+d)} L_1^{-1/2}(n) G_\infty^{(1)}(0) \sum_{t=1}^n X_t,$$

where $L_1(n) = (d(2d+1))^{-1} L_\gamma(n)$, and $L_\gamma(n)$ given in (4.39). If the power rank is greater than one, then one has to apply a higher-order expansion ($p \ge 2$). The limiting behaviour of the partial sum follows from the corresponding limit theorem for $U_{n,p}$. The latter was considered in (4.51) and (4.52).

Corollary 4.3 *Let* $X_t = \sum_{j=0}^\infty a_j \varepsilon_{t-j}$ $(t \in \mathbb{Z})$ *be a linear process defined by* (4.37) *with coefficients satisfying assumption* (B1), *i.e.* $a_j \sim L_a(j) j^{d-1}$, $d \in (0, 1/2)$. *Assume that G has the power rank m. If* $m(1-2d) < 1$, *then, under the conditions of Theorem* 4.9,

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) \sum_{t=1}^n \big\{G(X_t) - E\big(G(X_1)\big)\big\} \xrightarrow{d} G_\infty^{(m)}(0) Z_{m,H}(1),$$

where

$$L_m(n) = m! C_m L_\gamma^m(n),$$

$$C_m = \frac{2}{((2d-1)m+1)((2d-1)m+2)},$$

and L_γ *is given by* (4.39):

$$L_\gamma(k) = L_a^2(k)\sigma_\varepsilon^2 \int_0^\infty v^{d-1}(v+1)^{d-1}\,dv$$
$$= L_a^2(k)\sigma_\varepsilon^2 B(1-2d,d).$$

Let us apply Corollary 4.3 to X_t^2, where X_t is a linear process such that $E(X_1^2)=1$. The example shows that in a sense, the power rank method is distribution free. In contrast, limiting results for Appell polynomials are not directly applicable to X_t^2-1, unless X_t are Gaussian.

Example 4.9 Consider a linear process $X_t = \sum_{j=0}^\infty a_j\varepsilon_{t-j}$ $(t\in\mathbb{Z})$ such that $\sum_{k=0}^\infty a_k^2 = 1$ and $E[\varepsilon_1^2]=1$. Let $G(x)=x^2$. Then recall from Example 4.6 that

$$\sum_{t=1}^n \left(X_t^2-1\right) = \sum_{t=1}^n\sum_{j=0}^\infty a_j^2\left(\varepsilon_{t-j}^2-1\right) + \sum_{t=1}^n \sum_{k,l=0;\ k\neq l}^\infty a_k a_l \varepsilon_{t-k}\varepsilon_{t-l}.$$

The first term can be represented as $\sum_{t=1}^n Y_t$, where Y_t $(t\in\mathbb{Z})$ is the linear process $Y_t = \sum_{j=0}^\infty c_k\xi_{t-j}$, $\xi_{t-j}=\varepsilon_{t-j}^2-1$, with summable coefficients $c_j=a_j^2$. Using Theorem 4.5, we have

$$n^{-1/2}\sum_{t=1}^n\sum_{j=0}^\infty a_j^2\left(\varepsilon_{t-j}^2-1\right) \overset{d}{\to} N\left(0,v^2\right),$$

where $v^2=\sigma_Y^2+2\sum_{k=1}^\infty \gamma_Y(k)$. The second term can be recognized as $U_{n,2}$, see (4.51), (4.52) and (4.53). Therefore,

$$n^{-2d}L_2^{-1/2}(n)U_{n,2} \overset{d}{\to} Z_{2,H}(1)$$

if $d\in(1/4,1/2)$, where $Z_{2,H}(u)$ is the Hermite–Rosenblatt process with $H=d+1/2$. On the other hand,

$$n^{-1/2}U_{n,2} \overset{d}{\to} \sigma_S N(0,1)$$

if $d<1/4$. Furthermore, the terms in (4.63) are uncorrelated. Therefore, if $d>1/4$, then

$$n^{-2d}L_2^{-1/2}(n)\sum_{t=1}^n\left(X_t^2-1\right) \overset{d}{\to} Z_{2,H}(1).$$

Otherwise, if $d<1/4$,

$$n^{-1/2}\sum_{j=1}^n\left(X_t^2-1\right) \overset{d}{\to} N\left(0,v+\sigma_S^2\right). \tag{4.63}$$

Example 4.10 (ARFIMA) Assume that X_t ($t \in \mathbb{N}$) is a FARIMA$(0, d, 0)$ process as in Examples 4.4 and 4.5. Then

$$\gamma_X(k) \sim c_\gamma k^{2d-1}, \qquad c_\gamma = \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1-2d) \sin(\pi d).$$

Hence, for $d \in (1/4, 1/2)$,

$$n^{-2d} L_2^{-1/2}(n) \sum_{t=1}^{n} \left(X_t^2 - 1\right) \xrightarrow{d} Z_{2,H}(1),$$

where

$$L_2(n) = 2C_2 c_\gamma^2, \quad C_2 = \frac{1}{(2(2d-1)+1)(2d+1)}.$$

Of course, this is comparable to the Gaussian case, see Example 4.1.

4.2.5.5 Technical Details for Theorem 4.9

We write the proof for $p = 1$ only, leaving out some technical details. They can be found in Ho and Hsing (1996, 1997) and Wu (2003). Using the notation $\mathcal{V}_t = (\varepsilon_t, \varepsilon_{t-1}, \ldots,)$, we may write $T_n(G; 1) = \sum_{t=1}^n U(\mathcal{V}_t)$, where $U(\cdot)$ is a suitable function. Let P_K be the conditional expectation operator

$$P_K Y = E[Y|\mathcal{V}_K] - E[Y|\mathcal{V}_{K-1}].$$

Noting that $P_K T_n(G; 1) = 0$ if $K > n$, we can write down the orthogonal decomposition

$$T_n(G; 1) = \sum_{K=-\infty}^{n} P_K T_n(G; 1).$$

Furthermore,

$$\begin{aligned} P_K T_n(G; 1) &= \sum_{t=1}^{n} \left\{E\left(U(\mathcal{V}_t)|\mathcal{F}_K\right) - E\left(U(\mathcal{V}_t)|\mathcal{F}_{K-1}\right)\right\} \\ &= \sum_{t=\max\{K,1\}}^{n} \left\{E\left(U(\mathcal{V}_t)|\mathcal{F}_K\right) - E\left(U(\mathcal{V}_t)|\mathcal{F}_{K-1}\right)\right\} \\ &= \sum_{t=\max\{K,1\}}^{n} P_K U(\mathcal{V}_t), \end{aligned}$$

since the terms corresponding to $t \le K-1$ vanish. Therefore,

$$\left\|T_n(G;1)\right\|_2^2 = \sum_{K=-\infty}^{n} \left\|P_K T_n(G;1)\right\|_2^2 = \sum_{K=-\infty}^{n} \left\| \sum_{t=\max\{K,1\}}^{n} P_K U(\mathcal{V}_t) \right\|_2^2 .$$

Now, for any stationary sequence Y_t $(t \in \mathbb{N})$, we have $\| \sum_{t=1}^{n} Y_t \|_2 \le \sum_{t=1}^{n} \|Y_t\|_2$. Therefore, if we define

$$\psi_{t-K}^2 = \left\|P_K U(\mathcal{V}_t)\right\|_2^2 = \left\|P_{-(t-K)} U(\mathcal{V}_0)\right\|_2^2$$

and use Lemma 4.17 below, we obtain

$$\left\|T_n(G;1)\right\|_2^2 \le \sum_{K=-\infty}^{n} \left(\sum_{t=\max\{K,1\}}^{n} \left\|P_{-(t-K)} U(\mathcal{V}_0)\right\|_2 \right)^2 \tag{4.64}$$

$$\le \sum_{K=-\infty}^{n} \left(\sum_{t=\max\{K,1\}}^{n} (t-K)^{2(d-1)+1/2} \right)^2 . \tag{4.65}$$

A rough bound for this expression can be established as follows:

$$\sum_{K=-\infty}^{n} \left(\sum_{t=\max\{K,1\}}^{n} (t-K)^{2(d-1)+1/2} \right)^2$$

$$\approx \int_{-\infty}^{n} \left(\int_{\max\{s,0\}}^{n} (v-s)^{2(d-1)+1/2}\, dv \right)^2 ds$$

$$= \int_{-\infty}^{0} \left(\int_{0}^{n} (v-s)^{2(d-1)+1/2}\, dv \right)^2 ds + \int_{0}^{n} \left(\int_{s}^{n} (v-s)^{2(d-1)+1/2}\, dv \right)^2 ds.$$

Let us evaluate the first term only:

$$\int_{-\infty}^{0} \left(\int_{0}^{n} (v-s)^{2(d-1)+1/2}\, dv \right)^2 ds$$

$$= C \int_{-\infty}^{0} \left((n-s)^{2(d-1)+3/2} - (-s)^{2(d-1)+3/2} \right)^2 ds$$

$$= \int_{0}^{\infty} \left((n+s)^{2(d-1)+3/2} - s^{2(d-1)+3/2} \right)^2 ds = O\left(n^{4(d-1)+3+1}\right) = O\left(n^{4d}\right).$$

This is statement (4.62) of Theorem 4.9 when $p=1$. We note that the integral above is well defined. For example, as $s \to \infty$, the integrand behaves like $\{s^{2(d-1)+1/2}\}^2$, which is integrable since $d < 1/2$. A detailed computation can be found in Lemma 5 in Wu (2003).

To finish the proof of Theorem 4.9, we have to prove the following lemma.

Lemma 4.17 *Assume that the conditions of Theorem 4.9 are satisfied. Then*

$$\|P_{-K}U(\mathcal{V}_0)\|_2^2 = O\big(K^{4(d-1)+1}\big), \quad K \geq 0.$$

Proof We have

$$\begin{aligned} P_{-K}U(\mathcal{V}_0) &= E\big[G(X_0)|\mathcal{F}_{-K}\big] - E\big[G(X_0)|\mathcal{F}_{-(K+1)}\big] \\ &\quad - G_\infty^{(1)}(0)\big\{E[X_0|\mathcal{F}_{-K}] - E[X_0|\mathcal{F}_{-(K+1)}]\big\}. \end{aligned}$$

Now we use the decomposition $X_0 = X_{0,K-1} + \tilde{X}_{0,K-1}$ and note that $X_{0,K-1}$ is independent of $\mathcal{F}_{-K}$, whereas $\tilde{X}_{0,K-1}$ is measurable w.r.t. this sigma field. Thus, recalling that $E(\varepsilon_1) = 0$, the second term in $P_{-K}U(\mathcal{V}_0)$ yields

$$E[X_0|\mathcal{F}_{-K}] - E[X_0|\mathcal{F}_{-(K+1)}] = \tilde{X}_{0,K-1} - \tilde{X}_{0,K} = a_K\varepsilon_{-K}.$$

The first term in $P_{-K}U(\mathcal{V}_0)$ is

$$G_{K-1}(\tilde{X}_{0,K-1}) - G_K(\tilde{X}_{0,K}).$$

Applying (4.57) and (4.58) with $(\xi_A, \xi_B, \mathcal{F}) = (\tilde{X}_{0,K-1}, X_{0,K-1}, \mathcal{F}_{0-K})$ and $(\xi_A, \xi_B, \mathcal{F}) = (\tilde{X}_{0,K}, X_{0,K}, \mathcal{F}_{0-(K+1)})$, our goal is to evaluate the bound

$$\|P_{-K}U(\mathcal{V}_0)\|_2^2 = \|G_{K-1}(\tilde{X}_{0,K-1}) - G_K(\tilde{X}_{0,K}) - G_\infty^{(1)}(0)a_K\varepsilon_{0-K}\|_2^2.$$

In the first step, we will replace G_{K-1} by G_K. Note first that for any $y \in \mathbb{R}$,

$$\begin{aligned} G_K(y) &= E\big[G(X_{0,K} + y)\big] = E\big[G(X_{0,K-1} + a_K\varepsilon_{-K} + y)\big] \\ &= E\big\{E\big[G(X_{0,K-1} + a_K\varepsilon_{-K} + y)|\varepsilon_{-K}\big]\big\} = E\big[G_{K-1}(y + a_K\varepsilon_{-K})\big]. \end{aligned} \tag{4.66}$$

Taking into account that $E(\varepsilon_{-K}) = 0$ and applying a Taylor expansion, we therefore obtain

$$\begin{aligned} G_{K-1}(y) - G_K(y) &= E\big[G_{K-1}(y) - G_{K-1}(y + a_K\varepsilon_{-K})\big] \\ &= E\big[G_{K-1}(y) - G_{K-1}(y + a_K\varepsilon_{j-K}) + G_{K-1}^{(1)}(y)a_K\varepsilon_{-K}\big] \\ &\leq a_K^2 E\big(\varepsilon_{-K}^2\big)\sup_y\big|G_{K-1}^{(2)}(y)\big|. \end{aligned}$$

Therefore,

$$\begin{aligned} \|P_{-K}U(\mathcal{V}_0)\|_2^2 &\leq C\big\{\|G_K(\tilde{X}_{0,K-1}) - G_K(\tilde{X}_{0,K}) + G_\infty^{(1)}(0)a_K\varepsilon_{-K}\|_2^2 + a_K^4\big\} \\ &\leq C\big\{\|G_K(\tilde{X}_{0,K-1}) - G_K(\tilde{X}_{0,K}) + G_K^{(1)}(\tilde{X}_{0,K})a_K\varepsilon_{-K}\|_2^2 + a_K^4\big\} \\ &\quad + C\|G_\infty^{(1)}(0)a_K\varepsilon_{-K} - G_K^{(1)}(\tilde{X}_{0,K})a_K\varepsilon_{-K}\|_2^2 =: I_1 + I_2. \end{aligned}$$

The first term I_1 is treated again using a Taylor approximation: it is bounded by $a_K^4 E^2(\varepsilon_1^2)\sup_y |G_K^{(2)}(y)|$. As for the second term, since $\tilde{X}_{0,K}$ and ε_{-K} are independent, we have

$$I_2 = a_K^2 E\big[\varepsilon^2\big]\big\|G_\infty^{(1)}(0) - G_K^{(1)}(\tilde{X}_{0,K})\big\|_2^2.$$

Thus, in analogy to (4.66), by conditioning on $\tilde{X}_{0,K}$,

$$G_\infty^{(1)}(y) = E\big[G^{(1)}(X+y)\big] = E\big[G_K^{(1)}(\tilde{X}_{0,K}+y)\big]. \tag{4.67}$$

Furthermore, for any two random variables η_A and η_B, we have $E[(\eta_A - E[\eta_B])^2] \le E[(\eta_A - \eta_B)^2]$. Therefore, using (4.67) with $\tilde{Y}_{0,K}$, an independent copy of $\tilde{X}_{0,K}$, we obtain

$$\begin{aligned} I_2 &\le a_K^2 E\big(\varepsilon_{-K}^2\big)\big\|G_K^{(1)}(\tilde{Y}_{0,K}) - G_K^{(1)}(\tilde{X}_{0,K})\big\|_2^2 \\ &\le 2a_K^2 E\big(\varepsilon_{-K}^2\big)\big\|G_K^{(1)}(\tilde{X}_{0,K}) - G_K^{(1)}(0)\big\|_2^2 \le C a_K^2 E\big(\tilde{X}_{0,K}^2\big)\sup_y\big|G_K^{(2)}(y)\big|. \end{aligned}$$

Hence,

$$I_2 \le C a_K^2 \sum_{j=K+1}^{\infty} a_j^2 \sim C a_K^2 \sum_{j=K+1}^{\infty} j^{2(d-1)} \sim C K^{4(d-1)+1}.$$

This finishes the proof of the lemma.

Note that we had to assume that, for $p=1$,

$$\max_{r=1,2}\sup_y \big|G_K^{(r)}(y)\big| < \infty.$$

This explains the conditions of Theorem 4.9. □

4.2.6 Stochastic Volatility Models and Their Modifications

In this section we consider limit theorems for partial sums of stochastic volatility models. Let $X_t = \sigma_t \xi_t$ $(t \in \mathbb{N})$, where

$$\sigma_t = \sigma(\zeta_t), \qquad \zeta_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j},$$

and $\sigma(\cdot)$ is a positive function. It is assumed that (ξ_t, ε_t) $(t \in \mathbb{Z})$ is a sequence of i.i.d. random vectors and $E(\varepsilon_1) = 0$. The linear process ζ_t is assumed to have long memory with autocovariance function $\gamma_\zeta(k) \sim L_\gamma(k)k^{2d-1}$, $d \in (0, 1/2)$. However, we do not assume at the moment that $E(\xi_1) = 0$. If the sequences ξ_t and ε_t are mutually independent, then the model is called LMSV (Long-Memory Stochastic Volatility), but for the purpose of this section, we do not need to make this assumption.

Let $\mathcal{G}_j$ be the sigma field generated by $\xi_l, \varepsilon_l, l \le j$. We consider partial sums

$$S_n(u) = \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \quad (u \in [0,1]),$$

where G is a measurable function such that $E[G^2(X_1)] < \infty$.

The asymptotic behaviour of partial sums is described in the following theorem. For simplicity, we formulate it in a Gaussian setting; however, it can be extended to linear processes, using the results of Sect. 4.2.5 instead of Theorem 4.4.

Theorem 4.10 *Consider the stochastic volatility model described above with* $v^2 = \mathrm{var}(G(X_1)) < \infty$ *(but possibly* $E(\xi_1) \neq 0$*). Assume in addition that* ε_t $(t \in \mathbb{Z})$ *are standard normal.*

- *If* $E[G(X_1)|\mathcal{G}_0] = 0$, *then*

$$n^{-1/2} S_n(u) \Rightarrow v B(u), \tag{4.68}$$

  where $B(u)$ $(u \in [0,1])$ *is a standard Brownian motion.*
- *If* $E[G(X_1)|\mathcal{G}_0] \neq 0$, *then*

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \Rightarrow \frac{J(m)}{m!} Z_{m,H}(u), \tag{4.69}$$

  where $\Rightarrow$ *denotes weak convergence in* $D[0,1]$, $Z_{m,H}(u)$ $(u \in [0,1])$ *is the Hermite–Rosenblatt process*, m *is the Hermite rank of*

$$\tilde{G}(y) = \int G(s\sigma(y))\, dF_\xi(s)$$

  with F_ξ *denoting the distribution of* ξ, $L_m(n) = m! C_m L_\gamma^m(n)$ *(cf.* (4.39), (4.21), (4.22)) *and* $J(m) = E[\tilde{G}(\zeta_1) H_m(\zeta_1)]$.

Proof Note that σ_t is measurable w.r.t. $\mathcal{G}_{t-1}$, whereas ξ_t is independent of $\mathcal{G}_{t-1}$. Thus,

$$\sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\}$$
$$= \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_t)|\mathcal{G}_{t-1}]\}$$
$$+ \sum_{t=1}^{[nu]} \{E[G(X_t)|\mathcal{G}_{t-1}] - E[G(X_t)]\} =: M_n(u) + R_n(u).$$

Note that the first part is a martingale. For this part, it suffices to verify the conditions of the martingale central limit theorem; see Lemma 4.2. Set $X_{t,n} = n^{-1/2}G(X_t)$. The Lindeberg condition is clearly satisfied since

$$E\big[\tilde{X}_{t,n}^2 1\big\{|\tilde{X}_{t,n}| > \delta\big\}\big] \le 4E\big[X_{t,n}^2 1\big\{|X_{t,n}| > \delta\big\}\big] \to 0$$

on account of $E[G^2(X_1)] < \infty$, where $\tilde{X}_{t,n} = X_{t,n} - E[X_{t,n}|\mathcal{G}_{t-1}]$. Furthermore, $E[G^2(X_t)|\mathcal{G}_{t-1}]$ is a measurable function of the random variable ζ_t and hence of the i.i.d. sequence $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. Therefore, the sequence $E[G^2(X_t)|\mathcal{G}_{t-1}]$ $(t \ge 1)$ is ergodic, and $n^{-1}\sum_{t=1}^{n} E[G^2(X_t)|\mathcal{G}_{t-1}]$ converges in probability to $E[G^2(X_1)]$. Therefore, we conclude (4.68) for the martingale part $M_n(u)$.

On the other hand, the second part $R_n(u)$ can be written as

$$R_n(t) = \sum_{t=1}^{[nu]} \big\{\tilde{G}(\zeta_t) - E\big[\tilde{G}(\zeta_t)\big]\big\},$$

and (4.69) can be concluded using Theorem 4.4. □

Several comments have to be made here. We note that the proof of (4.68) does not involve a particular structure of the model. Consider for example the standard stochastic volatility model where $E(\xi_1) = 0$. If we take $G(x) = x$, then $n^{-1/2}\sum_{t=1}^{[nu]} X_t$ converges to a Brownian motion without the assumption of Gaussianity on ε_t. Furthermore, it is worth mentioning that this approach works (in the case (4.68) only) for partial sums of GARCH, ARCH(∞) or LARCH(∞) models; for the latter, see Beran (2006).

Example 4.11 Assume that $G(y) = y^2$. Then $\tilde{G}(y) = E[\xi_1^2]\sigma^2(y)$. Therefore, m is the Hermite rank of $\sigma^2(y)$. In particular, if $\sigma(y) = \exp(y)$, then $m = 1$. We conclude

$$n^{-(d+1/2)} L_1^{-1/2}(n) \sum_{t=1}^{[nu]} \big(X_t^2 - E\big(X_1^2\big)\big) \Rightarrow J(1) B_H(u),$$

where $J(1) = E(\zeta_1 \exp(2\zeta_1))E(\xi_1^2)$. This is analogous to Surgailis and Viano (2002); note however that the authors considered general linear processes.

If $E(\xi_1) \ne 0$ and $G(x) = x$, then (4.68) is no longer valid; rather (4.69) holds with $m = 1$.

Example 4.12 (Long-Memory Stochastic Duration, LMSD) For the purpose of this example, we assume that random variables ξ_t $(t \in \mathbb{N})$ are strictly positive and hence non-centred. Furthermore, it is assumed that the sequences ξ_t and σ_t are independent. Then $X_t = \xi_t \sigma_t$ inherits the dependence structure from σ_t, i.e.

$$cov(X_0, X_k) = E(X_0 X_k) - E(X_0)E(X_k) = E^2[\xi_1] cov(\sigma_0, \sigma_k).$$

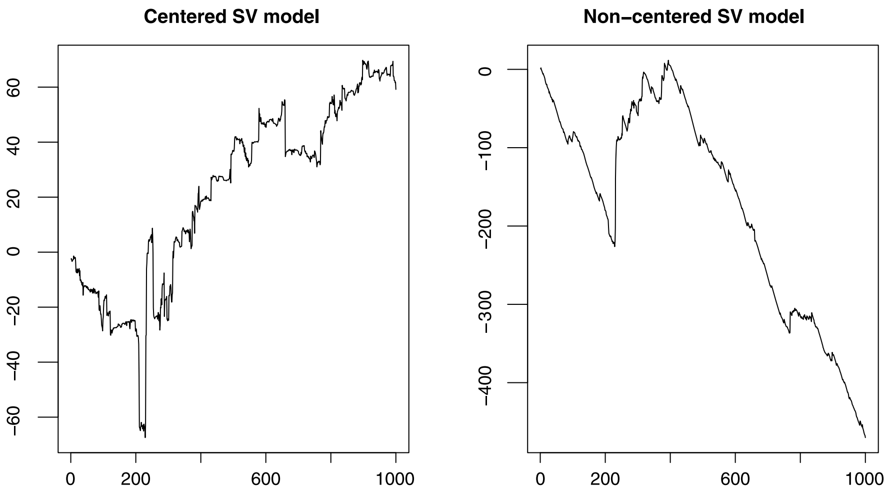

Fig. 4.3 Partial sums for a centred and a non-centred stochastic volatility model

Assume that $G(x) = x$ and $\sigma(x) = \exp(x)$. Then $\tilde{G}(y) = E(\xi_1)\exp(y)$ and $m = 1$. Application of Theorem 4.10 yields

$$n^{-(d+1/2)} L_1^{-1/2}(n) \sum_{t=1}^{[nu]} \big(X_t - E(X_1)\big) \Rightarrow J(1) B_H(u)$$

weakly in $D[0, 1]$, where $B_H(\cdot)$ is a fractional Brownian motion with $H = d + 1/2$, and $J(1) = E[\zeta_1 \exp(\zeta_1)]E[\xi_1]$.

Example 4.13 We illustrate the centering effect with a simulation example. First, we generate $n = 1000$ i.i.d. standard normal random variables ξ_t. Then we simulate independently $n = 1000$ observations ζ_t from a Gaussian FARIMA$(0, d, 0)$ process with $d = 0.4$ and compute $\sigma_t = \exp(\zeta_t)$. Then, we construct two stochastic volatility models: a centred one, $X_t = \xi_t \sigma_t$ and a non-centred one, $\tilde{X}_t = (\xi_t + 1)\sigma_t$. Finally, we plot the partial sum sequences $S_k = \sum_{t=1}^{k} X_t$ and $\tilde{S}_k = \sum_{t=1}^{k} (\tilde{X}_t - E(\tilde{X}_1))$, $k = 1, \dots, n$. The corresponding partial sum processes are plotted in Fig. 4.3. The smoother path in the second, non-centred, case indicates an influence of long memory (cf. Fig. 4.1).

4.2.7 ARCH(∞) *Models*

Recall from Definition 2.1 that the ARCH(∞) model has the form $X_t = \sigma_t \xi_t$, where ξ_t $(t \in \mathbb{Z})$ are i.i.d. zero mean random variables with variance σ_ξ^2. Also,

$$\sigma_t^2 = b_0 + \sum_{j=1}^{\infty} b_j X_{t-j}^2.$$

Furthermore, if $\sigma_\xi^2 \sum_{j=1}^{\infty} b_j < 1$, then X_t $(t \in \mathbb{Z})$ is stationary, and $E(X_1^2) < \infty$. The sequence X_t $(t \in \mathbb{Z})$ is a martingale. Using the martingale central limit theorem (see Lemma 4.2), we conclude the following result. It can also be stated in a functional form (as convergence to a Brownian motion).

Corollary 4.4 *Consider an* ARCH(∞) *model as in Definition* 2.1. *Assume that* $\sigma_\xi^2 \sum_{j=1}^{\infty} b_j < 1$. *Then*

$$n^{-1/2} \sum_{t=1}^{n} X_t \xrightarrow{d} N\big(0, \sigma_X^2\big),$$

where

$$\sigma_X^2 = \frac{\sigma_\xi^2 b_0}{1 - \sigma_\xi^2 \sum_{j=1}^{\infty} b_j}.$$

Next, we are interested in the asymptotic behaviour of

$$S_n = \sum_{t=1}^{n} \big(X_t^2 - E\big(X_1^2\big)\big).$$

To deal with this, we will use the general Definition 2.2 of ARCH(∞) models and set $Y_t = X_t^2 = v_t \zeta_t = \sigma_t^2 \xi_t^2$. In contrast to X_t $(t \in \mathbb{Z})$, the squared sequence is not a martingale. However, we recall from Theorem 2.3 that, under the existence condition $\mu_\zeta^{1/2} \sum_{j=1}^{\infty} b_j < 1$ (which guarantees $E(Y^2) < \infty$), we have the summability of the covariances, $\sum_{k=\infty}^{\infty} |\gamma_Y(k)| < \infty$. Thus, we may expect a central limit for partial sum S_n with the rate $n^{-1/2}$. Indeed, we will argue that the ARCH(∞) model $Y_t = v_t \zeta_t$, $v_t = b_0 + \sum_{j=1}^{\infty} b_j Y_{t-j}$, can be written using the Wold decomposition with respect to a martingale difference.

To see this, assume that $E(\zeta_1) = E(\xi_1^2) = 1$ and let $\psi(z) = 1 - \sum_{j=1}^{\infty} b_j z^j$. Since $\sum_{j=1}^{\infty} b_j < 1$, we conclude that $\psi(\cdot)$ is analytic on $\{z : |z| < 1\}$ and has no zeros in $\{z : |z| \le 1\}$. Hence, it is invertible, and $\psi^{-1}(z) = \sum_{j=0}^{\infty} \tilde{b}_j z^j$ with $\sum_{j=0}^{\infty} |\tilde{b}_j| < \infty$. Now, $v_t = b_0 + (1 - \psi(B))Y_t$, which leads to

$$\psi(B)Y_t = Y_t - v_t + b_0 = v_t(\zeta_t - 1) + b_0.$$

On the other hand,

$$E(Y_1) = E(v_1)E(\zeta_1) = E(v_1)$$
$$= \frac{b_0}{1 - \sum_{j=1}^{\infty} b_j},$$

so that

$$E(Y_1)\psi(B) = E(Y_1)\psi(1) = b_0.$$

Hence, $\psi(B)(Y_j - E(Y_1)) = v_t(\zeta_t - 1)$ and

$$Y_t - E(Y_1) = \sum_{j=0}^{\infty} \tilde{b}_j v_t(\zeta_t - 1).$$

We note that $v_t(\zeta_t - 1)$ $(t \in \mathbb{Z})$ is a martingale difference sequence. Therefore, the centred Y_t has a Wold decomposition with summable coefficients $\sum_{j=0}^{\infty} \tilde{b}_j$, where the innovations $v_t(\zeta_t - 1)$ are uncorrelated and martingale differences. Consequently, we could in principle apply the same method as in the proof of Theorem 4.5, provided that it can be generalized to possibly dependent innovations that are martingale differences. Since this is possible, we can conclude the following result.

Theorem 4.11 *Consider an* ARCH(∞) *process as in Definition* 2.2. *Assume that* $\sqrt{E[\xi_1^2]}\sum_{j=1}^{\infty} b_j < 1$. *Then*

$$n^{-1/2}\sum_{t=1}^{n}\bigl(Y_t - E(Y_1)\bigr) \xrightarrow{d} N\bigl(0, \sigma_Y^2\bigr),$$

where $\sigma_Y^2 = \sum_{k=-\infty}^{\infty} \gamma_Y(k)$.

4.2.8 LARCH Models

Recall that a LARCH(∞) process is defined as

$$X_t = \sigma_t \xi_t,$$

$$\sigma_t = b_0 + \sum_{j=1}^{\infty} b_j X_{t-j},$$

where $b_0 \neq 0$, and ξ_t $(t \in \mathbb{Z})$ are i.i.d. zero mean random variables with $\sigma_\xi^2 = E(\xi_1^2) = 1$. As in the case of ARCH(∞) processes, the sequence X_t is a martingale difference. Therefore, the statement of Corollary 4.4 still holds with $\sigma_X^2 = E[\sigma_1^2] = b_0^2/(1 - \|b\|_2^2)$ (cf. (2.51)).

The situation is different when we consider X_t^2. We can use the decomposition (cf. (2.56))

$$\sum_{t=1}^{n}\bigl(X_t^2 - E\bigl(X_1^2\bigr)\bigr) = \sum_{t=1}^{n}\bigl(\sigma_t^2 - E\bigl(\sigma_1^2\bigr)\bigr) + \sum_{t=1}^{n}\bigl(\xi_t^2 - 1\bigr)\sigma_t^2. \qquad (4.70)$$

The second term is a martingale and therefore of the order $O_P(\sqrt{n})$. Therefore, in the case of a long-memory LARCH(∞) process, the asymptotic behaviour of $\sum_t (X_t^2 - E(X_t^2))$ is the same as that of $\sum_t (\sigma_t^2 - E(\sigma_t^2))$. On the other hand, (2.57) of Theorem 2.7 suggests that $\sum_t (\sigma_t^2 - E(\sigma_1^2))$ behaves (up to a constant) like $\sum_t (\sigma_t - E(\sigma_1))$. This will be justified below. We then obtain the following result.

Theorem 4.12 *Consider a* LARCH(∞) *process. Let* $\mu_p = E[|\xi_1|^p] < \infty$. *Assume that* $11\mu_4^{1/2} b^2 < 1$, *where* $b = \sum_{j=1}^{\infty} b_j^2$, *and that*

$$b_j \sim c_b j^{d-1} \quad (j \to \infty), \tag{4.71}$$

where $c_b > 0$, $d \in (0, 1/2)$. *Then*

$$n^{-(d+1/2)} \sum_{t=1}^{[nu]} \big(X_t^2 - E\big(X_1^2\big)\big) \Rightarrow 2b_0^{-1} E\big(\sigma_1^2\big) c_1 \left(\frac{1}{d(2d+1)}\right)^{1/2} B_H(u),$$

where $\Rightarrow$ *denotes weak convergence in* $D[0,1]$, $B_H(u)$ *is a fractional Brownian motion with the Hurst parameter* $H = d + 1/2$, *and*

$$c_1 = \left(\frac{b_0^2}{1 - \|b\|^2}\right)^{1/2} \sqrt{B(d, 1-2d)} c_b.$$

Remark 4.1 According to Theorem 2.7, the condition $11\mu_4^{1/2} b^2 < 1$ implies that the fourth moment of X_t is finite.

Proof

Step 1: First, we look at $\sum_{t=1}^{[nu]} (\sigma_t - E(\sigma_1))$. It can be written as

$$\sum_{t=1}^{n} \big(\sigma_t - E(\sigma_1)\big) = \sum_{t=1}^{n} \sum_{l=1}^{\infty} b_l \sigma_{t-l} \xi_{t-l} = \sum_{t=1}^{n} \sum_{l=-\infty}^{t-1} b_{t-l} \sigma_l \xi_l.$$

We note that $\sigma_t \xi_t$ ($t \in \mathbb{Z}$) are uncorrelated and martingale differences. Therefore, we have the partial sum of a process $\sum b_{t-l} \sigma_l \xi_l$ that is a weighted linear sum with innovations being martingale differences. This is similar, though not identical, to the sum studied in Sect. 4.2.5 (the difference is that the innovations are only uncorrelated, not independent, i.e. we do not have a linear process). To identify asymptotic constants, rewrite the sum as $\sum_{t=1}^{n} \sum_{l=1}^{\infty} b_l \xi_{t-l} \sigma_{t-l}$. Then for $t < t'$,

$$cov\left(\sum_{l=1}^{\infty} b_l \xi_{t-l} \sigma_{t-l}, \sum_{l=1}^{\infty} b_l \xi_{t'-l} \sigma_{t'-l}\right) = \mathrm{var}(\xi_1 \sigma_1) \sum_{l=1}^{\infty} b_l b_{l+t'-t}.$$

If (4.71) holds, then, as $|j' - j| \to \infty$, the covariance behaves like

$$\mathrm{var}(\xi_1 \sigma_1) c_b^2 \int_0^{\infty} v^{d-1} (1+v)^{d-1} \, dv \big|j' - j\big|^{2d-1}$$
$$= \mathrm{var}(\xi_1 \sigma_1) c_b^2 B(d, 1-2d) \big|j' - j\big|^{2d-1}.$$

Using known results for linear processes (see Lemma 4.9), we obtain, as $n \to \infty$,

$$\mathrm{var}\left(\sum_{t=1}^{n} \sigma_t\right) \sim \mathrm{var}(\xi_1\sigma_1)\frac{1}{d(2d+1)}c_b^2 B(d, 1-2d)n^{2d+1}$$

(note that these results are applicable as long as the innovations are uncorrelated). Now,

$$\mathrm{var}(\xi_0\sigma_0) = \frac{b_0^2}{1-\|b\|_2^2}.$$

Theorem 4.6 can be generalized to the case where innovations are martingale differences. Setting

$$c_1 = \left(\frac{b_0^2}{1-\|b\|^2}\right)^{1/2}\bigl(B(d, 1-2d)\bigr)^{1/2}c_b,$$

one then can apply the generalized version of Theorem 4.6 to obtain

$$\frac{1}{n^{d+1/2}}\sum_{t=1}^{[nu]}\bigl(\sigma_t - E(\sigma_1)\bigr) \Rightarrow c_1\left(\frac{1}{d(2d+1)}\right)^{1/2} B_H(u). \qquad (4.72)$$

Step 2: To deal with $\sum_{t=1}^{[nu]}(\sigma_t^2 - E(\sigma_1^2))$, we recall that (cf. (2.57))

$$cov\bigl(\sigma_t^2, \sigma_{t+k}^2\bigr) \sim \left(\frac{2E(\sigma_1^2)}{b_0}\right)^2 cov(\sigma_t, \sigma_{t+k}) \quad (k \to \infty).$$

The implication is that the asymptotic behaviour of the partial sum is the same as that of

$$2b_0^{-1}E\bigl[\sigma_1^2\bigr]\sum_{t=1}^{[nu]}\bigl(\sigma_t - E(\sigma_1)\bigr)$$

(though more detailed arguments are required to obtain a similar linear representation as for σ_t). Hence,

$$n^{-(d+1/2)}\sum_{t=1}^{[nu]}\bigl(\sigma_t^2 - E\bigl(\sigma_1^2\bigr)\bigr) \Rightarrow 2b_0^{-1}E\bigl(\sigma_1^2\bigr)c_1\left(\frac{1}{d(2d+1)}\right)^{1/2} B_H(u).$$

Using this and decomposition (4.70), we obtain the result. □

4.2.9 Summary of Limit Theorems for Partial Sums

We summarize the main results for partial sums under long memory in Table 4.1. For simplicity, the slowly varying functions are assumed to be constant in this summary. Also, only X_t^2 is considered as a representative of nonlinear transformations.

Table 4.1 Limits for partial sums with finite moments

	Partial sums—finite moments	
	$S_n(u) = \sum_{t=1}^{[nu]} X_t$	$T_n(u) = \sum_{t=1}^{[nu]} (X_t^2 - E(X_1^2))$
Linear processes	$n^{-(1/2+d)} S_n(u) \Rightarrow cB_H(u)$ (Theorems 4.2, 4.6)	$n^{-1/2} T_n(u) \Rightarrow cB(u)$ ($d \in (0, 1/4)$) $n^{-2d} T_n(u) \Rightarrow cZ_{2,H}(u)$ ($d \in (1/4, 1/2)$) (Theorem 4.3, Corollary 4.3, Examples 4.1, 4.9)
Stochastic volatility $X_t = \xi_t \sigma_t$, $E[\xi_t] = 0$	$n^{-1/2} S_n(u) \Rightarrow cB(u)$ (Theorem 4.10)	$n^{-(1/2+d)} T_n(u) \Rightarrow cB_H(u)$ (Theorem 4.10)
LARCH	$n^{-1/2} S_n(u) \Rightarrow cB(u)$	$n^{-(1/2+d)} T_n(u) \Rightarrow cB_H(u)$ (Theorem 4.12)

4.3 Limit Theorems for Sums with Infinite Moments

4.3.1 Introduction

In this section we present limit theorems for partial sums of long-memory processes with infinite moments. Although the theory is quite well understood for weakly dependent random variables (Davis and Resnick 1985, Davis and Hsing 1995, Denker and Jakubowski 1989, Dabrowski and Jakubowski 1994, Bartkiewicz et al. 2011), the case of long memory is less well developed yet, except in the linear case. Results for linear processes with long memory were proven already several decades ago in Astrauskas (1983) and Kasahara and Maejima (1988). Subordinated linear processes were studied in Hsing (1999), Koul and Surgailis (2001), Surgailis (2002, 2004), Vaičiulis (2003). Surprisingly, the martingale decomposition method, used for finite-variance random variables in Theorem 4.9, works also here. Subordinated Gaussian processes were considered for instance in Davis (1983) and Sly and Heyde (2008). Limiting results for infinite-variance stochastic volatility models with long memory are almost non-existing; see McElroy and Politis (2007), Surgailis (2008), Kulik and Soulier (2012). In particular, both subordinated Gaussian processes and stochastic volatility models can be treated using a point process methodology. A complete list of the meanwhile quite extended literature would be too long to be included here. However, some important results and more references can be found for instance in Astrauskas et al. (1991), Benassi et al. (2002), Heath et al. (1998), Houdré and Kawai (2006), Kokoszka and Taqqu (1995a, 1995b, 1996, 1997, 1999), Koul and Surgailis (2001), Samorodnitsky (2004), Samorodnitsky and Taqqu (1994), Surgailis (2004), Zhou and Wu (2010).

First, we will summarize (with some details) results on regularly varying distributions, stable laws and point processes, referring the reader for details to standard textbooks such as Bingham et al. (1989), Feller (1971), Kallenberg (1997), Resnick (2007), Samorodnitsky and Taqqu (1994), Embrechts et al. (1997).

4.3.2 General Tools: Regular Variation, Stable Laws and Point Processes

4.3.2.1 Regular Variation

Let X_t ($t \in \mathbb{N}$) be an i.i.d. sequence whose marginal distribution has regularly varying tails:

$$P(X_1 > x) \sim \frac{1+\beta}{2} x^{-\alpha} L_X(x), \qquad P(X_1 < -x) \sim \frac{1-\beta}{2} x^{-\alpha} L_X(x) \quad (x \to \infty), \tag{4.73}$$

where $L_X(\cdot)$ is slowly varying at infinity, and $\beta \in [-1, 1]$. Condition (4.73) is the balanced tail condition. It is equivalent to $P(|X_1| > x) \sim x^{-\alpha} L_X(x)$ and

$$\lim_{x\to\infty} \frac{P(X_1 > x)}{P(|X_1| > x)} = \frac{1+\beta}{2}, \qquad \lim_{x\to\infty} \frac{P(X_1 < -x)}{P(|X_1| > x)} = \frac{1-\beta}{2}.$$

A typical example is a random variable with Cauchy density $p_X(x) = \pi(1+x^2)^{-1}$. This random variable is symmetric, and $P(X_1 > x) \sim (\pi x)^{-1}$, $x > 0$. Therefore, the Cauchy distribution is regularly varying with index $\alpha = 1$. Another example is a (two-sided) Pareto distribution where

$$P\big(|X_1| > x\big) = x^{-\alpha} \quad (x > 1).$$

We note that if $\alpha \in (0, 2)$, then random variable X has an infinite second moment. The case $\alpha = 2$ requires special attention.

Example 4.14 Assume that $L_X(x) \equiv 1$ and that for $x > x_0 > 0$, we have $\bar{F}_{|X|}(x) := P(|X| > x) = x^{-\alpha}$ with $\alpha = 2$. Then

$$\int_{x_0}^{\infty} x \bar{F}_{|X|}(x)\, dx = \int_{x_0}^{\infty} x x^{-\alpha}\, dx = \int_{x_0}^{\infty} x^{-1}\, dx = +\infty.$$

On the other hand, if $L_X(x) = (\log x)^{-2}$, then

$$\int_{x_0}^{\infty} x x^{-\alpha} \frac{1}{(\log x)^2}\, dx = \int_{x_0}^{\infty} \frac{1}{x(\log x)^2}\, dx = \int_{\log x_0}^{\infty} \frac{1}{u^2}\, du < +\infty.$$

Therefore, we have infinite and finite variance, respectively, in the first and the second case. This means that for $\alpha = 2$, the slowly varying function plays an important role.

The following result is the appropriately modified Karamata theorem. It provides extremely useful estimates for truncated moments (see e.g. Resnick 2007, pp. 25, 36).

Lemma 4.18 *Assume that X is a random variable such that* (4.73) *holds. Let* $\bar{F}(x) = P(X > x)$.

- *If* $\alpha < \eta$, *then*

$$E\big[X^{\eta} 1\big\{|X| \le x\big\}\big] \sim \frac{\alpha}{\eta - \alpha} x^{\eta} \bar{F}(x).$$

Finally note that

$$c_n = \inf\big\{x : P\big(|X| > x\big) \le n^{-1}\big\} \tag{4.74}$$

will be the appropriate normalization sequence used to establish convergence of partial sums and point process convergence. In particular, this sequence can be chosen as $c_n = n^{1/\alpha} L(n)$, where L is a slowly varying at infinity. If $L_X(x) \equiv A$ (i.e. L is constant), then $c_n = A^{1/\alpha} n^{1/\alpha}$.

4.3.2.2 Stable Random Variables

Stable random variables can be considered as a special case of (4.73). There are several equivalent definitions of stable random variables.

Definition 4.2 A random variable X is stable if for any $n \ge 2$, there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \stackrel{d}{=} c_n X + d_n,$$

where $X_1, X_2, \ldots$ are independent copies of X. Necessarily, $c_n = n^{1/\alpha}$, where $\alpha \in (0, 2]$. If $d_n = 0$, then X is called strictly stable.

Equivalently, stable random variables are characterized in terms of *domains of attraction*:

Definition 4.3 A random variable X is stable if there exists an i.i.d. sequence Y_t ($t \in \mathbb{N}$) and constants $c_n > 0$, $d_n \in \mathbb{R}$ such that

$$\frac{Y_1 + \cdots + Y_n}{c_n} + d_n \stackrel{d}{\to} X.$$

The characteristic function of a stable random variable X is given by

$$E\big[e^{i\theta X}\big] = \begin{cases} \exp(-\eta^{\alpha}|\theta|^{\alpha}(1 - i\beta \mathrm{sign}(\theta) \tan \frac{\pi\alpha}{2}) + i\mu\theta) & \text{if } \alpha \ne 1, \\ \exp(-\eta|\theta|(1 + i\beta \frac{2}{\pi} \mathrm{sign}(\theta) \ln(\theta)) + i\mu\theta) & \text{if } \alpha = 1. \end{cases}$$

Here, $0 < \alpha \le 2$, $\eta > 0$ is the scale parameter, $-1 \le \beta \le 1$ is a skewness, and $\mu \in \mathbb{R}$ a shift parameter. We write $X \sim S_{\alpha}(\eta, \beta, \mu)$. In particular, X is symmetric α-stable (written as $X \sim S\alpha S$) if $X \sim S_{\alpha}(\eta, 0, 0)$. If $\beta = 1$, then the random variable X is called *totally skewed to the right*. If $\alpha \in (1, 2]$, then $-\infty < \mu = E(X) < \infty$. In what follows, we will omit the case $\alpha = 1$ from our discussion.

If $\alpha \in (0, 2)$, then stable random variables are heavy tailed in the sense of (4.73). Indeed, if $X \sim S_\alpha(\eta, \beta, \mu)$, then

$$\lim_{x\to\infty} x^\alpha P(X > x) = C_\alpha \frac{1+\beta}{2}\eta^\alpha, \qquad \lim_{x\to\infty} x^\alpha P(X < -x) = C_\alpha \frac{1-\beta}{2}\eta^\alpha, \tag{4.75}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x\right)^{-1} = \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} \quad (\alpha \neq 1).$$

Therefore, (4.73) holds with $L_X(x) \equiv C_\alpha \eta^\alpha$. If $\eta = 1$, then the scaling constant c_n defined in (4.74) is $c_n = C_\alpha^{1/\alpha} n^{1/\alpha}$.

In what follows, we will use several properties of stable random variables. They can be obtained by considering the characteristic function. If $X_j \overset{\mathrm{d}}{=} S_\alpha(\eta_j, \beta_j, \mu_j)$ $(j = 1, 2)$ are independent, then

$$X_1 + X_2 \overset{\mathrm{d}}{=} S_\alpha\left(\left(\eta_1^\alpha + \eta_2^\alpha\right)^{1/\alpha}, \frac{\beta_1\eta_1^\alpha + \beta_2\eta_2^\alpha}{\eta_1^\alpha + \eta_2^\alpha}, \mu_1 + \mu_2\right) \tag{4.76}$$

and

$$cX_1 \overset{\mathrm{d}}{=} S_\alpha\big(|c|\eta_1, \mathrm{sign}(c)\beta_1, c\mu_1\big). \tag{4.77}$$

Due to the scaling property, it is sufficient to consider $S_\alpha(1, \beta, \mu)$ random variables.

4.3.2.3 Stable Convergence

Stable random variables play a crucial in the asymptotic theory for heavy-tailed random variables (with $\alpha \in (0, 2)$; see Gnedenko and Kolmogorov 1968, Feller 1971). Assume that X_t $(t \in \mathbb{N})$ is an i.i.d. sequence of $S_\alpha(1, \beta, \mu)$ random variables. Using (4.76) and (4.77), we have

$$n^{-1/\alpha} \sum_{t=1}^n X_t \overset{\mathrm{d}}{=} S_\alpha\left(1, \beta, \frac{n\mu}{n^{1/\alpha}}\right).$$

Thus, if $\alpha \in (0, 1)$, then $n/n^{1/\alpha} \to 0$ and

$$n^{-1/\alpha} \sum_{t=1}^n X_t \overset{\mathrm{d}}{\to} S_\alpha(1, \beta, 0). \tag{4.78}$$

If $\alpha \in (1, 2)$, a centering is required:

$$n^{-1/\alpha} \sum_{t=1}^n (X_t - \mu_n) \overset{\mathrm{d}}{=} S_\alpha\left(1, \beta, \frac{n(\mu - \mu_n)}{n^{1/\alpha}}\right).$$

Thus, we may choose $\mu_n = \mu$ (recall from Definition 4.3 that for $\alpha \in (1,2)$, we have $\mu = E(X)$) to obtain

$$n^{-1/\alpha} \sum_{t=1}^{n} (X_t - \mu) \xrightarrow{\text{d}} S_\alpha(1, \beta, 0). \tag{4.79}$$

However, we may also choose $\mu_n = E[X1\{|X| < n^{1/\alpha}\}]$. Then from the Karamata theorem, as $n \to \infty$,

$$\frac{n(\mu - \mu_n)}{n^{1/\alpha}} = \frac{nE[X \cdot 1\{|X| \geq n^{1/\alpha}\}]}{n^{1/\alpha}} \to C_\alpha \frac{\alpha}{\alpha - 1}.$$

Consequently,

$$n^{-1/\alpha} \sum_{t=1}^{n} \left(X_t - E\left[X \cdot 1\left\{|X| < n^{1/\alpha}\right\}\right]\right) \xrightarrow{\text{d}} S_\alpha\left(1, \beta, C_\alpha \frac{\alpha}{\alpha - 1}\right).$$

Of course, we can restate these results using $c_n = C_\alpha^{1/\alpha} n^{1/\alpha}$ instead of $n^{1/\alpha}$. The convergence results can be proven formally using the characteristic functions.

More generally, a classical result by Skorokhod (1957) states that if the i.i.d. random variables X_t ($t \in \mathbb{N}$) fulfill (4.73) with $L_X(x) \equiv A$, then

$$n^{-1/\alpha} S_n(u) := n^{-1/\alpha} \sum_{t=1}^{[nu]} (X_t - \mu) \Rightarrow A^{1/\alpha} C_\alpha^{-1/\alpha} Z_\alpha(u), \tag{4.80}$$

where $Z_\alpha(\cdot)$ is an α-stable Lévy motion with $Z_\alpha(u) \stackrel{\text{d}}{=} u^{1/\alpha} S_\alpha(1, \beta, 0)$, $\Rightarrow$ denotes weak convergence in $D[0,1]$ w.r.t. J_1 topology, and $\mu = E(X)$ if $\alpha \in (1,2)$ and $\mu = 0$ if $\alpha \in (0,1)$. We say then that random variables X_t ($t \in \mathbb{N}$) are in the domain of attraction of the α-stable law. Of course, if the random variables X_t are stable $S_\alpha(1, \beta, 0)$ and $u = 1$, then (4.80) reduces to (4.79) since then $A = C_\alpha$.

4.3.2.4 Point Processes

Point processes are a useful tool to study limit theorems for partial sums, sample covariances and some other functionals such as extremes. Here, we summarize (with some details) results on convergence of point processes. For a detailed exposition, the reader is referred to Resnick (2007) or Embrechts et al. (1997).

Let X_t ($t \in \mathbb{N}$) be a stationary sequence, and c_n a sequence of constants. Define the point process as

$$N_n = \sum_{t=1}^{n} \delta_{(t/n, c_n^{-1} X_t)}.$$

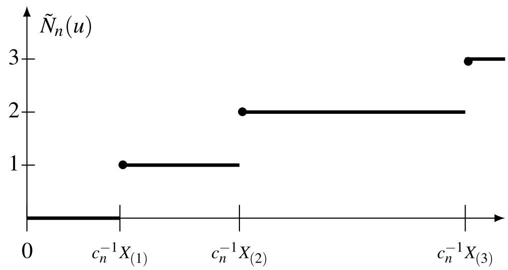

Fig. 4.4 Counting process: $X_{(1)} \le X_{(2)} \le X_{(3)}$ are the smallest observations in the sample $X_1, \dots, X_n$

Here, δ is a Dirac measure, which means that $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. A point process N_n can be viewed as a random element defined on $[0,1] \times (-\infty, \infty)$, with values in $\mathbb{N}$. In other words, this is a random element with values in $M_p(E)$, the set of all Radon point measures on $E = \mathbb{R}^2$. In particular, if we choose a set $U = [0,1] \times (0, u)$, then $N_n(U) = \tilde{N}_n(u) = \sum_{t=1}^n 1\{0 < c_n^{-1} X_t < u\}$ counts points $c_n^{-1} X_t$ that lie between 0 and u. The process $\tilde{N}_n(u)$ ($u \in \mathbb{R}_+$) is called a *counting process* and is depicted on Fig. 4.4.

There are several ways to establish convergence of point processes. The first one is referred to as Kallenberg's theorem (see Theorem 14.17 in Kallenberg 1997, or Theorem 5.2.2 in Embrechts et al. 1997).

Proposition 4.2 *Let N_n, $n \in \mathbb{N}$, and N be point processes on $\mathbb{R}^d$ such that N has no multiple points. Assume that*

$$\lim_{n\to\infty} E\big[N_n(U)\big] = E\big[N(U)\big], \tag{4.81}$$

$$\lim_{n\to\infty} P\big(N_n(U) = 0\big) = P\big(N(U) = 0\big) \tag{4.82}$$

for $U = \bigcup_{i=1}^K (k_i, l_i) \times (s_i, t_i)$, $K \ge 1$, $0 \le k_i < l_i \le 1$, and arbitrary relatively compact open intervals (s_i, t_i) of $(-\infty, 0) \cup (0, \infty)$. Then N_n converges weakly to N in $M_p(\mathbb{R}^d)$.

We illustrate this theorem by proving convergence of point processes based on i.i.d. sequences. The proof will be easily adapted to models with (long-range) dependence, such as stochastic volatility or subordinated Gaussian sequences. Define the measure λ on $(-\infty, \infty) \setminus \{0\}$ by

$$d\lambda(x) = \alpha \left[\frac{1+\beta}{2} x^{-(\alpha+1)} 1\{0 < x < \infty\} + \frac{1-\beta}{2} (-x)^{-(\alpha+1)} 1\{-\infty < x < 0\} \right] dx, \tag{4.83}$$

where $\beta \in [-1, 1]$. We say that $ds \times d\lambda(x)$ is an intensity measure of a Poisson process N on $[0,1] \times (-\infty, \infty)$ if for any $A \subset [0,1]$, $B \subset (-\infty, \infty)$, we have

$$E\big[N(A \times B)\big] = \int_B \int_A d\lambda(x)\, ds.$$

In particular, we note that $E[N([0,1] \times (-\infty, \infty))] < \infty$.

Theorem 4.13 *Let X_t $(t \in \mathbb{N})$ be a sequence of i.i.d. random variables such that* (4.73) *holds. Let*

$$P\big(|X_1| > c_n\big) \sim n^{-1}.$$

Then N_n converges weakly in $M_p([0,1] \times \mathbb{R})$ to a Poisson process N on $[0,1] \times ((-\infty, \infty) \setminus \{0\})$ with intensity measure $ds \times d\lambda(x)$.

Before we prove this result, let us state some of its consequences. First, the result can be restated as

$$\sum_{t=1}^{n} \delta_{c_n^{-1} X_t} \Rightarrow \sum_{l=0}^{\infty} \delta_{j_l},$$

where $\Rightarrow$ denotes weak convergence in $M_p(\mathbb{R})$, and j_l are points of a Poisson process with intensity measure $d\lambda(x)$. If $\alpha \in (0,1)$, then the continuous mapping theorem yields that

$$c_n^{-1} \sum_{j=1}^{n} X_t \overset{\mathrm{d}}{\to} \sum_{l=0}^{\infty} j_l.$$

If we assume for a moment that X_t $(t \in \mathbb{N})$ fulfill (4.73) with $L_X \equiv A$, then the scaling constants defined in (4.74) become $c_n = n^{1/\alpha} A^{1/\alpha}$, and so

$$n^{-1/\alpha} \sum_{t=1}^{n} X_t \overset{\mathrm{d}}{\to} A^{1/\alpha} \sum_{l=0}^{\infty} j_l.$$

For the α-stable random variables X_t, we have $A = C_\alpha$. Comparing this expression with (4.78) and using the scaling property (4.77), we conclude that $\sum_{l=0}^{\infty} j_l$ is a series representation of $S_\alpha(C_\alpha^{-1/\alpha}, \beta, 0)$. However, this consideration is not valid for the case where $\alpha \in (1,2)$.

Analogously,

$$\sum_{t=1}^{n} \delta_{c_n^{-2} X_t^2} \Rightarrow \sum_{l=0}^{\infty} \delta_{j_l^2},$$

and for $\alpha \in (0,2)$,

$$c_n^{-2} \sum_{t=1}^{n} X_t^2 \overset{\mathrm{d}}{\to} \sum_{l=0}^{\infty} j_l^2 = S_{\alpha/2}\big(C_{\alpha/2}^{-2/\alpha}, 1, 0\big),$$

or

$$n^{-2/\alpha} \sum_{t=1}^{n} X_t^2 \overset{\mathrm{d}}{\to} A^{2/\alpha} S_{\alpha/2}\big(C_{\alpha/2}^{-2/\alpha}, 1, 0\big).$$

We note that for X_t^2, the skewness parameter is $\beta = 1$. Then the stable random variable is called *totally skewed to the right*. This means that the heavy-tailed property

(4.75) of the limiting stable distribution is related to the heavy-tailed behaviour of

$$P\big(X^2 > x\big) = P(X > \sqrt{x}) + P(X < -\sqrt{x}) \sim Ax^{-\alpha},$$

which is valid for positive values of x only. In contrast, when considering X_t, the heavy-tailed behaviour of the limiting random variable $S_\alpha(C_\alpha^{-1/\alpha}, \beta, 0)$ is attributed to the heavy-tailed behaviour of $P(X > x)$ $(x > 0)$ and $P(X < x)$ $(x < 0)$.

Proof of Theorem 4.13 We verify (4.81). It is enough to consider $U = \bigcup_{i=1}^K (k_i, l_i) \times (s_i, t_i)$ for $K = 1$. We have

$$\begin{aligned} E\big[N_n(U)\big] &= \sum_{t=1}^n E[\delta_{(t/n, c_n^{-1} X_t)}] \\ &= (l_1 - k_1) P\big(c_n^{-1} X_t \in (s_1, t_1)\big) \\ &\to (k_1 - l_1)\lambda\big((s_1, t_1)\big), \end{aligned}$$

where we recall that $\lambda((s_i, t_i)) = \int_{s_i}^{t_i} d\lambda(x)$, and the measure $\lambda(\cdot)$ is given by (4.83). To prove (4.82), write

$$\begin{aligned} P\big(N_n(U) = 0\big) &= P\left(\sum_{i=1}^K \sum_{nk_i < t < nl_i}^n 1\big\{c_n^{-1} X_t \in (s_i, t_i)\big\} = 0\right) \\ &= \prod_{i=1}^K \prod_{nk_i < t < nl_i}^n P\big(c_n^{-1} X_t \notin (s_i, t_i)\big). \end{aligned}$$

Let

$$Q_n = \prod_{i=1}^K \prod_{nk_i < t < nl_i} e^{-n^{-1}\lambda((s_i, t_i))}$$

and note that

$$\begin{aligned} Q_n &= \exp\left(-\sum_{i=1}^K n^{-1} \sum_{nk_i < t < nl_i} \lambda\big((s_i, t_i)\big)\right) \to \exp\left(-\sum_{i=1}^K (l_i - k_i)\lambda\big((s_i, t_i)\big)\right) \\ &= P\big(N(U) = 0\big) \end{aligned}$$

as $n \to \infty$. Recall the two elementary inequalities

$$\left|\prod_{i=1}^K (s_i - t_i)\right| \le \sum_{i=1}^K |s_i - t_i| \quad \text{and} \quad \big|1 - e^{-x} - x\big| \le x^{1+\varepsilon}$$

for any $\varepsilon > 0$. Then we obtain

$$\begin{aligned}
&\left|P\big(N_n(U)=0\big)-Q_n\right| \\
&\quad=\left|\prod_{i=1}^{K}\prod_{nk_i<t<nl_i}\big(1-P\big(c_n^{-1}X\in(s_i,t_i)\big)\big)-\prod_{i=1}^{K}\prod_{nk_i<t<nl_i}e^{-n^{-1}\lambda((s_i,t_i))}\right| \\
&\quad\le\sum_{i=1}^{K}(l_i-k_i)n\left|\big(1-P\big(c_n^{-1}X\in(s_i,t_i)\big)\big)-e^{-n^{-1}\lambda((s_u,t_u))}\right| \\
&\quad\le\sum_{i=1}^{K}(l_i-k_i)\left|nP\big(c_n^{-1}X\in(s_i,t_i)\big)-\lambda\big((s_i,t_i)\big)\right| \\
&\qquad+\sum_{i=1}^{K}n(l_i-k_i)\left|1-e^{-n^{-1}\lambda((s_i,t_i))}-\frac{\lambda((s_i,t_i))}{n}\right| \\
&\quad=o(1)+Cn^{-\varepsilon}=o(1)
\end{aligned}$$

for some $\varepsilon>0$. □

Another result, due to Davis and Resnick (1988, Proposition 2.1), is useful when studying processes that can be approximated by sequences with finite memory. Their result is stated in fact in a much more general setting, which is omitted here.

We say that a sequence ν_n of measures converges vaguely to ν ($\nu_n \overset{v}{\to} \nu$) if for all continuous functions $g: E\to\mathbb{R}^d$ with compact support (written as $g\in C^+(E)$), we have

$$\int g(x)\nu_n(dx)\to\int g(x)\nu(dx).$$

We refer to Appendix A for additional precise notions related to vague convergence.

Proposition 4.3 *Assume that X_t ($t\in\mathbb{N}$) is a stationary K-dependent sequence with values in $\mathbb{R}^d$ and $c_n\to\infty$ is a sequence of constants such that for the marginal distribution, we have*

$$nP\big(c_n^{-1}X\in\cdot\big)\overset{v}{\to}\lambda(\cdot).$$

Furthermore, assume that for any $g\in C^+(\mathbb{R}^d)$,

$$\lim_{k\to\infty}\limsup_{n\to\infty}n\sum_{t=2}^{[n/k]}E\big[g\big(c_n^{-1}X_1\big)g\big(c_n^{-1}X_t\big)\big]=0.$$

Then

$$N_n=\sum_{t=1}^{n}\delta_{(t/n,c_n^{-1}X_t)}$$

converges weakly in $M_p([0,1] \times \mathbb{R})$ *to a Poisson process* N *on* $[0,1] \times (-\infty, \infty)$ *with intensity measure* $ds \times d\lambda(x)$.

This result is applicable to sequences X_t with regularly varying tails as in (4.73). In fact (see Theorem 3.6 in Resnick 2007), the vague convergence of $nP(c_n^{-1}X \in \cdot)$ is equivalent to regular variation of the distribution of X.

4.3.3 Sums of Linear and Subordinated Linear Processes

In this section we discuss limit theorems for partial sums of linear processes

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

where $a_j \sim c_a j^{d-1}$, $d \in (0, 1/2)$, and ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that

$$P(\varepsilon_1 > x) \sim A \frac{1+\beta}{2} x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A \frac{1-\beta}{2} x^{-\alpha}. \tag{4.84}$$

In both, the coefficients a_j and the tail $P(\varepsilon_1 > x)$, we assume for simplicity that possible slowly varying functions are constant. If $\alpha \in (1,2)$, we assume also that $E(\varepsilon_1) = 0$.

The infinite series above converges if $\sum_{j=0}^{\infty} |a_j|^{\delta} < \infty$ for some $\delta < \alpha$ (see e.g. Avram and Taqqu 1992). In our case this is possible if an only if $\alpha(d-1) < -1$ and hence $d < 1 - 1/\alpha$. Thus, if $\alpha \in (0,1)$, then the existence condition implies that $\sum_{j=0}^{\infty} |a_j| < \infty$. Consequently, for $\alpha \in (0,1)$, long memory (in the sense of non-summability of the coefficients) is excluded.

Linear processes are the easiest models to describe the interplay between dependence and heavy tails. The asymptotic theory for partial sums is well developed and includes approaches such as convergence of stochastic integrals (Astrauskas 1983, Kasahara and Maejima 1986, 1988) or K-dependent approximations, together with the point process methodology (Davis and Resnick 1985, Davis and Hsing 1995). Interesting results on functional convergence are given in Avram and Taqqu (1992), among others.

4.3.3.1 Tail Behaviour

First, we analyse the tail behaviour of linear processes. We note that if ε_t $(t \in \mathbb{Z})$ are $S_\alpha(1,0,0)$, so that (4.84) holds with $\beta = 0$ and $A = C_\alpha$, then

$$X_1 \stackrel{\mathrm{d}}{=} \left(\sum_{j=0}^{\infty} |a_j|^{\alpha} \right)^{1/\alpha} S_\alpha(1,0,0) =: D_\alpha^{1/\alpha} S_\alpha(1,0,0) \stackrel{\mathrm{d}}{=} D_\alpha^{1/\alpha} \varepsilon_1,$$

which follows directly from properties (4.76) and (4.77). Therefore, we may conclude that, as $x \to \infty$,

$$P\left(|X_1| > x\right) \sim P\left(D_\alpha^{1/\alpha}|\varepsilon_1| > x\right) \sim D_\alpha C_\alpha x^{-\alpha} \sim D_\alpha P\left(|\varepsilon_1| > x\right).$$

This property is valid in fact under the general assumption (4.84).

Lemma 4.19 *Assume that X_t $(t \in \mathbb{N})$ is a linear process, ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that (4.84) holds, and $E(\varepsilon_1) = 0$ if $\alpha \in (1, 2)$.*

- *If for some $\delta < \alpha$,*

$$\sum_{j=0}^{\infty} |a_j| + \sum_{j=0}^{\infty} |a_j|^\delta < \infty, \tag{4.85}$$

  then

$$\lim_{x\to\infty} \frac{P(|X_1| > x)}{P(|\varepsilon_1| > x)} = \sum_{j=0}^{\infty} |a_j|^\alpha. \tag{4.86}$$

- *If $a_j \sim c_a j^{d-1}$, $d \in (0, 1 - 1/\alpha)$, and ε_t $(t \in \mathbb{Z})$ are symmetric with $\alpha \in (1, 2)$, then (4.86) holds.*

Note that in the second part of the theorem, the coefficients a_j are not absolutely summable, however $\sum |a_j|^\alpha$ is finite. This turns out to be sufficient. The first part was proven in Cline (1983); see also Davis and Resnick (1985). The second part was proven (under special assumptions with symmetry of the innovations) in Kokoszka and Taqqu (1996).

4.3.3.2 Point Process Convergence

In what follows we show that, under the conditions of Lemma 4.19, a point process based on X_t $(t \in \mathbb{N})$ converges. Its behaviour is the same under short memory (4.85) and under long memory.

Theorem 4.14 *Under the assumptions of Lemma 4.19, we have*

$$\sum_{t=1}^{n} \delta_{c_n^{-1}(X_t,\ldots,X_{t-K})} \Rightarrow \sum_{l=1}^{\infty}\sum_{r=0}^{\infty} \delta_{j_l(a_r,a_{r-1},\ldots,a_{r-K})}$$

in $M_p(\mathbb{R}^{K+1})$, where c_n is such that $P(|\varepsilon_1| > c_n) \sim n^{-1}$, i.e. $c_n \sim A^{1/\alpha} n^{1/\alpha}$.

Proof We give the proof for $K = 0$ only. For details, we refer to Davis and Resnick (1985, Theorem 2.4). We note that the authors prove the results under condition (4.85). However, a crucial part of the proof relies on (4.86) only, which due to

Lemma 4.19 is valid under more general conditions on a_j. We restate Theorem 4.13 in terms of i.i.d. random variables ε_t ($t \in \mathbb{Z}$),

$$\sum_{t=1}^{n} \delta_{c_n^{-1}\varepsilon_t} \Rightarrow \sum_{l=1}^{\infty} \delta_{j_l}$$

where $c_n \sim A^{1/\alpha} n^{1/\alpha}$. Moreover (see Theorem 2.2. in Davis and Resnick 1985), this convergence can be extended to

$$\sum_{t=1}^{n} \delta_{c_n^{-1}(\varepsilon_t,\ldots,\varepsilon_{t-K})} \Rightarrow \sum_{l=1}^{\infty}\sum_{r=0}^{K} \delta_{j_l \mathbf{e}_r}, \tag{4.87}$$

where $\mathbf{e}_r$ is a unit vector in $\mathbb{R}^{K+1}$ with the rth coordinate equal to one. In other words, the limiting process has the following structure. It is a Poisson process with values in $\{0,\ldots,K\} \times \mathbb{R}$ such that it is a univariate Poisson process on the horizontal line $\{0\} \times \mathbb{R}$ and its points are repeated on the other horizontal lines. Since the mapping $(z_t,\ldots,z_{t-K}) \to \sum_{r=0}^{K} b_k z_{t-k}$ from $M_p(\mathbb{R}^{K+1})$ to $M_p(\mathbb{R} \setminus \{0\})$ is continuous, (4.87) implies

$$\sum_{t=1}^{n} \delta_{c_n^{-1} X_{t,K}} \Rightarrow \sum_{l=1}^{\infty}\sum_{r=0}^{K} \delta_{j_l a_r},$$

where $X_{t,K} = \sum_{r=0}^{K} a_r \varepsilon_{t-k}$. Letting $K \to \infty$, we obtain

$$\sum_{l=1}^{\infty}\sum_{r=0}^{K} \delta_{j_l a_r} \overset{\mathrm{p}}{\to} \sum_{l=1}^{\infty}\sum_{r=0}^{\infty} \delta_{j_l a_r}.$$

Therefore, to apply Proposition 4.1, we need to verify that the sequence X_t can be approximated by the K-dependent sequence $X_{t,K}$, in the sense that for each $\gamma > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\Big(c_n^{-1} \sup_{1\le t\le n} |X_t - X_{t,K}| > \gamma\Big) = 0.$$

The latter probability is bounded by $nP(c_n^{-1}|X_0 - X_{0,K}| > \gamma)$. Since $P(|\varepsilon_1| > c_n) \sim n^{-1}$, applying (4.86), we have, as $n \to \infty$,

$$nP\big(c_n^{-1}|X_0 - X_{0,K}| > \gamma\big) \sim \frac{P(|X_0 - X_{0,K}| > c_n\gamma)}{P(|\varepsilon_1| > c_n)} = \gamma^{-\alpha} \sum_{r=K+1}^{\infty} |a_r|^{\alpha}.$$

The last expression converges to zero as $K \to \infty$. □

4.3.3.3 Convergence of Partial Sums

Recall our comments following Theorem 4.13. If the innovations ε_t have tail index $\alpha \in (0, 1)$, then we may conclude directly from Theorem 4.14 that

$$c_n^{-1} \sum_{t=1}^{n} X_t \overset{d}{\to} \left(\sum_{j=0}^{\infty} a_j \right) \sum_{l=1}^{\infty} j_l \overset{d}{=} \left(\sum_{j=0}^{\infty} a_j \right) S_\alpha\big(C_\alpha^{-1/\alpha}, \beta, 0\big),$$

where j_l are points of a Poisson process, and $\sum_{l=1}^{\infty} j_l$ is a series representation of $S_\alpha(C_\alpha^{-1/\alpha}, \beta, 0)$. Equivalently,

$$n^{-1/\alpha} \sum_{t=1}^{n} X_t \overset{d}{\to} A^{1/\alpha} \left(\sum_{j=0}^{\infty} a_j \right) S_\alpha\big(C_\alpha^{-1/\alpha}, \beta, 0\big) \overset{d}{=} A^{1/\alpha} C_\alpha^{-1/\alpha} \left(\sum_{j=0}^{\infty} a_j \right) S_\alpha(1, \beta, 0).$$

The situation is more complicated for $\alpha \in (1, 2)$. Convergence of partial sums does not follow directly from point process convergence (however, as in Davis and Resnick 1985, an implication of point process convergence may serve as an intermediate tool—this will be illustrated for stochastic volatility models in the following section). In particular, for a long-memory sequence, the scaling for partial sums $\sum_{t=1}^{n} X_t$ of linear processes may differ from c_n.

Theorem 4.15 *Assume that X_t $(t \in \mathbb{Z})$ is a linear process such that $a_j \sim c_a j^{d-1}$, $d \in (0, 1/2)$ and ε_t $(t \in \mathbb{Z})$ are i.i.d random variables such that (4.84) holds with $\alpha \in (1, 2)$ and $E(\varepsilon_1) = 0$.*

- *If for some $\delta < \alpha$,*

$$\sum_{j=0}^{\infty} |a_j| + \sum_{j=0}^{\infty} |a_j|^{\delta} < \infty, \tag{4.88}$$

then

$$n^{-1/\alpha} S_n(u) = n^{-1/\alpha} \sum_{t=1}^{[nu]} X_t \overset{\text{f.d.}}{\to} A^{1/\alpha} C_\alpha^{-1/\alpha} \left(\sum_{j=0}^{\infty} a_j \right) Z_\alpha(u),$$

where $Z_\alpha(\cdot)$ is an α-stable Lévy motion (with independent increments) such that $Z_\alpha(1) \overset{d}{=} S_\alpha(1, \beta, 0)$, and $\overset{\text{f.d.}}{\to}$ denotes finite-dimensional convergence.

- *If $0 < d < 1 - 1/\alpha$, then*

$$n^{-H} S_n(u) = n^{-H} \sum_{t=1}^{[nu]} X_t \Rightarrow A^{1/\alpha} C_\alpha^{-1/\alpha} \frac{c_a}{d} \tilde{Z}_{H,\alpha}(u),$$

where $H = d + \alpha^{-1}$, $\tilde{Z}_{H,\alpha}(\cdot)$ is a Linear Fractional stable motion, and $\Rightarrow$ denotes weak convergence in $D[0, 1]$ w.r.t. the Skorokhod J_1-topology.

Before we present a proof, we make several comments.

Remark 4.2 If condition (4.88) holds, then the scaling factor and the limiting process are (up to a constant) the same as for i.i.d. random variables; see (4.80). The limiting Lévy process has independent increments and discontinuous sample paths. Thus, in this case the particular structure of the coefficients a_j is not really important. On the other hand, if $d \in (0, 1 - 1/\alpha)$, then the scaling factor involves the memory parameter d. This is one reason why such a process is said to have long-range dependence. Also, the limiting process has dependent increments but continuous sample paths. We illustrate this in Example 4.15. Note also that the theorem can be stated more generally by allowing slowly varying functions in both a_j and the tail of ε_1.

Remark 4.3 It should be pointed out that in the long-memory case ($d \in (0, 1-1/\alpha)$) we have weak convergence w.r.t. the standard J_1-topology and the limiting process has continuous paths. In contrast, in the case of summable coefficients we have finite-dimensional convergence only, and this cannot be extended to J_1-convergence. This can be seen as follows. Assume for a moment that $X_t = b_0\varepsilon_t + b_1\varepsilon_{t-1}$ ($t \in \mathbb{N}$). The limiting behaviour of $S_n = \sum_{t=1}^n X_t$ is determined by large values of X_t ($t \in \mathbb{N}$). Now, there is a small chance that both ε_t and ε_{t+1} are large since $P(\varepsilon_t > x, \varepsilon_{t+1} > x) = o(P(\varepsilon_1 > x))$ as $x \to \infty$. Therefore, we have one large value of a particular ε_{t^*}, say which implies $X_{t^*} \approx b_0\varepsilon_{t^*}$ and $X_{t^*+1} \approx b_1\varepsilon_{t^*}$. This produces two "clustered" large jumps in the limiting process, which contradicts a heuristic explanation of J_1-topology in the Appendix A. However, it is possible to have weak convergence w.r.t. different topologies. We refer to Avram and Taqqu (1992).

Proof In the case of weak dependence (i.e. where (4.88) holds), the proof mimics the one for normal convergence (see Theorem 4.5). Let $X_{t,K} = \sum_{j=0}^K a_j\varepsilon_{t-j}$. Note that (4.80) can be restated for $u = 1$ as

$$n^{-1/\alpha}\left(\sum_{t=1}^n \varepsilon_t, \ldots, \sum_{t=1}^n \varepsilon_{t-m}\right) \xrightarrow{\mathrm{d}} A^{1/\alpha}C_\alpha^{-1/\alpha}\big(Z_\alpha(1), \ldots, Z_\alpha(1)\big).$$

The continuous mapping theorem implies

$$n^{-1/\alpha}\sum_{t=1}^n X_{t,K} \xrightarrow{\mathrm{d}} A^{1/\alpha}C_\alpha^{-1/\alpha}\left(\sum_{j=0}^K a_j\right)Z_\alpha(1).$$

Furthermore, $(\sum_{j=0}^K a_j)Z_\alpha(1) \xrightarrow{\mathrm{p}} (\sum_{j=0}^\infty a_j)Z_\alpha(1)$. We finish the proof by verifying

$$\limsup_{K\to\infty}\lim_{n\to\infty} P\left(n^{-1/\alpha}\left|S_n(1) - S_{n,K}(1)\right| > \gamma\right) = 0$$

for each $\gamma > 0$. This requires precise calculations on the tail behaviour of X_t. In particular, (4.86) plays a crucial role. We refer to Davis and Resnick (1985) for details. The result then follows from Proposition 4.1.

As for the long-memory case, we assume for simplicity that ε_t $(t \in \mathbb{Z})$ are $S_\alpha(1, \beta, 0)$. We may write

$$S_n = \sum_{t=1}^{n} X_t = \sum_{l=-\infty}^{n} \varepsilon_l \sum_{j=1-l}^{n-l} a_j =: \sum_{l=-\infty}^{n} \tilde{a}_{l,n} \varepsilon_l$$

with $\tilde{a}_{l,n} = \sum_{j=1-l}^{n-l} a_j$. If $a_j \sim c_a j^{d-1}$, then

$$\tilde{a}_{l,n} \sim \frac{c_a}{d} \left\{ (n-l)^d - (1-l)^d \right\}.$$

Therefore, since S_n is a sum of independent stable random variables, on account of (4.76), we expect that

$$\sum_{l=-\infty}^{n} \tilde{a}_{l,n} \varepsilon_l \overset{\mathrm{d}}{=} S_\alpha(\eta_n, \beta, 0)$$

with the scale parameter such that

$$\eta_n^\alpha = \sum_{l=-\infty}^{n} \tilde{a}_{l,n}^\alpha = \left(\frac{c_a}{d}\right)^\alpha \sum_{l=-\infty}^{n} \left\{ (n-l)^d - (1-l)^d \right\}^\alpha$$

$$\sim \left(\frac{c_a}{d}\right)^\alpha \frac{1}{n^{d\alpha+1}} \int_{-\infty}^{1} \left\{ (1-v)_+^d - (-v)_+^d \right\}^\alpha dv.$$

Here, note that the integral above is defined only if $0 < d < 1 - 1/\alpha$. Therefore, with $b_n = (c_\alpha/d) n^H$ (recall that now $C_\alpha = A$ since we consider stable innovations), the distribution of $b_n^{-1} S_n(1)$ agrees asymptotically with the distribution of a stable random variable with the scale

$$\eta = \left(\int_{-\infty}^{1} \left\{ (1-v)_+^d - (-v)_+^d \right\}^\alpha dv \right)^{1/\alpha}$$

and skewness β. Now, if we have a stable integral $\int g(x)\, dM(x)$, then it is a stable random variable with the scale $(\int |g(x)|^\alpha \, dx)^{1/\alpha}$. Thus, for each u, the Linear Fractional Stable Motion $\tilde{Z}_{H,\alpha}(\cdot)$ (see Sect. 3.7.2 for additional details)

$$\int_{-\infty}^{u} \left\{ (u-v)_+^{H-1/\alpha} - (-v)_+^{H-1/\alpha} \right\} dZ_\alpha(v)$$

is a stable random variable with the scale

$$u^{1/\alpha} \left(\int_{-\infty}^{1} \left\{ (u-v)_+^{H-1/\alpha} - (-v)_+^{H-1/\alpha} \right\}^\alpha dv \right)^{1/\alpha}.$$

Consequently, the result follows for $u = 1$. In this argument we replaced the coefficients $\tilde{a}_{l,n}$ by the asymptotically equivalent expressions. This approximation can be made more precise by computing the characteristic function. □

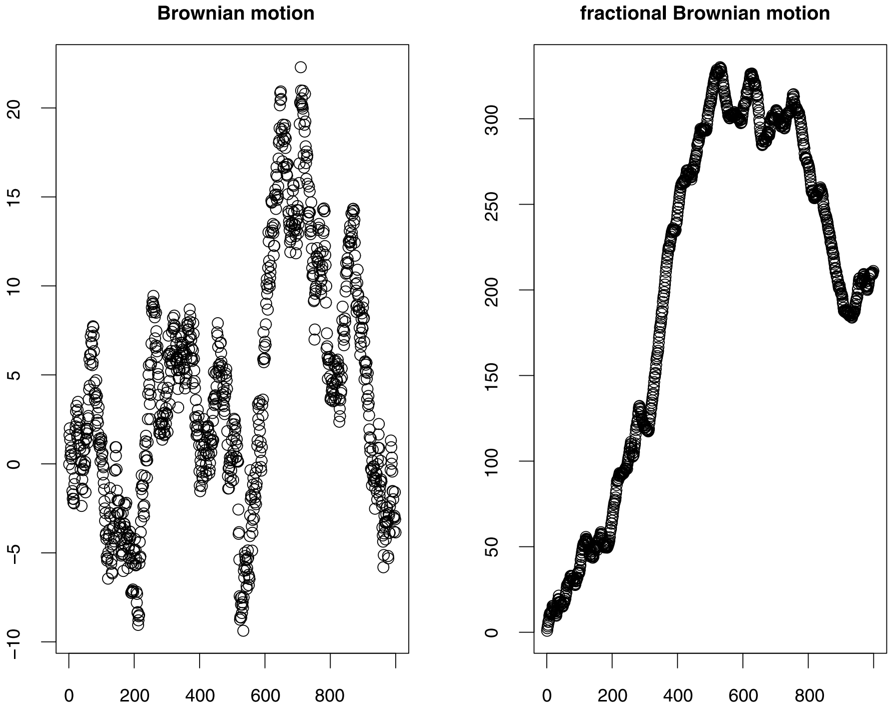

Fig. 4.5 Paths of a partial sum sequence $S_k = \sum_{t=1}^{k} X_t$ with X_t i.i.d. $N(0,1)$ (*left*) and X_t generated by a FARIMA(0, 0.4, 0) process

Example 4.15 We illustrate Theorem 4.15 by a simulation study. First, as in Example 4.2, we generate $n = 1000$ i.i.d. standard normal random variables X_t and plot the partial sum sequence $S_k = \sum_{t=1}^{k} X_t$, $k = 1, \ldots, n$. This procedure is repeated for Gaussian FARIMA$(0, d, 0)$ process with $d = 0.4$. The path of the fractional Brownian motion is much smoother than of the Brownian motion. This is due to the influence of long memory. The corresponding partial sum processes are plotted in Fig. 4.5. For comparison, we simulate i.i.d. random variables from a t-distribution with $3/2$ degrees of freedom (hence, with a finite mean and infinite variance) and a FARIMA(0, 0.4, 0) process where the innovations have a t-distribution with $3/2$ degrees of freedom. The partial sum processes are depicted on Fig. 4.6. In the i.i.d. case, the process has clearly discontinuous sample paths, whereas this effect does not seem to be present in the long-memory case.

4.3.3.4 Subordinated Case

Consider the partial sum

$$S_{n,G}(u) = \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \quad \bigl(u \in [0, 1]\bigr),$$

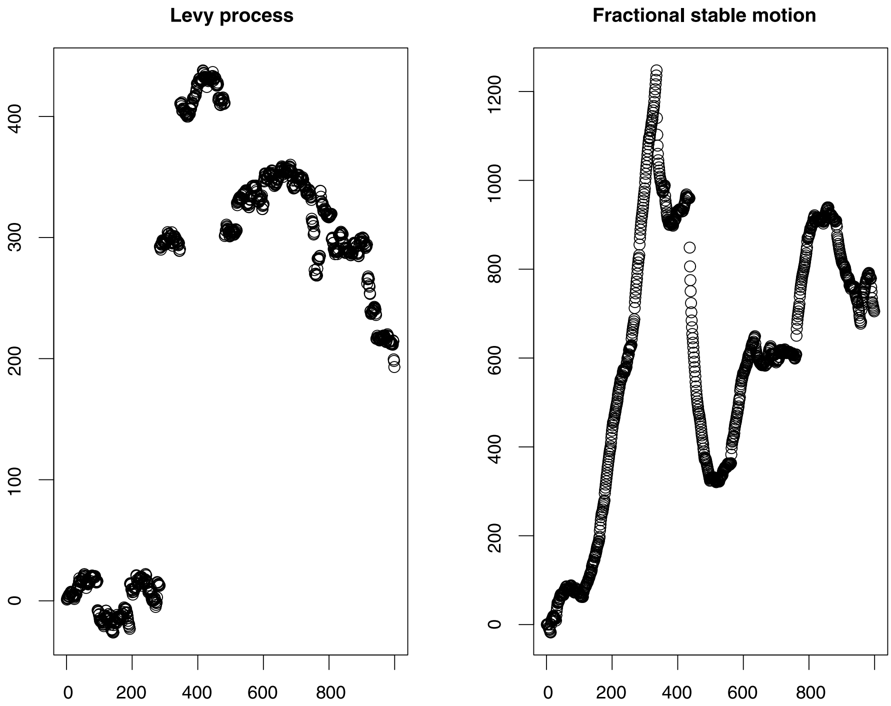

Fig. 4.6 Paths of a Lévy stable motion and a fractional stable motion with Hurst parameter $H = d + 1/\alpha$, $d = 0.4$, $\alpha = 3/2$

where G is a measurable function. Subordinated linear processes with infinite second moments were studied in Hsing (1999), Koul and Surgailis (2001), Surgailis (2002, 2004), Vaičiulis (2003). Surprisingly, the martingale decomposition method, used in Theorem 4.9 for variables with finite variance, works also here.

We start with the simple case of polynomials. Let us focus on a quadratic function $G(x) = x^2$. If $\alpha \in (0,2)$, then we can repeat the argument following point process convergence in Theorem 4.14. First (see the discussion following Theorem 4.13), we can also write

$$\sum_{t=1}^{n} \delta_{c_n^{-2} X_t^2} \Rightarrow \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \delta_{j_l^2 a_j^2}.$$

This is valid as long as the conditions of Lemma 4.19 hold. Now, if $\alpha \in (0,2)$, the random variables X_t^2 $(t \in \mathbb{N})$ have infinite means. Therefore, for $\alpha \in (0,2)$,

$$c_n^{-2} \sum_{t=1}^{n} X_t^2 \overset{\mathrm{d}}{\to} \left(\sum_{j=0}^{\infty} a_j^2 \right) \sum_{l=0}^{\infty} j_l^2 = \left(\sum_{j=0}^{\infty} a_j^2 \right) S_{\alpha/2}\left(C_{\alpha/2}^{-2/\alpha}, 1, 0\right),$$

or equivalently,

$$n^{-2/\alpha}\sum_{t=1}^{n}X_t^2 \stackrel{\mathrm{d}}{\to} \left(\sum_{j=0}^{\infty}a_j^2\right)A^{2/\alpha}C_{\alpha/2}^{-2/\alpha}S_{\alpha/2}(1,1,0).$$

The case $\alpha \in (0,1)$ was proven in Davis and Resnick (1985, Theorem 4.2), whereas the case $\alpha \in (1,2)$ is addressed in Kokoszka and Taqqu (1996, Theorem 2.1). In other words, long memory does not influence the limiting behaviour.

Now, the situation changes when $2 < \alpha < 4$. The partial sum

$$S_{n,G}(u) = \sum_{t=1}^{[nu]}\left(X_t^2 - E\left(X_1^2\right)\right)$$

can be decomposed as (cf. Example 4.9)

$$S_{n,G,1}(u) + S_{n,G,2}(u) := \sum_{t=1}^{[nu]}\sum_{j=0}^{\infty}a_j^2\left(\varepsilon_{t-j}^2 - E\left(\varepsilon_1^2\right)\right) + \sum_{t=1}^{[nu]}\sum_{j,k=0;\ j\neq k}^{\infty} a_j a_l \varepsilon_{t-j}\varepsilon_{t-k}.$$

The first part $S_{n,G,1}(u)$ is a partial sum process based on the linear process with summable coefficients a_j^2. Therefore, on account of the first part of Theorem 4.15,

$$n^{-2/\alpha}S_{n,G,1}(u) \stackrel{\text{f.d.}}{\to} A^{2/\alpha}C_{\alpha/2}^{-2/\alpha}\left(\sum_{j=0}^{\infty}a_j^2\right)Z_{\alpha/2}(u),$$

where $Z_{\alpha/2}(\cdot)$ is a Lévy process such that $Z_{\alpha/2}(1) \stackrel{\mathrm{d}}{=} S_{\alpha/2}(1,1,0)$, i.e. $Z_{\alpha/2}(1)$ is an $\alpha/2$-stable random variable that is completely skewed to the right.

Convergence of the second term follows exactly as in Example 4.9. First, since $2 < \alpha < 4$, the random variables ε_t have a finite variance where under the assumption $a_j \sim c_a j^{d-1}$ we have $\gamma_X(k) = cov(X_t, X_{t+k}) \sim L_\gamma(k)k^{2d-1}$ with

$$L_\gamma(k) = c_a^2\sigma_\varepsilon^2\int_0^\infty v^{d-1}(v+1)^{d-1}\,dv,$$

see Lemma 4.13. If $1/4 < d < 1/2$, then

$$n^{-2d}L_2^{-1/2}(n)S_{n,G,2}(u) \Rightarrow Z_{2,H}(u),$$

where $H = d + 1/2$, $Z_{2,H}(u)$ is the Hermite–Rosenblatt process, and

$$L_2(n) = m!C_m L_\gamma^m(n).$$

Otherwise, if $0 < d < 1/4$, then $n^{-1/2}S_{n,G,2}(u) = O_P(1)$. Therefore, we have a dichotomous behaviour depending on a relation between the "memory parameter" d and tails. Such consideration can be carried out for instance for Appell polynomials (see Vaičiulis 2003). Before we state our theorem, we recall for convenience the heavy-tail condition (4.84):

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}. \tag{4.89}$$

Theorem 4.16 *Assume that X_t $(t \in \mathbb{Z})$ is a linear process such that $a_j \sim c_a j^{d-1}$, $d \in (0, 1/2)$ and ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that (4.89) holds with $\alpha \in (2, 4)$. Also, assume that $E(\varepsilon_1) = 0$.*

- *If $0 < d < 1/\alpha$, then*

$$n^{-2/\alpha}\sum_{t=1}^{[nu]}\bigl(X_t^2 - E\bigl(X_1^2\bigr)\bigr) \overset{\text{f.d.}}{\to} A^{2/\alpha}C_{\alpha/2}^{-2/\alpha}\left(\sum_{j=0}^{\infty}a_j^2\right)Z_{\alpha/2}(u),$$

  where $Z_{\alpha/2}(\cdot)$ is an $\alpha/2$-stable Lévy motion such that $Z_{\alpha/2}(1) \overset{\mathrm{d}}{=} S_{\alpha/2}(1, 1, 0)$.
- *If $1/\alpha < d < 1/2$, then*

$$n^{-2d}L_2^{-1/2}(n)\sum_{t=1}^{[nu]}\bigl(X_t^2 - E\bigl(X_1^2\bigr)\bigr) \Rightarrow Z_{2,H}(u),$$

  where $\Rightarrow$ denotes weak convergence in $D[0, 1]$, $Z_{2,H}(\cdot)$ is the Hermite–Rosenblatt process, and $H = d + 1/2$.

The next theorem follows from Theorem 4.15 and a reduction principle along the lines of Theorem 4.9. We assume that the innovations in the linear process are symmetric.

Theorem 4.17 *Assume that X_t $(t \in \mathbb{Z})$ is a linear process such that $a_j \sim c_a j^{d-1}$, $d \in (-\infty, 1/2)$, ε_t $(t \in \mathbb{Z})$ are i.i.d. symmetric random variables such that (4.89) holds with $\alpha \in (1, 2)$ and $\beta = 0$, i.e.*

$$P(\varepsilon_1 > x) \sim \frac{A}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim \frac{A}{2}x^{-\alpha}.$$

Furthermore, assume that the distribution F_ε of ε_1 fulfills

$$\bigl|F_\varepsilon^{(2)}(x)\bigr| \le C\bigl(1 + |x|\bigr)^{-\alpha}, \qquad \bigl|F_\varepsilon^{(2)}(x) - F_\varepsilon^{(2)}(y)\bigr| \le C|x - y|\bigl(1 + |x|\bigr)^{-\alpha},$$

where $|x - y| < 1$, $x \in \mathbb{R}$.

- *If $0 < d < 1 - 1/\alpha$ and G is bounded, then*

$$n^{-H}\sum_{t=1}^{[nu]}\bigl\{G(X_t) - E\bigl[G(X_1)\bigr]\bigr\} \Rightarrow A^{1/\alpha}C_\alpha^{-1/\alpha}\frac{c_a}{d}G_\infty^{(1)}(0)\tilde{Z}_{H,\alpha}(u), \tag{4.90}$$

where $\Rightarrow$ denotes weak convergence in $D[0,1]$, and $\tilde{Z}_{H,\alpha}(\cdot)$ is a linear fractional stable motion with $H = d + \alpha^{-1}$ such that $\tilde{Z}_{H,\alpha}(1)$ is a symmetric α-stable random variable with scale

$$\eta = \left(\int_{-\infty}^{1} \left\{ (1-v)_+^d - (-v)_+^d \right\}^\alpha dv \right)^{1/\alpha}$$

and $G_\infty(x) = E[G(X+x)]$.

- *If $1 - 2/\alpha < d < 0$ and $A = 1$ in (4.89) and G is bounded, then*

$$n^{-1/\alpha(1-d)} \sum_{t=1}^{[nu]} \left\{ G(X_t) - E\left[G(X_1)\right] \right\} \Rightarrow c_G^+ \tilde{Z}^+_{\alpha(1-d)}(u) + c_G^- \tilde{Z}^-_{\alpha(1-d)}(u), \quad (4.91)$$

where $\tilde{Z}^+_{\alpha(1-d)}(\cdot)$, $\tilde{Z}^-_{\alpha(1-d)}(\cdot)$ are independent copies of an $\alpha(1-d)$-stable Lévy motion such that $Z_{\alpha(1-d)}(1) \overset{d}{=} S_{\alpha(1-d)}(1,1,0)$ and

$$c_G^\pm = C_{\alpha(1-d)}^{-1/\alpha(1-d)} \frac{c_a^{1/(1-d)}}{1-d} \int_0^\infty \left[G_\infty(\pm v) - G_\infty(0) \right] v^{-1-1/(1-d)} \, dv,$$

where $G_\infty(x) = E[G(X_1 + x)]$.

- *If $-\infty < d < 1 - 2/\alpha$ and G is bounded, then*

$$n^{-1/2} \sum_{t=1}^{[nu]} \left\{ G(X_t) - E\left[G(X_1)\right] \right\} \Rightarrow \sigma_S B(u), \quad (4.92)$$

where $B(\cdot)$ is a standard Brownian motion, and σ_S is a finite positive constant.

This theorem was proven in Koul and Surgailis (2001), Surgailis (2002) and Hsing (1999). Remarkably, in (4.90) and (4.91), we may obtain a stable limit arising from a summation of bounded random variables. The convergence in (4.90) can be thought of as a *long-memory-type behaviour* since the scaling involves the memory parameter d and the limiting process has dependent increments. The convergence in (4.91) is a sort of an *intermediate case*: the scaling involves d, but the limiting process has independent increments. Finally, (4.92) represents a *standard behaviour*: as in the i.i.d. case, the limiting process is a Brownian motion since $\operatorname{var}(G(X_1))$ is finite.

Below, we give an outline of the proof of (4.90). As for (4.91), the limiting process has independent increments, but the scaling factor involves the memory parameter d. The reason for this is that the process $S_{n,G}(u)$ can be approximated by a sum $\sum_{t=1}^n \eta_G(\varepsilon_t)$ of i.i.d. random variables, where

$$\eta_G(\varepsilon_t) = \sum_{j=0}^\infty \left\{ G_\infty(a_j \varepsilon_t) - E\left[G_\infty(a_j \varepsilon_1)\right] \right\},$$

and the variables η_G have a tail decaying like $|x|^{-\alpha(1-d)}$.

In (4.90) it may happen that the quantity $G_\infty^{(1)}(0)$ vanishes. It is an open question, whether it is possible to obtain a nondegenerate limit in this case with $1<\alpha<2$. Let us recall that in the case of linear processes with finite moments the solution to this problem is given for example in Theorem 4.4. In the case of infinite moments, this question was studied in Surgailis (2004) under the assumption $2<\alpha<4$. It may happen that the limit is an $\alpha(1-d)$-Lévy stable motion, Hermite–Rosenblatt process or Brownian motion.

Proof of Theorem 4.17 Recall the notation from the proof of Theorem 4.9. We denote by $\mathcal{V}_t$ the sigma field generated by $(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and set

$$T_n(G;1)=\sum_{t=1}^{n}\bigl(G(X_t)-E\bigl[G(X_1)\bigr]-G_\infty^{(1)}(0)X_t\bigr)$$

and $P_K Y = E(Y|\mathcal{V}_K)-E(Y|\mathcal{V}_{K-1})$. We can repeat the computation there, using the rth norm with $r<\alpha$ instead of $r=2$:

$$\begin{aligned}\bigl\|T_n(G;1)\bigr\|_r^r &\le 2\sum_{K=-\infty}^{n}\Biggl\|\sum_{t=\max\{K,1\}}^{n}P_K U(\mathcal{V}_t)\Biggr\|_r^r\\ &\le \sum_{K=-\infty}^{n}\left(\sum_{t=\max\{K,1\}}^{n}\bigl\|P_{-(t-K)}U(\mathcal{V}_0)\bigr\|_r\right)^r.\end{aligned}$$

The first inequality follows from a result for martingale differences Y_t $(t\in\mathbb{N})$, namely

$$\Biggl\|\sum_{t=1}^{n}Y_t\Biggr\|_r^r\le 2\sum_{t=1}^{n}\|Y_j\|_r^r$$

for any $1\le r\le 2$. The second one is the norm inequality used in the proof of Theorem 4.9. Now, instead of Lemma 4.17, we use

$$\bigl\|P_{-(t-K)}U(\mathcal{V}_0)\bigr\|_r\le (t-K)^{-(1-d)(1+\gamma)},$$

where $(1+\gamma)r<\alpha$. Computations leading to this expression are quite involved; we refer the reader to Koul and Surgailis (2001). Then one obtains

$$\bigl\|T_n(G;1)\bigr\|_r^r\le C\sum_{K=-\infty}^{n}\left(\sum_{t=K\vee 1}^{n}(t-K)^{-(1-d)(1+\gamma)}\right)^r\le Cn^{r+1}n^{-(1-d)(1+\gamma)r}$$

by similar calculations as those leading to (4.64), (4.65). Choosing γ sufficiently close to 0, we conclude that

$$\bigl\|T_n(G;1)\bigr\|_r^r=o\bigl(n^{r(d+1/\alpha)}\bigr).$$

In particular, $\|T_n(G;1)\|_r^r = o(v_n^r)$, where

$$v_n = C_\alpha^{-1/\alpha} A^{1/\alpha} \frac{c_a}{d} n^H$$

with $H = d + \frac{1}{\alpha}$. Therefore, on account of Theorem 4.15, the limiting behaviour of

$$v_n^{-1} \sum_{t=1}^{n} \{G(X_t) - E[G(X_1)]\}$$

is the same as that of $v_n^{-1} G_\infty^{(1)}(0) \sum_{t=1}^{n} X_t$. □

4.3.4 Stochastic Volatility Models

In this section we consider Long-Memory Stochastic Volatility (LMSV) sequences with infinite moments. Let $X_t = \sigma_t \xi_t$ $(t \in \mathbb{N})$, where

$$\sigma_t = \sigma(\zeta_t), \quad \zeta_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j},$$

$\sigma(\cdot)$ is a positive function, $\sum_{j=1}^{\infty} a_j^2 < \infty$, and ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables. It is further assumed that ξ_t $(t \in \mathbb{Z})$ is a sequence of i.i.d. random variables such that

$$P(\xi_1 > x) \sim A \frac{1+\beta}{2} x^{-\alpha}, \qquad P(\xi_1 < -x) \sim A \frac{1-\beta}{2} x^{-\alpha}. \tag{4.93}$$

Also, we assume that the sequences ε_t $(t \in \mathbb{Z})$ and ξ_t $(t \in \mathbb{Z})$ are mutually independent. At the moment we do not assume anything about the mean of ξ_t.

Limiting results for infinite-variance volatility models with long memory are almost non-existing; see Kulik and Soulier (2012) or Surgailis (2008); the latter in a quadratic LARCH case. In particular, we will show below that stochastic volatility models can be treated using a point process methodology.

4.3.4.1 Tail Behaviour

The first question we have to answer is the following. If ξ is like in (4.93), what is the consequence on the tail of X? The next lemma shows that if the random variables ε and σ are independent, then $\sigma\varepsilon$ is still regularly varying. The result is often referred to as Breiman's lemma (Breiman 1965), and a proof can be found for example in Resnick (2007, Proposition 7.5).

Lemma 4.20 *Assume that* (4.93) *holds. If* σ_1 *is a positive random variable independent of* ξ_1 *and such that for some* $\delta > 0$,

$$E\left(\sigma_1^{\alpha+\delta}\right) < \infty, \tag{4.94}$$

then the distribution of $\sigma\xi$ *is regularly varying, and*

$$\lim_{x\to\infty} \frac{P(\sigma_1\xi_1 > x)}{P(|\xi_1| > x)} = \frac{1+\beta}{2} E\left(\sigma_1^{\alpha}\right), \qquad \lim_{x\to\infty} \frac{P(\sigma_1\xi_1 < -x)}{P(|\xi_1| > x)} = \frac{1-\beta}{2} E\left(\sigma_1^{\alpha}\right). \tag{4.95}$$

Lemma 4.20 implies for the LMSV model and arbitrary $p > 0$ that

$$P\left(|X_1|^p > x\right) = P\left(X_1 > x^{1/p}\right) + P\left(X_1 < -x^{1/p}\right) \sim A\ E\left(\sigma_1^{\alpha}\right) x^{-\alpha/p}. \tag{4.96}$$

Thus, if we consider the LMSV model, we may take ξ_t as in (4.93), $\sigma(x) = e^x$ and ζ_t $(t \in \mathbb{N})$ to be e.g. long-memory Gaussian. Then the random variables X_t $(t \in \mathbb{N})$ have heavy tails and long memory.

4.3.4.2 Point Process Convergence

Point process convergence results play a crucial role when proving asymptotic results for partial sums based on infinite-variance sequences. Here, we assume that the reader is familiar with material presented in Sect. 4.3.2.4.

We start with a simple generalization of Theorem 4.13 to the LMSV model. Recall the intensity measure

$$d\lambda(x) = \alpha\left[\frac{1+\beta}{2} x^{-(\alpha+1)} 1\{0 < x < \infty\} + \frac{1-\beta}{2}(-x)^{-(\alpha+1)} 1\{-\infty < x < 0\}\right] dx,$$

where $\beta \in [-1, 1]$, and consider the point processes

$$N_n = \sum_{t=1}^{n} \delta_{(t/n, c_n^{-1} X_t)},$$

where c_n is chosen to fulfill $P(|\xi_1| > c_n) \sim n^{-1}$, i.e.

$$c_n = A^{1/\alpha} n^{1/\alpha}.$$

The next result shows that the point process based on the LMSV sequence X_t behaves as if the random variables were independent. It will be clear from the proof that the same applies to $|X_t|^r$ where r is any power. Furthermore, we do not really need the particular structure $\sigma_t = \sigma(\zeta_t)$, where ζ_t $(t \in \mathbb{Z})$ is a linear process. Only the ergodicity of σ_t $(t \in \mathbb{N})$ is needed.

Theorem 4.18 *Consider the LMSV model $X_t = \sigma_t \xi_t$ $(t \in \mathbb{N})$ such that (4.93) and Breiman's condition (4.94) hold. Then N_n converges weakly in $M_p([0,1] \times \mathbb{R})$ to a Poisson process N with intensity measure $E(\sigma_1^\alpha) ds \times d\lambda(x)$.*

Proof (Personal communication with P. Soulier) The proof is basically the same as in the i.i.d. case, see Theorem 4.13. We also use the same notation as in Theorem 4.13. Let $U = \bigcup_{i=1}^K (k_i, l_i) \times (s_i, t_i)$. Then

$$P\big(N_n(U) = 0\big) = P\left(\sum_{i=1}^K \sum_{nk_i < t < nl_i}^n 1\big\{c_n^{-1} X_t \in (s_i, t_i)\big\} = 0\right)$$
$$= E\left[\prod_{i=1}^K \prod_{nk_i < t < nl_i}^n P\big(c_n^{-1} X_t \notin (s_i, t_i) | \mathcal{F}_\sigma\big)\right] =: mE[P_n],$$

where $\mathcal{F}_\sigma$ is the sigma field generated by the entire sequence σ_t. Let $\theta_t((s_i, t_i))$ be the limit of $nP(c_n^{-1} X_t \in (s_i, t_i) | \mathcal{F}_\sigma)$ and write

$$Q_n = \prod_{i=1}^K \prod_{nk_i < t < nl_i} \exp\big\{-n^{-1} \theta_t\big((s_i, t_i)\big)\big\}.$$

Note that θ_t is a random variable since it depends on the sequence σ_t $(t \in \mathbb{N})$. Therefore, the only difference between the LMSV setting and the i.i.d. one is that Q_n here is a random variable and $\lambda((s_i, t_i))$ is replaced by $\theta_t((s_i, t_i))$. Nevertheless, Q_n converges in probability to

$$\exp\left\{-E\big(\sigma_1^\alpha\big) \sum_{i=1}^K (l_i - k_i) \lambda\big((s_i, t_i)\big)\right\} = P\big(N(U) = 0\big).$$

It remains to prove that $|P_n - Q_n|$ converges in probability to 0 and apply the bounded convergence theorem. To prove that $|P_n - Q_n| \to_P 0$, we proceed as in Theorem 4.13:

$$E|P_n - Q_n|$$
$$\le \sum_{i=1}^K (l_i - k_i) E\big[\big|nP\big(c_n^{-1} X_1 \in (s_i, t_i) | \mathcal{F}_\sigma\big) - \theta_1\big((s_i, t_i)\big)\big|\big]$$
$$+ \sum_{i=1}^K n(l_i - k_i) E\left[\left|1 - e^{-n^{-1}\theta_1((s_i, t_i))} - \frac{\theta_1((s_i, t_i))}{n}\right|\right].$$

For the second term, we have

$$nE\left[\left|1 - e^{-n^{-1}\theta_1((s_i, t_i))} - \frac{\theta_1((s_i, t_i))}{n}\right|\right] \le Cn^{-\delta} E\big[\sigma_1^{\alpha+\delta}\big].$$

Furthermore, let us recall the so-called Potter's bound (see Theorem 1.5.6. in Bingham et al. 1989), namely: for $v > 0$,

$$nP\big(c_n^{-1} v\xi_1 \in (s_i, t_i)\big) \le C\big(\max\{v, 1\}\big)^{\alpha+\delta},$$

where $\delta > 0$. For the first term, we apply Potter's bound to get

$$nP\big(c_n^{-1} X_1 \in (s_i, t_i)|\mathcal{F}_\sigma\big) = nP\big(c_n^{-1} \xi_1\sigma_1 \in (s_i, t_i)|\mathcal{F}_\sigma\big) \le \big(\max\{\sigma_1, 1\}\big)^{\alpha+\delta},$$

and the same bound holds for $\theta_1(s_i, t_i)$. We then can apply bounded convergence to get

$$\lim_{n\to\infty} E\big[\big|nP\big(c_n^{-1} X_1 \in (s_i, t_i)\big) - \theta_1\big((s_i, t_i)\big)\big|\big] = 0. \qquad \square$$

4.3.4.3 Convergence of Partial Sums

Having established point process convergence, we proceed with its consequences for partial sums. Assume that ξ_1 fulfills (4.93) and $E(\xi_1) = 0$ or ξ_1 is symmetric if $\alpha \in (0, 1)$. Define

$$S_n(u) = \sum_{t=1}^{[nu]} X_t$$

and

$$S_{n,p}(u) = \sum_{t=1}^{[nu]} \big(|X_t|^p - E\big[|X_1|^p\big]\big),$$

assuming that $E[|X_1|^p] < \infty$ but $E[|X_1|^{2p}] = \infty$. Due to Lemma 4.20, this is achieved when $p < \alpha < 2p$. In the next theorem we show that depending on an interplay between long memory and tails, partial sums based on the LMSV sequence may converge either to a Lévy process (weakly dependent behaviour) or to a Hermite process (long-memory behaviour).

Theorem 4.19 *Consider the LMSV model $X_t = \sigma_t\xi_t$ ($t \in \mathbb{N}$) and assume that the conditions of Theorem 4.18 hold. In addition, we assume that $\alpha > 1$, $E(\xi_1) = 0$ and ζ_t ($t \in \mathbb{N}$) is a Gaussian linear process with coefficients a_j satisfying* (B1), *i.e. $a_j = L_a(j)j^{d-1}$, $d \in (0, 1/2)$, and covariance function $\gamma_\zeta(k) \sim L_\gamma(k)k^{2d-1}$. Let $m \ge 1$ be the Hermite rank of the function $\sigma^p(\cdot)$ and assume further that $E(\sigma_1^{2\alpha+2\varepsilon}) < \infty$.*

- *If $1 < \alpha < 2$, then*

$$n^{-1/\alpha} S_n(u) \Rightarrow A^{1/\alpha} C_\alpha^{-1/\alpha} \big(E\big[\sigma_1^\alpha\big]\big)^{1/\alpha} Z_\alpha(u), \tag{4.97}$$

  where $Z_\alpha(\cdot)$ is an α-stable Lévy process such that $Z_\alpha(1) \overset{\mathrm{d}}{=} S_\alpha(1, \beta, 0)$, and $\Rightarrow$ denotes weak convergence in $D[0, 1]$.

- *If $p < \alpha < 2p$ and $1 - m(1/2 - d) < p/\alpha$, then*

$$n^{-p/\alpha} S_{n,p}(u) \Rightarrow A^{p/\alpha} C_{\alpha/p}^{-p/\alpha} \big(E\big[\sigma_1^\alpha\big]\big)^{p/\alpha} Z_{\alpha/p}(u), \tag{4.98}$$

where $Z_\alpha(\cdot)$ is an α/p-stable Lévy process such that $Z_\alpha(1) \overset{d}{=} S_{\alpha/p}(1,1,0)$, and $\Rightarrow$ denotes weak convergence in $D[0,1]$.

- *If $p < \alpha < 2p$ and $1 - m(1/2 - d) > p/\alpha$, then*

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) S_{n,p}(u) \Rightarrow \frac{J(m)E[|\xi_1|^p]}{m!} Z_{m,H}(u), \tag{4.99}$$

where $Z_{m,H}(\cdot)$ is a Hermite process of order m, $H = d + \frac{1}{2}$,

$$L_m(n) = m! C_m L_\gamma(n),$$

$J(m)$ is the Hermite coefficient of $\sigma^p(\cdot)$, and $\Rightarrow$ denotes weak convergence in $D[0,1]$.

When $\alpha \in (1,2)$, the partial sum $S_n(u)$ is a martingale because $E(X_t) = E(\xi_t)E(\sigma_t) = 0$. Hence, only the stable Lévy limit arises, and (4.97) holds. This can be concluded from a general theory by Surgailis (2008). If $S_{n,p}(\cdot)$ is considered, then we observe a dichotomous behaviour. Assume for simplicity that $m = 1$. If long memory is strong enough, then it influences the limiting behaviour. Interestingly, the infinite variance sequence $|X_t|^p$ yields a limiting process with finite variance. Furthermore, results are readily extendable to the case where ζ_t is a general linear process. Instead of Theorem 4.4, one has to use corresponding results for subordinated linear processes; see Theorem 4.6. Furthermore, in contrast to Theorem 4.15 for linear processes with infinite variance, we note that we have weak convergence w.r.t. J_1-topology in all three cases.

Example 4.16 (Cf. Example 4.11) Assume that $X_t = \xi_t \exp(\zeta_t)$, where ζ_t is a standard normal sequence with covariance $\gamma_\zeta(k) \sim L_\gamma k^{2d-1}$, $d \in (0, 1/2)$. If $\alpha \in (2,4)$ and $d + 1/2 < 2/\alpha$, then $n^{-2/\alpha} S_{n,2}(u)$ converges to a Lévy process. Otherwise, if $\alpha \in (2,4)$ and $d + 1/2 > 2/\alpha$, then

$$n^{-(1/2+d)} L_1(n)^{-1/2}(n) S_{n,2}(u) \Rightarrow J(1)E\big(\xi_1^2\big) B_H(u),$$

where $L_1(n) = (d(2d+1))^{-1} L_\gamma(n)$ and $J(1) = E[\zeta_1 \exp(2\zeta_1)]$.

In the spirit of Example 4.12, if $\alpha \in (1,2)$ and $E(\xi_t) \neq 0$, then long memory appears already in $\sum_{t=1}^{[nu]} X_t$.

Example 4.17 (LMSD with Infinite Variance) As in Example 4.12, we assume that the random variables ξ_t ($t \in \mathbb{N}$) are strictly positive. Suppose that we have heavy tails

$$P(\xi_1 > x) \sim A x^{-\alpha} \quad (x \to \infty)$$

with $\alpha \in (1,2)$. Furthermore, it is assumed that the sequences ξ_t and ζ_t are independent and the covariance of ζ_t is of the asymptotic form $\gamma_\zeta(k) \sim L_\gamma(k)k^{2d-1}$, $d \in (0,1/2)$. Let $G(x) = x$ and $\sigma(x) = \exp(x)$, so that the Hermite rank $m = 1$. Then we have a dichotomous behaviour for $S_n(u) := \sum_{t=1}^{[nu]}(X_t - E(X_1))$. Specifically, (4.98) and (4.99) hold with $p = 1$:

- If $1/2 + d < 1/\alpha$, then

$$n^{-1/\alpha} S_n(u) \Rightarrow A^{1/\alpha} C_\alpha^{-1/\alpha} \left(E\left[\sigma_1^\alpha\right]\right)^{1/\alpha} Z_\alpha(u), \tag{4.100}$$

  where $Z_\alpha(\cdot)$ is an α-stable Lévy process such that $Z_\alpha(1) \overset{\mathrm{d}}{=} S_\alpha(1,1,0)$.
- If $1/2 + d > 1/\alpha$, then

$$n^{-(1/2+d)} L_1^{-1/2}(n) S_n(u) \Rightarrow J(1) E[\xi_1] B_H(u), \tag{4.101}$$

  where $B_H(\cdot)$ is a fractional Brownian motion, $H = d + \frac{1}{2}$, $L_1(n) = C_1 L_\gamma(n)$ and $J(1) = E[\zeta_1 \exp(\zeta_1)]$.

Proof of Theorem 4.19 Let $\mathcal{F}_t$ be a sigma field generated by ξ_j, ε_j $(j \le t)$. We start by studying $S_{n,p}(\cdot)$. Write

$$\sum_{t=1}^{[nu]} \left(|X_t|^p - E\left[|X_t|^p\right]\right) = \sum_{t=1}^{[nu]} \left(|X_t|^p - E\left[|X_t|^p | \mathcal{F}_{t-1}\right]\right)$$
$$+ \sum_{t=1}^{[nu]} \left(E\left[|X_t|^p | \mathcal{F}_{t-1}\right] - E\left[|X_1|^p\right]\right) =: M_n(u) + R_n(u).$$

Note that $E[|X_t|^p | \mathcal{F}_{t-1}] = E(|\xi_1|^p)\sigma^p(\zeta_t)$ is a function of ζ_t and does not depend on ξ_t. Therefore, for the long-memory part $R_n(u)$, we have

$$n^{-(1-m(\frac{1}{2}-d))} L_1^{-1/2}(n) R_n(u) \Rightarrow \frac{J(m) E[|\xi_1|^p]}{m!} Z_{m,H}(u) \tag{4.102}$$

if $m(1/2 - d) < 1$, where $Z_{m,H}(\cdot)$ is a Hermite–Rosenblatt process, and L_1 is a slowly varying function defined in Theorem 4.4. If $m(1/2 - d) > 1$, then

$$n^{-1/2} R_n(u) \Rightarrow v E\left[|\xi_1|^p\right] B(u), \tag{4.103}$$

where $B(\cdot)$ is a standard Brownian motion, and v is a constant.

We will show that under the assumptions we have,

$$c_n^{-p} M_n(u) \Rightarrow C_{\alpha/p}^{-p/\alpha} \left(E\left[\sigma_1^\alpha\right]\right)^{p/\alpha} Z_{\alpha/p}(u), \tag{4.104}$$

or equivalently,

$$n^{-1/\alpha} M_n(u) \Rightarrow A^{p/\alpha} C_{\alpha/p}^{-p/\alpha} \left(E\left[\sigma_1^\alpha\right]\right)^{p/\alpha} Z_{\alpha/p}(u).$$

From (4.102), (4.103) and (4.104) we conclude the proof of the theorem. First we prove (4.104). The proof is very similar to the proof of convergence of the partial sum of an i.i.d. sequence in the domain of attraction of a stable law to a Lévy stable process. The difference consists of some additional technicalities (see e.g. the proof of Theorem 71 in Resnick 2007 for additional details).

Step 1: For $0 < \varepsilon < 1$, decompose $M_n(u)$ further as

$$\begin{aligned}M_n(u) &= \sum_{t=1}^{[nu]}\Big(|X_t|^p 1\big\{|X_t| < \varepsilon c_n\big\} - E\big[|X_t|^p 1\big\{|X_t| < \varepsilon c_n\big\}|\mathcal{F}_{t-1}\big]\Big)\\ &\quad + \sum_{t=1}^{[nu]}\Big(|X_t|^p 1\big\{|X_t| > \varepsilon c_n\big\} - E\big[|X_t|^p 1\big\{|X_t| > \varepsilon c_n\big\}|\mathcal{F}_{t-1}\big]\Big)\\ &=: M_n^{(\varepsilon)}(u) + \tilde{M}_n^{(\varepsilon)}(u).\end{aligned}$$

The term $\tilde{M}_n^{(\varepsilon)}(\cdot)$ is treated using point process convergence. It excludes *small jumps* X_t defined by $c_n^{-1}|X_t| < \varepsilon$. The reason for this is that the summation functional is not continuous on the entire real line; one has to exclude small jumps. For any $\varepsilon > 0$, the summation functional is an almost surely (with respect to the distribution of the Poisson point process, see e.g. p. 215 in Resnick 2007) continuous mapping from the set of Radon measures on $[0, 1] \times [\varepsilon, \infty)$ to $D([0, 1], \mathbb{R})$. From Theorem 4.18 we then conclude

$$c_n^{-p}\sum_{t=1}^{[nu]}|X_t|^p 1\big\{|X_t| > \varepsilon c_n\big\} \Rightarrow \sum_{k:t_k\le u}|j_k|^p 1\big\{|j_k| > \varepsilon\big\} \tag{4.105}$$

in $([0, 1], \mathbb{R})$, where we recall that (t_k, j_k) are points of the limiting Poisson process. Taking expectations in (4.105), we obtain

$$\lim_{n\to\infty}[nu]c_n^{-p}E\big[|X_1|^p 1\big\{|X_1| > \varepsilon c_n\big\}\big] = u\int_{|x|>\varepsilon}|x|^p\, d\lambda(x)$$

uniformly with respect to $u \in [0, 1]$, since this is a sequence of increasing functions with a continuous limit. Furthermore, we claim that

$$c_n^{-p}\left|\sum_{t=1}^{[nu]}\Big(E\big[|X_1|^p 1\big\{|X_1| > \varepsilon c_n\big\}\big] - E\big[|X_t|^p 1\big\{|X_t| > \varepsilon c_n\big\}\big|\mathcal{F}_{t-1}\big]\Big)\right| \overset{\mathrm{p}}{\to} 0$$

uniformly in $u \in [0, 1]$. The variance of the last expression is in fact bounded by

$$\begin{aligned}&c_n^{-2p}[nu]^2\gamma_\zeta^m\big([nu]\big)\operatorname{var}\Big(E\big[|X_1|^p 1\big\{|X_1| > \varepsilon c_n\big\}\big|\mathcal{F}_0\big]\Big)\\ &\quad\le c_n^{-2p}[nu]^2\gamma_\zeta^m\big([nu]\big)E\Big[E^2\big[|X_1|^p 1\big\{|X_1| > \varepsilon c_n\big\}\big|\mathcal{F}_0\big]\Big],\end{aligned}$$

where $\gamma_\zeta(k)$ is the covariance function of the Gaussian sequence ζ_t $(t \in \mathbb{Z})$, and m is the Hermite rank of $\sigma^p(\cdot)$. Recall Potter's bound (see Theorem 1.5.6. in Bingham et al. 1989): for $v > 0$,

$$nP\left(c_n^{-1} v\xi_1 \in (s_i, t_i)\right) \le C\left(\max\{v, 1\}\right)^{\alpha+\delta},$$

where $\delta > 0$. Now, if $p < \alpha < 2p$, then we combine Karamata's theorem with Potter's bound to obtain

$$\begin{aligned} E\left[\sigma^p(x)|\xi_1|^p 1\left\{\left|\sigma(x)\xi_1\right| > \varepsilon c_n\right\}\right] &\le Cn^{-1}c_n^p \frac{\bar{F}_\xi(\varepsilon c_n/\sigma(x))}{\bar{F}_\xi(c_n)} \\ &\le Cn^{-1}c_n^p \sigma^{\alpha+\varepsilon}(x). \end{aligned}$$

Since by assumption $E[\sigma_1^{2\alpha+2\varepsilon}] < \infty$ for some $\varepsilon > 0$, we have for each t,

$$\begin{aligned} &\operatorname{var}\left(c_n^{-p}\sum_{j=1}^{[nu]}\left\{E\left[|X_0|^p 1\left\{|X_0| > \varepsilon c_n\right\}\right] - E\left[|X_t|^p 1\left\{|X_t| > \varepsilon c_n\right\}\middle|\mathcal{F}_{j-1}\right]\right\}\right) \\ &\quad \le Cn^{-2}[nu]^2\gamma_\zeta\left([nu]\right) \le Cn^{2-2H+\varepsilon}u^{2H-\varepsilon}, \end{aligned} \tag{4.106}$$

where the last bound is obtained for some $\varepsilon > 0$ by Potter's bound. This proves the convergence of finite-dimensional distributions to 0 and tightness in $D([0, 1])$. We now argue that the bounds obtained above imply

$$c_n^{-p}\tilde{M}_n^{(\varepsilon)}(u) \Rightarrow C_{\alpha/p}^{-p/\alpha}\left(E\left[\sigma_1^\alpha\right]\right)^{p/\alpha} Z_{\alpha/p}^{(\varepsilon)}(u)$$

and also $Z_{\alpha/p}^{(\varepsilon)}(u) \Rightarrow Z_{\alpha/p}(u)$ as $\varepsilon \to 0$. Therefore, it is suffices to show the negligibility of $c_n^{-p}M_n^{(\varepsilon)}$, i.e. that small jumps are negligible. By Doob's martingale inequality we obtain

$$\begin{aligned} &E\left[\left(\sup_{u\in[0,1]} c_n^{-p}\sum_{t=1}^{[nu]}\left\{|X_t|^p 1\left\{|X_t| < \varepsilon c_n\right\} - E\left[|X_t|^p 1\left\{|X_t| < \varepsilon c_n\right\}\middle|\mathcal{F}_{t-1}\right]\right\}\right)^2\right] \\ &\quad \le Cnc_n^{-2p}E\left[\left(|X_1|^p 1\left\{|X_1| < \varepsilon c_n\right\} - E\left[|X_1|^p 1\left\{|X_1| < \varepsilon c_n\right\}\middle|\mathcal{F}_0\right]\right)^2\right] \\ &\quad \le 4Cnc_n^{-2p}E\left[\left(|X_1|^{2p} 1\left\{|X_1| < \varepsilon c_n\right\}\right)\right]. \end{aligned}$$

Recall that $\alpha < 2p$. By Karamata's theorem (Lemma 4.18),

$$E\left[|X_1|^{2p}1\left\{|X_1| < \varepsilon c_n\right\}\right] \sim \frac{2\alpha}{2p-\alpha}(\varepsilon c_n)^{2p}\bar{F}_X(\varepsilon c_n) \sim \frac{2\alpha}{2p-\alpha}\varepsilon^{2p-\alpha}c_n^{2p}n^{-1}.$$

Applying this and letting $\varepsilon \to 0$, we conclude that $c_n^{-p}M_n^{(\varepsilon)}$ is uniformly negligible in L^2 and therefore also in probability. Thus,

$$c_n^{-p}M_n(u) \Rightarrow C_{\alpha/p}^{-p/\alpha}\left(E\left[\sigma_1^\alpha\right]\right)^{p/\alpha} Z_{\alpha/p}(u).$$

This finishes the proof of (4.98) and (4.99).

As for the sum S_n, the long-memory part R_n vanishes since $E(X_1) = E(\xi_1)E(\sigma_1) = 0$. Thus, in this case also only the stable limit arises. □

The reader is referred to Kulik and Soulier (2012) for more discussion, a detailed proof and extensions to stochastic volatility with leverage.

4.3.5 Subordinated Gaussian Processes with Infinite Variance

Previously (see Theorem 4.16 or Theorem 4.19, Eq. (4.99)) we have seen that it is possible to obtain limiting distributions with finite variance although we start with innovations with infinite second moments. In this section we illustrate that this type of behaviour can also be achieved in the context of Gaussian subordination with infinite variance. This rather peculiar result depends on specific circumstances to be explained below.

Let X_t $(t \in \mathbb{Z})$ be a stationary centred Gaussian process with covariance $\gamma_X(K) \sim L_\gamma(k)k^{2d-1}$, $d \in (0, 1/2)$. Assume that G is a function such that, as $x \to \infty$,

$$P\big(G(X_1) > x\big) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P\big(G(X_1) < -x\big) \sim A\frac{1-\beta}{2}x^{-\alpha}, \quad (4.107)$$

where $\beta \in [-1, 1]$. If $\alpha \in (0, 2)$, then $G(X_t)$ have infinite (or non-existing) variance. Furthermore, if $\alpha \in (0, 1)$, then $E(|G(X_1)|) = +\infty$. A typical example is $G(x) = |x|^{-1/\alpha}$. After the transformation $|x|^{-1/\alpha}$ the mass from zero is "sent" to infinity (since for a standard normal density, $\phi(0) \neq 0$). Another example is $G(x) = b\exp(cx^2)$ for some constants $b \in \mathbb{R}$ and $c > 0$.

In this section we shall assume that $\alpha \in (1, 2)$. Again we consider

$$S_{n,G}(u) = \sum_{t=1}^{[nu]} \big\{G(X_t) - E\big[G(X_1)\big]\big\}.$$

With a similar trick as in the proof of Theorem 4.19, i.e. the decomposition into a martingale and a long-memory part, $S_{n,G}$ will be studied using techniques available for weakly dependent processes with infinite variance (see $M_n(\cdot)$ in the proof of Theorem 4.19) and finite-variance subordinated Gaussian processes (see Sect. 4.2.3). This method was used in Sly and Heyde (2008) for $\alpha \in (1, 2)$. The result for $\alpha \in (0, 1)$ was proven in Davis (1983).

4.3.5.1 Point Process Convergence

Assume that $\alpha \in (1, 2)$, so that $\operatorname{var}(G(X_t)) < \infty$. As in case of the LMSV model, we start with the convergence of point processes

$$N_n = \sum_{t=1}^{n} \delta_{(t/n, c_n^{-1}G(X_t))},$$

where in the present context

$$c_n = \inf\{x : P(|G(X_1)| > x) \le n^{-1}\}.$$

Recall that

$$d\lambda(x) = \alpha\left[\frac{1+\beta}{2}x^{-(\alpha+1)}1\{0 < x < \infty\} + \frac{1-\beta}{2}(-x)^{-(\alpha+1)}1\{-\infty < x < 0\}\right].$$

We state the following result without proof. In principle, as in the LMSV case, it says that the random variables $G(X_t)$ behave as if they were independent.

Theorem 4.20 *Consider a Gaussian sequence X_t ($t \in \mathbb{N}$) and a real-valued function G such that (4.107) holds. Then N_n converges weakly in $M_p([0,1] \times \mathbb{R})$ to a Poisson process N with intensity measure $ds \times d\lambda(x)$.*

4.3.5.2 Hypercontraction Principle for Gaussian Random Variables

We shall explain how it is possible to obtain a finite-variance random variable from infinite-variance variables $G(X_t)$. Recall that for a function G such that $E[G^2(X_1)] < \infty$, we have the following expansion:

$$G(x) = E[G(X_1)] + \sum_{l=m}^{\infty} \frac{J(l)}{l!} H_l(x),$$

where m is the Hermite rank of G, and $J(l) = E[G(X_1)H_l(X_1)]$. This expansion is also valid for a function G with $E[|G(X_1)|^{1+\theta}] < \infty$, where $\theta \in (0, 1)$. Indeed, the Hermite coefficients $J(l)$ are still well defined. Applying the Hölder inequality, we obtain with $r = (1+\theta)/\theta$,

$$|J(l)| \le E^{\frac{1}{1+\theta}}\left[|G(X_1)|^{1+\theta}\right] E^{\frac{1}{r}}\left[|H_l(X_1)|^r\right] = \|G\|_{1+\theta}\|H_l\|_r < \infty, \tag{4.108}$$

where $\|G\|_r^r = \int G^r(u)\phi(u)\,du$. Now, let $X = a_1X_1 + \theta X_2$, where $a_1^2 + \theta^2 = 1$, and X_1, X_2 are independent standard normal random variables. Let $\mathcal{F}$ be the sigma field generated by X_2. We will argue below that although $E[G^2(X)] = +\infty$, we have

$$\operatorname{var}\big(E[G(X_1)|\mathcal{F}]\big) < \infty.$$

We start with the following result.

Lemma 4.21 *Assume that $E[|G(X_1)|^{1+\theta}] < \infty$, where $\theta \in (0, 1)$. Then*

$$\sum_{l=m}^{\infty} \frac{J^2(l)}{l!}\theta^{2l} < \infty.$$

Proof From Lemma 3.1 in Taqqu (1977) we have the following bound:

$$\|H_l\|_r \le (r-1)^{l/2}\sqrt{l!}.$$

Applying (4.108) (recall that $r=(1+\theta)/\theta$), we obtain

$$\frac{J^2(l)\theta^{2l}}{l!} \le \frac{\theta^{2l}}{l!}\|G\|_{1+\theta}^2(r-1)^l l! = \theta^{2l}\|G\|_{1+\theta}^2\theta^{-l} = \|G\|_{1+\theta}^2\theta^l. \qquad \square$$

The consequence of this simple lemma is quite remarkable. Applying formula (3.16) and recalling that X_2 is $\mathcal{F}$-measurable and Hermite polynomials H_l ($l \ge 1$) are centred, we obtain

$$E\big[H_l(X)|\mathcal{F}\big] = E\big[H_l(a_1X_1+\theta X_2)|\mathcal{F}\big] = \sum_{j=0}^{l}\binom{l}{j}a_1^j\theta^{l-j}E\big[H_j(X_1)H_{l-j}(X_2)|\mathcal{F}\big]$$

$$= \sum_{j=0}^{l}\binom{l}{j}a_1^j\theta^{l-j}H_{l-j}(X_2)E\big[H_j(X_1)|\mathcal{F}\big] = \theta^l H_l(X_2).$$

We recall that $E[H_l^2(X_2)] = l!$. From Lemma 4.21 we have

$$\sum_{l=m}^{\infty}\left(\frac{J(l)}{l!}\right)^2\theta^{2l}l! < \infty.$$

This expression is however equal to

$$\operatorname{var}\left(\sum_{l=m}^{\infty}\frac{J(l)}{l!}\theta^l H_l(X_2)\right) = \operatorname{var}\left(\sum_{l=m}^{\infty}\frac{J(l)}{l!}E\big[H_l(X)|\mathcal{F}\big]\right).$$

Thus, $\sum_{l=m}^{\infty}E[H_l(X)|\mathcal{F}]J(l)/l!$ is a well-defined Hermite expansion of a function

$$\tilde{g}(X_2) := E\big[G(X)|\mathcal{F}\big] = E\big[\tilde{g}(X_2)\big] + \sum_{l=m}^{\infty}\frac{J(l)}{l!}\theta^l H_l(X_2)$$

with finite variance. Note also that, since X_2 is $\mathcal{F}$-measurable,

$$E\big[\tilde{g}(X_2)H_l(X_2)\big] = E\big\{E\big[G(X)|\mathcal{F}\big]H_l(X_2)\big\} = E\big[G(X)H_l(X_2)\big].$$

4.3.5.3 Partial Sums Convergence

Theorem 4.21 *Assume that X_t ($t \in \mathbb{Z}$) is a stationary standard normal sequence with covariance $\gamma_X(K) \sim L_\gamma(k)k^{2d-1}$, $d \in (0,1/2)$. Let G be a function with Hermite rank m such that (4.107) holds with $1<\alpha<2$.*

- *If* $1 < \alpha < 2$ *and* $1 - m(1/2 - d) < 1/\alpha$, *then*

$$n^{-1/\alpha} \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \overset{\text{f.d.}}{\to} A^{1/\alpha} C_\alpha^{-1/\alpha} Z_\alpha(u), \tag{4.109}$$

where $Z_\alpha(\cdot)$ *is an* α*-stable Lévy process such that* $Z_\alpha(1) \overset{\text{d}}{=} S_\alpha(1, \beta, 0)$.

- *If* m *is the Hermite rank of* G *and* $1 - m(\frac{1}{2} - d) > 1/\alpha$, *then*

$$n^{-(1-m(\frac{1}{2}-d))} L_m^{-1/2}(n) \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \Rightarrow Z_{m,H}(u) \quad \big(u \in [0,1]\big),$$

where $H = d + \frac{1}{2}$, $L_m(n) = m! C_m L_\gamma^m(n)$, $Z_{m,H}(u)$ *is the Hermite–Rosenblatt process, and* $\Rightarrow$ *denotes weak convergence in* $D[0, 1]$.

Proof We present just a short heuristic derivation. The Gaussian sequence can be written as a linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$, where ε_t $(t \in \mathbb{Z})$ are i.i.d. standard normal, and $\sum_{j=0}^{\infty} a_j^2 = 1$. Let $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. Then

$$\begin{aligned}
&\sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_1)]\} \\
&\quad = \sum_{t=1}^{[nu]} \{G(X_t) - E[G(X_t)|\mathcal{F}_{t-l}]\} + \sum_{t=1}^{[nu]} \{E[G(X_t)|\mathcal{F}_{t-l}] - E[G(X_1)]\} \\
&\quad =: M_n(u) + R_n(u),
\end{aligned}$$

where l is such that $\theta := \sqrt{\sum_{j=l}^{\infty} a_j^2} < \alpha - 1$. The first part $M_n(\cdot)$ is a martingale. Therefore, its limiting properties are studied in the very same way as $M_n(\cdot)$ in the proof of Theorem 4.19. As for the second part, write

$$X_t := \sum_{j=0}^{l-1} a_j \varepsilon_{t-j} + \theta \tilde{X}_{t,l},$$

where $\tilde{X}_{t,l} := \theta^{-1} \sum_{j=l}^{\infty} a_j \varepsilon_{t-j}$. The random variables $\tilde{X}_{t,l}$ $(t \in \mathbb{N})$ are standard normal. Applying Lemma 4.21, the function

$$g(\tilde{X}_{t,l}) := E[G(X_t)|\mathcal{F}_{t-l}] - E[G(X_1)]$$

has finite variance. Therefore, the convergence of the second part $R_n(u)$ follows from Theorem 4.4. □

4.3.6 Quadratic LARCH Models

We recall (cf. (2.58)) that the quadratic LARCH(∞) (or LARCH$_+$) process is the unique solution of

$$X_t = b_0\eta_t + \xi_t \sum_{j=1}^{\infty} b_j X_{t-j}, \tag{4.110}$$

where (η_t, ξ_t) $(t \in \mathbb{Z})$ is a sequence of i.i.d. random vectors. We assume that $b_j \sim c_b j^{d-1}$ $(d \in (0, 1/2))$ and that the random variables η_t are heavy tailed in the sense that

$$P\big(|\eta_1| > x\big) \sim Ax^{-\alpha}$$

for some $\alpha \in (2, 4)$. In other words, $E(\eta_1^2) < \infty$, but $E(\eta_1^4) = \infty$. Furthermore, we assume that $E(\xi_1^4 + \xi_1^2\eta_1^2) < \infty$. Surgailis (2008) considers convergence of the sum of the squares and proves that under appropriate technical assumptions we have a dichotomous behaviour as in case of the stochastic volatility model (cf. Theorem 4.19) or the subordinated Gaussian sequence with heavy tails (cf. Theorem 4.21): if $d + \frac{1}{2} < 2/\alpha$, then

$$n^{-2/\alpha} \sum_{t=1}^{[nu]} \big(X_t^2 - E\big(X_1^2\big)\big)$$

converges in a finite-dimensional sense to a Lévy process. Otherwise, if $d + \frac{1}{2} > 2/\alpha$, then

$$n^{-(d+\frac{1}{2})} \sum_{t=1}^{[nu]} \big(X_t^2 - E\big(X_1^2\big)\big)$$

converges to a fractional Brownian motion.

Also, if $\alpha \in (1, 2)$, then $n^{-1/\alpha} \sum_{t=1}^{n} X_t$ converges to a stable limit. As in the case of LMSV processes (see Sect. 4.3.4), this can be concluded from a general theory by Surgailis (2008).

4.3.7 Summary of Limit Theorems for Partial Sums

We summarize the main limit theorems. We consider centred linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ such that, as $x \to \infty$,

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}$$

Table 4.2 Limits for partial sums with infinite moments

	Partial sums—infinite moments	
	$S_n(u) = \sum_{t=1}^{[nu]} X_t$	$T_n(u) = \sum_{t=1}^{[nu]} (X_t^2 - E(X_1^2))$
Linear processes	$n^{-1/\alpha} S_n(1) \xrightarrow{d} c\tilde{Z}_\alpha(1)$ if $\sum \lvert a_j \rvert < \infty$ $n^{-(d+1/\alpha)} S_n(u) \Rightarrow c\tilde{Z}_{H,\alpha}(u)$ if $0 < d < 1 - 1/\alpha$ (Theorem 4.15)	$n^{-2/\alpha} T_n(1) \xrightarrow{d} c\tilde{Z}_{\alpha/2}(1)$ if $d \in (0, 1/\alpha)$ $n^{-2d} T_n(u) \Rightarrow cZ_{2,H}(u)$ if $d \in (1/\alpha, 1/2)$ (Theorem 4.16)
Stochastic volatility	$n^{-1/\alpha} S_n(u) \Rightarrow c\tilde{Z}_\alpha(u)$ (Theorem 4.19)	$n^{-2/\alpha} T_n(u) \Rightarrow c\tilde{Z}_{\alpha/2}(u)$ if $d \in (0, 2/\alpha - 1/2)$ $n^{-(1/2+d)} T_n(u) \Rightarrow cB_H(u)$ if $d \in (2/\alpha - 1/2, 1/2)$ (Theorem 4.19)

with $\alpha \in (1, 2)$ and appropriate regularity conditions (that assure the existence of the process) hold. When the sum of the squares X_t^2 is considered, then we assume instead that α is in the range $\alpha \in (2, 4)$.

Another class of processes considered above are stochastic volatility models with infinite second moments. As a representative, we look at $X_t = \xi_t \exp(\sum_{j=1}^{\infty} a_j \varepsilon_{t-j})$, where the sequences ξ_t and ε_t are mutually independent. We assume that

$$P(\xi_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\xi_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}$$

with $\alpha \in (1, 2)$ and $E[\xi_1] = 0$. Again, if the sum of X_t^2 is considered, then this tail behaviour is assumed to hold for $\alpha \in (2, 4)$. Furthermore, the random variables ε_t are assumed to be standard normal. We use the notation $B(\cdot)$ for a Brownian motion on $[0, 1]$, $B_H(\cdot)$ denotes a fractional Brownian motion on $[0, 1]$, $Z_{2,H}(\cdot)$ is the Hermite–Rosenblatt process on $[0, 1]$, and $\tilde{Z}_{H,\alpha}$ is a linear fractional stable motion with Hurst parameter $H = d + 1/\alpha$. Furthermore, c is a generic constant. We summarize the results for partial sums in Table 4.2. For simplicity, the slowly varying functions are assumed to be constant.

4.4 Limit Theorems for Sample Covariances

In a preliminary analysis of a time series, sample autocovariances play a crucial role. Moreover, limit theorems for quadratic forms can often be deduced from those for sample covariances. In this section we therefore study the limiting behaviour of sample covariances and, more generally, of multivariate functions applied to long-memory sequences. Surprisingly, this theory is not well developed beyond Gaussian (Rosenblatt 1979; Ho and Sun 1987, 1990; Arcones 1994) and linear processes with finite (Hosking 1996; Horváth and Kokoszka 2008) and infinite moments (Kokoszka and Taqqu 1996; Horváth and Kokoszka 2008). Some recent results were developed for stochastic volatility models (Davis and Mikosch 2001; McElroy and Politis 2007; Kulik and Soulier 2012).

4.4.1 Gaussian Sequences

In what follows, all vectors are considered as column vectors. Consider a stationary centred sequence of Gaussian vectors

$$\mathbf{X}_t = \left(X_t^{(1)}, \ldots, X_t^{(q)}\right)^T \quad (t \in \mathbb{Z})$$

with the marginal covariance matrix Σ and autocovariance function $\gamma_{i,j}(k) = E[X_0^{(i)} X_k^{(j)}]$ $(i, j = 1, \ldots, q)$, and assume either

$$\sum_{k=-\infty}^{\infty} \left|\gamma_{i,j}(k)\right| < \infty \tag{4.111}$$

or the existence of a parameter $d \in (0, 1/2)$ and a slowly varying function L_γ such that

$$\gamma_{i,j}(k) \sim a_{i,j} k^{2d-1} L_\gamma(k) \quad (i, j = 1, 2, \ldots, q), \tag{4.112}$$

where the constants $a_{i,j}$ are not all equal to zero. We will then use the same notation $\gamma(k) = k^{2d-1} L_\gamma(k)$ as in the univariate case.

Example 4.18 Let $q = 2$ and assume that $\tilde{X}_t^{(1)}$ $(t \in \mathbb{N})$ and $\tilde{X}_t^{(2)}$ $(t \in \mathbb{N})$ are mutually independent long-memory standard Gaussian sequences with the same covariances $\gamma_X(k) = \gamma_{\tilde{X}}(k) = \gamma(k)$. Then (4.112) holds with $a_{1,1} = a_{2,2} = 1$ and $a_{1,2} = a_{2,1} = 0$.

Example 4.19 Let X_t $(t \in \mathbb{N})$ be a stationary standard Gaussian sequence with covariance $\gamma_X(k) = c_\gamma k^{2d-1}$. Fix $s > 0$, and let

$$\left(X_t^{(1)}, X_t^{(2)}\right)^T = (X_t, X_{t+s})^T \quad (t \in \mathbb{N}).$$

Then

$$\gamma_{1,1}(k) = \gamma_{2,2}(k) = E[X_0 X_k] = \gamma_X(k),$$

so that $a_{1,1} = a_{2,2} = 1$. Furthermore,

$$\gamma_{1,2}(k) = E[X_0 X_{s+k}] = \gamma_X(k+s) \sim \gamma_X(k)$$

as $k \to \infty$, so that $a_{1,2} = 1$. Similarly, $a_{2,1} = 1$.

Example 4.20 Assume that $\tilde{X}_t^{(1)}$ and $\tilde{X}_t^{(2)}$ ($t \in \mathbb{N}$) are as in Example 4.18. Fix $s > 0$, and let

$$\left(X_t^{(1)}, X_t^{(2)}\right)^T = \left(\tilde{X}_t^{(1)}, \rho \tilde{X}_t^{(2)} + \sqrt{1-\rho^2}\tilde{X}_t^{(2)}\right)^T,$$

where $\rho = \gamma_X(s)$. Note that for a fixed t, the vectors $(X_t^{(1)}, X_t^{(2)})^T$ in Example 4.19 and here have the same covariance matrix. Now, $a_{1,1} = a_{2,2} = 1$, whereas

$$\gamma_{1,2}(k) = \rho\gamma_X(k),$$

so that $a_{1,2} = \rho$. Similarly, $a_{2,1} = \rho$.

After explaining basic structures of dependent Gaussian vectors, we turn our attention to limit theorems. It turns out that limit theorems for multivariate Gaussian vectors can be reduced to the case where the vectors have the identity covariance matrix I_q. Therefore, we start with the case of independent components.

4.4.1.1 Independent Components

Consider the collection $\{\tilde{X}_t^{(l)}, l \in \mathbb{N}, t \in \mathbb{N}\}$ of long-memory Gaussian sequences. For any $l \neq k$, the sequences $X_t^{(l)}$ and $X_t^{(k)}$ $(t \in)$ are assumed to be independent. Recall the following notation from Sect. 4.2.3 (see also Sect. 4.1.3) the following notation. Assume for a moment that $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ is the Gaussian process, where ε_t $(t \in \mathbb{Z})$ are i.i.d. standard normal random variables. Consider the following random measures: $M_\varepsilon(\cdot)$ is a Gaussian random measure with independent increments, associated with the sequence ε_t, that is $E[|dM_\varepsilon(\lambda)|^2] = \sigma_\varepsilon^2/(2\pi)\,d\lambda$, $dM_0(\lambda) = \sqrt{2\pi}\,dM_\varepsilon(\lambda)$,

$$dM_X(\lambda) = \left(\sum_{j=0}^{\infty} a_j e^{-ij\lambda}\right) dM_\varepsilon(\lambda) = A\left(e^{-i\lambda}\right) dM_\varepsilon(\lambda) = a(\lambda)\,dM_0(\lambda)$$

is the spectral random measure associated with a sequence X_t $(t \in \mathbb{N})$. Recall further that $n^{1/2}M_0(n^{-1}A)$ is another Gaussian random measure with the same distribution as $M_0(A)$. Then

$$\frac{L_f^{1/2}((n\lambda)^{-1})}{L_f^{1/2}(n^{-1})}|\lambda|^{-d} n^{1/2}\,dM_0\left(n^{-1}\lambda\right)$$

converges vaguely to $W_X(d\lambda) := |\lambda|^{-d}\,dM_0(\lambda)$.

As in Sect. 4.2.3, we can represent the Gaussian sequences $\tilde{X}_t^{(l)}$ $(t \in \mathbb{N})$ as (cf. (4.28))

$$\tilde{X}_t^{(l)} = \int_{-\pi}^{\pi} e^{it\lambda}\,dM_{\tilde{X}^{(l)}}(\lambda) \quad (t \geq 1),$$

where

$$dM_{\tilde{X}^{(l)}}(\lambda) = a^{(l)}(\lambda)\, dM_0^{(l)}(\lambda),$$

and $M_0^{(l)}(\cdot)$ $(l \geq 1)$ are independent Gaussian random measures. Furthermore, $|a^{(l)}(\lambda)|^2 = f_{\tilde{X}^{(l)}}(\lambda)$, where $f_{(l)} = f_{\tilde{X}^{(l)}}$ is the spectral density associated with the sequence $\tilde{X}_t^{(l)}$ $(t \in \mathbb{N})$. Also, $n^{1/2} M_0^{(l)}(n^{-1}A) \overset{d}{=} M_0(A)$, and

$$\frac{L_{f_{(l)}}^{1/2}((n\lambda)^{-1})}{L_{f_{(l)}}(n^{-1})} |\lambda|^{-d} n^{1/2}\, dM_0^{(l)}\big(n^{-1}\lambda\big) \tag{4.113}$$

converges vaguely to a measure $dW_{\tilde{X}^{(l)}}(\lambda) = |\lambda|^{-d}\, dM_0^{(l)}(\lambda)$.

As in the alternate proof of Theorem 4.2 (see also the proof of Theorem 4.3), we may write

$$\begin{aligned}\sum_{t=0}^{n-1} \tilde{X}_t^{(1)} \tilde{X}_t^{(2)} &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{e^{in(\lambda_1+\lambda_2)}-1}{e^{i(\lambda_1+\lambda_2)}-1} a^{(1)}(\lambda_1) a^{(2)}(\lambda_2)\, dM_0^{(1)}(\lambda_1)\, dM_0^{(2)}(\lambda_2)\\ &= \int_{-n\pi}^{n\pi}\int_{-n\pi}^{n\pi} D_n\big((\lambda_1+\lambda_2)/n\big)\\ &\quad \times \prod_{l=1}^{2} a^{(l)}\left(\frac{\lambda_l}{n}\right) n^{1/2}\, dM_0^{(1)}\big(n^{-1}\lambda_1\big) n^{1/2}\, dM_0^{(2)}\big(n^{-1}\lambda_2\big)\end{aligned}$$

with

$$D_n(\lambda) = \frac{e^{i\lambda n}-1}{n(e^{i\lambda}-1)} 1\big\{|\lambda| < \pi n\big\}.$$

The functions above converge to

$$D(\lambda) = \frac{e^{i\lambda}-1}{i\lambda}.$$

Thus, if

$$a^{(l)}(\lambda) = a_{l,l} L_f^{1/2}\big(\lambda^{-1}\big)|\lambda|^{-d} \quad (l = 1, 2),$$

then we may conclude that for $d \in (1/4, 1/2)$,

$$\begin{aligned}& n^{-2d} L_f^{-1}\big(n^{-1}\big) \sum_{t=0}^{n-1} \tilde{X}_t^{(1)} \tilde{X}_t^{(2)}\\ &\quad \overset{d}{\to} a_{1,1} a_{2,2} \int_{\mathbb{R}^2} D(\lambda_1+\lambda_2) \prod_{l=1}^{2} \frac{1}{|\lambda_l|^d}\, dM_0^{(1)}(\lambda_1)\, dM_0^{(2)}(\lambda_2).\end{aligned}$$

This convergence can be extended to nonlinear functionals. The following theorem is adapted from Arcones (1994). For simplicity, we assume that all $a_{l,l}$ in (4.112) are one. (Recall from Example 4.18 that the terms $a_{i,l}$, $i \neq l$, vanish.)

Theorem 4.22 *Let $\tilde{X}_t = (\tilde{X}_t^{(1)}, \ldots, \tilde{X}_t^{(q)})^T$ $(t \in \mathbb{N})$, be a stationary sequence of centred Gaussian vectors with the marginal covariance matrix I_q, such that (4.112) holds. Let $G : \mathbb{R}^q \to \mathbb{R}$ be a function with the Hermite rank $m = \tilde{m}(G)$. If $m(1 - 2d) > 1$, then*

$$n^{-(1-m(1/2-d))} L_f^{m/2}\left(n^{-1}\right) \sum_{t=1}^{n} \left\{G(\tilde{X}_t) - E\left[G(\tilde{X}_1)\right]\right\}$$

$$\overset{\mathrm{d}}{\to} \sum_{r_1,\ldots,r_m=1}^{q} \tilde{c}_{r_1,\ldots,r_m} \tilde{Z}_{(r_1,\ldots,r_m),H}(1),$$

where

$$\tilde{Z}_{(r_1,\ldots,r_m),H}(1) = \int_{\mathbb{R}^m} D(\lambda_1 + \cdots + \lambda_m) \prod_{l=1}^{m} \frac{1}{|\lambda_l|^{r_l}} \, dM_0^{(r_1)}(\lambda_1) \cdots dM_0^{(r_m)}(\lambda_m),$$

$\int_{\mathbb{R}^m}$ is the m-fold multiple Wiener–Ito integral, and

$$\tilde{c}_{r_1,\ldots,r_m} = \frac{1}{m!} E\left[G(\tilde{X}_1) \prod_{l=1}^{q} H_{k(r_1,\ldots,r_m)}\left(\tilde{X}_1^{(l)}\right) \right],$$

where $k(r_1, \ldots, r_m)$ is the number of components among $r_1, \ldots, r_m$ that are equal to l.

Again, as in (4.33), the limiting random variable $\tilde{Z}_{(r_1,\ldots,r_m),H}(1)$ can be expressed as

$$\int_{\mathbb{R}^m} \frac{e^{iu(\lambda_1+\cdots+\lambda_m)} - 1}{i(\lambda_1 + \cdots + \lambda_m)} \, dW_{\tilde{X}^{(r_1)}}(\lambda_1) \cdots dW_{\tilde{X}^{(r_m)}}(\lambda_m), \tag{4.114}$$

where $dW_{\tilde{X}^{(r)}}(\lambda) = |\lambda|^{-d} \, dM_0^{(r)}(\lambda)$.

Example 4.21 Consider $G(y_1, y_2) = H_2(y_2)H_2(y_2)$. Then (see Example 3.8) its Hermite rank with respect to a vector $\tilde{X}_1 = (\tilde{X}_1^{(1)}, \tilde{X}_1^{(2)})^T$ of independent standard normal random variables is $m(G) = 4$. Then

$$c_{1,1,2,2} = \frac{1}{4!} E\left[G(\tilde{X}_1) H_2\left(\tilde{X}_1^{(1)}\right) H_2\left(\tilde{X}_1^{(2)}\right)\right] = \frac{1}{4!} \tilde{J}\left(G, (2, 2)\right) = \frac{4}{4!}.$$

Also, this computation is invariant under permutation of indices $(1,1,2,2)$. All other coefficients c_{r_1,r_2,r_3,r_4} vanish. Note that $k(1,1,2,2)=2$ for $l=1,2$. Thus,

$$n^{-(1-4(1/2-d))}L_f^{4/2}(n^{-1})\sum_{t=1}^{n}H_2(\tilde{X}_t^{(1)})H_2(\tilde{X}_t^{(2)})$$

converges in distribution to

$$\frac{6\times 4}{4!}\int_{\mathbb{R}^4}\frac{e^{iu(\lambda_1+\cdots+\lambda_4)}-1}{i(\lambda_1+\cdots+\lambda_4)}\,dW_{\tilde{X}^{(1)}}(\lambda_1)\,dW_{\tilde{X}^{(1)}}(\lambda_2)\,dW_{\tilde{X}^{(2)}}(\lambda_3)\,dW_{\tilde{X}^{(2)}}(\lambda_4).$$

This can be also seen by expanding

$$\sum_{t=1}^{n}H_2(\tilde{X}_t^{(1)})H_2(\tilde{X}_t^{(2)})$$

and using a representation for $H_m(X_t)$, see the proof of Theorem 4.3. The convergence is valid for $d\in(1/4,1/2)$.

Example 4.22 Let $G(y)=H_m(y)$. Then one can see that $Z_{m,H}(1)$ in Theorem 4.22 is exactly the Hermite–Rosenblatt random variable.

4.4.1.2 From Independent to Dependent Components

In general, let $X_t=(X_t^{(1)},\dots,X_t^{(q)})^T$ $(t\in\mathbb{N})$ be a long-memory Gaussian sequence with cross-autocovariance function $\gamma_{i,j}(k)=E(X_0^{(i)}X_k^{(j)})$ as in (4.112) and marginal covariance matrix Σ. Then the statement of Theorem 4.22 remains valid if we replace $m=\tilde{m}(G)$ by $m=m(G,X_1)$, where $m(G,X_1)$ is the Hermite rank of G with respect to the Gaussian vector X_1; the spectral measures $W_{\tilde{X}^{(r_l)}}$ are replaced by the so-called joint spectral measure

$$\left(dW_{X^{(1)}}(\lambda_1),\dots,dW_{X^{(q)}}(\lambda_q)\right),$$

and

$$c_{r_1,\dots,r_m}=\frac{1}{m!}E\left[G(X_1)\prod_{l=1}^{q}H_{k(r_1,\dots,r_m)}(X_1^{(l)})\right].$$

We do not provide details here; the reader is referred to Arcones (1994). However, we will consider the special case of the covariance matrix Σ since this leads to study of sample covariances.

Example 4.23 Recall Example 3.13. We consider the function

$$G(X_t,X_{t+s})=e^{pX_t}e^{pX_{t+s}}.$$

Then the Hermite rank is one. Thus, we have to evaluate $c_{r_1}, r_1 = 1, 2$. We compute

$$c_1 = E\big[G(X_t, X_{t+s})X_t\big] = p\big(1 + \gamma_X(s)\big)e^{p^2(1+\gamma_X(s))}.$$

Also, $c_2 = E[G(X_t, X_{t+s})X_{t+s}] = c_1$. Thus,

$$n^{-(d+1/2)}L_f^{-1/2}\big(n^{-1}\big)\sum_{t=1}^{n} e^{pX_t}e^{pX_{t+s}} \xrightarrow{d} 2c_1 \int D(\lambda)\,dW_X(\lambda),$$

where W_X is the spectral random measure associated with X_t $(t \in \mathbb{N})$, see (4.34).

4.4.1.3 From Independent to Dependent Components: Sample Covariances

We go back to the original problem of sample covariances. Our vectors $X_t = (X_1^{(1)}, X_t^{(2)})^T$ are as in Example 4.19:

$$\big(X_t^{(1)}, X_t^{(2)}\big)^T = (X_t, X_{t+s})^T \quad (t \in \mathbb{N}).$$

We write

$$X_t = \int_{-\pi}^{\pi} e^{ij\lambda}a(\lambda)\,dM_0(\lambda) = \int_{-\pi}^{\pi} e^{ij\lambda}\,dM_X(\lambda),$$

$$X_{t+s} = \int_{-\pi}^{\pi} e^{it\lambda}e^{is\lambda}a(\lambda)\,dM_0(\lambda) = \int_{-\pi}^{\pi} e^{it\lambda}e^{is\lambda}\,dM_X(\lambda).$$

Recall now the proof of Theorems 4.2 and 4.3. Like in the proof of Theorem 4.3

$$\begin{aligned}
&\sum_{t=0}^{n-1}\big(X_tX_{t+s} - E(X_tX_{t+s})\big)\\
&\quad= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{e^{in(\lambda_1+\lambda_2)} - 1}{e^{i(\lambda_1+\lambda_2)} - 1}\prod_{r=1}^{2} a(\lambda_r)e^{is\lambda_2}\,dM_0(\lambda_1)\,dM_0(\lambda_2)\\
&\quad= \int_{-n\pi}^{n\pi}\int_{-n\pi}^{n\pi} D_n\big((\lambda_1 + \lambda_2)/n\big)e^{is\lambda_2/n}\\
&\qquad\times \prod_{r=1}^{2} a\left(\frac{\lambda_r}{n}\right)n^{1/2}\,dM_0\big(n^{-1}\lambda_1\big)n^{1/2}\,dM_0\big(n^{-1}\lambda_2\big).
\end{aligned} \tag{4.115}$$

Note that, as $n \to \infty$, $e^{is\lambda_2/n} \to 1$. Therefore, omitting technical details, the limiting behaviour of

$$n^{-2d}L_f^{-1}\big(n^{-1}\big)\sum_{t=0}^{n-1}\big(X_tX_{t+s} - E(X_tX_{t+s})\big)$$

or, equivalently, of

$$n^{-2d}L_2^{-1/2}\big(n^{-1}\big)\sum_{t=0}^{n-1}\big(X_tX_{t+s}-E(X_tX_{t+s})\big)$$

is the same as that of $n^{-2d}L_2^{-1/2}(n^{-1})\sum_{t=0}^{n-1}(X_t^2-E(X_1^2))$, i.e. it does not involve s. Hence, using Theorem 4.3 with $m=2$, one can argue that for $d\in(1/4,1/2)$,

$$\begin{aligned}&n^{1-2d}L_2^{-1/2}\big(n^{-1}\big)\big(\hat{\gamma}_n(1)-\gamma_X(1),\dots,\hat{\gamma}_n(K)-\gamma_X(K)\big)\\&\quad\overset{\mathrm{d}}{\to}\big(Z_{2,H}(1),\dots,Z_{2,H}(1)\big),\end{aligned}\tag{4.116}$$

where

$$\hat{\gamma}_n(s)=\frac{1}{n}\sum_{t=0}^{n-s}X_tX_{t+s}\quad(s=1,\dots,K)$$

is the sample covariance at lag s and $H=d+1/2$. Thus, the limiting random vector has totally dependent components.

We extend this to arbitrary Hermite polynomials. Recall Example 3.15. One can derive the equation (see Lemma 3.4 in Fox and Taqqu 1985)

$$H_m(X_t)H_m(X_{t+s})=m!\gamma_X^m(s)+\sum_{r=1}^{m}(m-r)!\binom{m}{r}^2\gamma_X^{m-r}(s)K_r(t,t+s),\tag{4.117}$$

where

$$K_r(j,l)=\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}e^{ij(\lambda_1+\cdots+\lambda_r)+il(\lambda_{r+1}+\cdots+\lambda_{2r})}\prod_{l=1}^{2r}a(\lambda_l)\,dM_0(\lambda_1)\cdots dM_0(\lambda_{2r}).$$

For $m=1$, the formula reduces to the formula for X_tX_{t+s}, used in deriving (4.115). For $m=2$, the formula yields

$$\begin{aligned}&2\gamma_X^2(s)+4\gamma_X(s)\int\int e^{ij\lambda_1+is\lambda_2}\prod_{r=1}^{2}a(\lambda_r)\,dM_0(\lambda_1)\,dM_0(\lambda_2)\\&\quad+\int\cdots\int e^{ij(\lambda_1+\lambda_2)+i(j+s)(\lambda_3+\lambda_4)}\prod_{r=1}^{4}a(\lambda_r)\,dM_0(\lambda_1)\cdots dM_0(\lambda_4).\end{aligned}$$

The important feature of decomposition (4.117) is that under the condition $d\in(1/4,1/2)$ only the term with $r=1$ will contribute. In other words, the limiting behaviour of

$$\hat{\gamma}_n(s;H_m):=\frac{1}{n}\sum_{t=1}^{n-s}H_m(X_t)H_m(X_{t+s})$$

is up to a constant the same for each $m \geq 1$. Noting that $(m-1)!\binom{m}{1}^2 = m!m$ and using (4.117), we have for $d \in (1/4, 1/2)$,

$$n^{1-2d} L_2^{-1}(n^{-1})\big(\hat{\gamma}_n(1; H_m) - m!\gamma_X^m(1), \dots, \hat{\gamma}_n(K; H_m) - m!\gamma_X^m(K)\big)$$
$$\xrightarrow{d} m!m\big(\gamma_X^{m-1}(1), \dots, \gamma_X^{m-1}(K)\big) Z_{2,H}(1), \quad (4.118)$$

where $H = d + 1/2$.

4.4.2 Linear Processes with Finite Moments

In this section we consider second-order stationary linear processes $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ $(t \in \mathbb{N})$, where ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that $E(\varepsilon_1) = 0$, $E(\varepsilon_1^2) = \sigma_\varepsilon^2 = 1$ and $E(\varepsilon_1^4) = \eta < \infty$.

Let

$$\hat{\gamma}_n(s) = \frac{1}{n} \sum_{t=0}^{n-s} X_t X_{t+s}.$$

It converges in probability to the population covariance

$$\gamma_X(s) = E(X_0 X_s) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j a_{j+s}.$$

Classical results for weakly dependent sequences under $E(\varepsilon_1^4) < \infty$ were obtained in Anderson (1971, p. 478); see also Brockwell and Davis (1991, Proposition 7.3.3). For long-memory linear processes, they were obtained in Hosking (1996) and Horváth and Kokoszka (2008).

Theorem 4.23 *Let* $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ $(t \in \mathbb{N})$ *be a linear process such that* $E(\varepsilon_1) = 0$, $E(\varepsilon_1^2) = \sigma_\varepsilon^2 = 1$ *and* $E(\varepsilon_1^4) = \eta < \infty$. *Furthermore, assume that* $\sum_{j=0}^{\infty} a_j^2 = 1$.

(a) *If* $a_j \sim L_a(j) j^{d-1}$, $d \in (0, 1/4)$ *or* $\sum_{j=0}^{\infty} |a_j| < \infty$, *then*

$$n^{1/2}\big(\hat{\gamma}_n(s) - \gamma_X(s)\big) \xrightarrow{d} N\big(0, \nu^2\big),$$

where the variance is

$$\nu^2 = (\eta - 3)\gamma_X^2(s) + \sum_{k=-\infty}^{\infty} \big(\gamma_X^2(k) + \gamma_X^2(k+s)\big).$$

(b) *If* $a_j \sim L_a(j) j^{d-1}$ *and* $d \in (1/4, 1/2)$, *then*

$$n^{1-2d} L_2^{-1/2}(n)\big(\hat{\gamma}_n(s) - \gamma_X(s)\big) \xrightarrow{d} Z_{2,H}(1),$$

where $Z_{2,H}(u)$ *is a Hermite–Rosenblatt process*, $L_2(n) = 2C_2 L_\gamma^2(n)$,

$$C_2 = \big[\big(2(2d-1)+1\big)(2d+1)\big]^{-1},$$

and $L_\gamma(n)$ *is given in* (4.39).

This theorem can be formulated in a multivariate setup. In the first case the limiting distribution is multivariate normal (with dependent components):

$$n^{1/2}\big(\hat{\gamma}_n(0) - \gamma_X(0), \ldots, \hat{\gamma}_n(q) - \gamma_X(q)\big) \xrightarrow{d} (G_0, \ldots, G_q), \tag{4.119}$$

where $(G_0, \ldots, G_q)$ is a zero-mean Gaussian vector with covariance

$$E[G_s G_t] = (\eta - 3)\gamma_X(s)\gamma_X(t) + \sum_{k=-\infty}^{\infty} \big(\gamma_X(k)\gamma_X(k+s-t) + \gamma_X(k+s)\gamma_X(k+t)\big). \tag{4.120}$$

In the second case, $d \in (1/4, 1/2)$, the limit has the form $(Z_{2,H}(1), \ldots, Z_{2,H}(1))$.

Proof For part (a), we use the standard truncation argument as illustrated in the proof of Theorem 4.5. Let

$$X_{t,K} = \sum_{j=0}^{K} a_j \varepsilon_{t-j},$$

$$\hat{\gamma}_n^{(K)}(s) = \frac{1}{n}\sum_{t=0}^{n-s} X_{t,K} X_{t+s,K}, \qquad \gamma_X^{(K)}(s) = E[X_{0,K} X_{s,K}] = \sigma_\varepsilon^2 \sum_{j=0}^{K} a_j a_{j+s}.$$

First, since the sequence $X_{t,K} X_{t+s,K}$ is $(K+s)$-dependent, its convergence is described by

$$n^{1/2}\big(\hat{\gamma}_n^{(K)}(s) - \gamma_X^{(K)}(s)\big) \xrightarrow{d} N\big(0, \nu_K^2\big),$$

where

$$\nu_K^2 = (\eta - 3)\big(\gamma_X^{(K)}(s)\big)^2 + \sum_{k=-\infty}^{\infty} \big[\big(\gamma_X^{(K)}(k)\big)^2 + \big(\gamma_X^{(K)}(k+s)\big)^2\big].$$

Since $\nu_K^2 \to \nu^2$ as $K \to \infty$, we also have $N(0, \nu_K^2) \xrightarrow{d} N(0, \nu^2)$. It suffices to verify that for all $\delta > 0$,

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\Big(\big|n^{1/2}\big(\hat{\gamma}_n^{(K)}(s) - \gamma_X^{(K)}(s)\big) - n^{1/2}\big(\hat{\gamma}_n(s) - \gamma_X(s)\big)\big| > \delta\Big) = 0.$$

By Markov's inequality, to do this, it suffices to verify that

$$\lim_{K\to\infty}\lim_{n\to\infty} n\cdot \operatorname{var}\big(\hat{\gamma}_n^{(K)}(s)-\hat{\gamma}_n(s)\big)=0.$$

In the case of Theorem 4.5 this was handled by introducing the random variable $\bar{X}_{t,K}=X_t-X_{t,K}$. In our situation here this is not straightforward since

$$\sum_{j,j'=0}^{\infty} a_j a_{j+s}-\sum_{j,j'=0}^{K} a_j a_{j+s}\neq \sum_{j,j'=K+1}^{\infty} a_j a_{j+s}.$$

We have to verify that

$$\lim_{n\to\infty} n\cdot \operatorname{var}\big(\hat{\gamma}_n(s)\big)=\nu^2,$$

$$\lim_{K\to\infty}\lim_{n\to\infty} n\cdot \operatorname{var}\big[\hat{\gamma}_n^{(K)}(s)\big]=\nu^2,\qquad \lim_{K\to\infty}\lim_{n\to\infty} n\cdot cov\big(\hat{\gamma}_n^{(K)}(s),\hat{\gamma}_n(s)\big)=\nu^2.$$

We prove the first part only. The expression is

$$\sum_{k=-(n-1)}^{n-1}\left(1-\frac{|k|}{n}\right)\left[(\eta-3)\sigma_\varepsilon^2\sum_{j=0} a_j a_{j+s}a_{j+k}a_{j+k+s}+\gamma_X^2(k)+\gamma_X^2(k+s)\right].$$

Then the relation follows by the dominated convergence theorem. For this, one needs, in particular, $\sum_k \gamma_X^2(k)<\infty$, which is achieved if $d\in(0,1/4)$ or $\sum_{j=0}^{\infty}|a_j|<\infty$.

As for part (b), we use the following decomposition:

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\big(X_tX_{t+s}-E(X_tX_{t+s})\big)\\ &\quad=\frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{\infty}a_ja_{j+s}\big(\varepsilon_{t-j}^2-\sigma_\varepsilon^2\big)+\frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{\infty}\sum_{l=0;\ l\neq j+s}^{\infty} a_ja_l\varepsilon_{t-j}\varepsilon_{t-l}\\ &\quad=: M_n+R_n.\end{aligned}$$

We may write the first part as $M_n=n^{-1}\sum_{t=1}^{n}Y_t$, where Y_t $(t\in\mathbb{N})$ is the linear process $Y_t=\sum_{j=0}^{\infty}c_j(\varepsilon_{t-j}-\sigma_\varepsilon^2)$ with summable coefficients $c_j=a_ja_{j+s}$. Indeed, by the Cauchy–Schwarz inequality,

$$\sum|c_j|\le\Big(\sum a_j^2\Big)^{1/2}\Big(\sum a_{j+s}^2\Big)^{1/2}<\infty.$$

Thus, $n^{1/2}M_n$ converges to a normal distribution on account of Theorem 4.5.

As for the second part, we may recognize that it has almost the same form as the therm $U_{n,2}$ in (4.51), so that its limiting distribution is of Hermite–Rosenblatt type.

If $d \in (1/4, 1/2)$, then

$$n^{1-2d} L_2^{-1/2}(n) R_n \xrightarrow{d} Z_{2,H}(1).$$

Thus, the second part dominates if $d \in (1/4, 1/2)$.

Note that formally the limit in part (b) may depend on s. However, this is not the case; a precise computation is given in Horváth and Kokoszka (2008). □

4.4.3 Linear Processes with Infinite Moments

Here we consider the same linear processes as in Sect. 4.4.2, however, instead of assuming $E[\varepsilon_1^4] < \infty$, we impose the regularly varying condition:

$$P(\varepsilon_1 > x) \sim A \frac{1+\beta}{2} x^{-\alpha}, \qquad P(\varepsilon_1 < -x) = A \frac{1+\beta}{2} x^{-\alpha}, \tag{4.121}$$

where $A > 0$, $\beta \in [-1, 1]$ and $\alpha \in (1, 4)$. In particular, $E[|\varepsilon_1|] < \infty$, $E[\varepsilon_1^4] = +\infty$.

There is a vast literature on sample covariances for weakly dependent linear processes with regularly varying innovations. Kanter and Steiger (1974) considered AR(p) models, Davis and Resnick (1985, 1986) considered processes with infinite variance and with finite variance, but infinite fourth moment, respectively. In the latter papers, the authors used point process techniques, as described in the section on partial sums with infinite moments; see Sect. 4.3. This technique was successfully applied to bilinear processes with infinite moments (Davis and Resnick 1996; Basrak et al. 1999) and to GARCH models (Davis and Mikosch 1998; Basrak et al. 2002)

As for long-memory linear processes, Kokoszka and Taqqu (1996) generalized the results by Davis and Resnick (1985) for $\alpha \in (1, 2)$, whereas Horváth and Kokoszka (2008) generalized Davis and Resnick (1986) for $\alpha \in (2, 4)$. (Recall that there is no long memory if $\alpha \in (0, 1)$).

Recall that the sample covariance is defined as

$$\hat{\gamma}_n(s) = \frac{1}{n} \sum_{t=1}^{n-s} X_t X_{t+s} \quad (s = 1, \ldots, q).$$

The first result deals with $\alpha \in (1, 2)$. There is no influence of long memory.

Theorem 4.24 *Assume that X_t $(t \in \mathbb{N})$ is a linear process and ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that (4.121) holds with $\alpha \in (1, 2)$ and $E(\varepsilon_1) = 0$. If $\alpha \in$*

$(1, 2)$, *then*

$$n^{1-2/\alpha}\big(\hat{\gamma}_n(0), \ldots, \hat{\gamma}_n(q)\big) \xrightarrow{d} A^{2/\alpha} C_{\alpha/2}^{-2/\alpha} \left(\sum_{j=0}^{\infty} a_j a_{j+0}, \ldots, \sum_{j=0}^{\infty} a_j a_{j+q} \right) S_{\alpha/2}(1, 1, 0), \quad (4.122)$$

where $S_\alpha(1, 1, 0)$ *is a stable random variable.*

Proof The proof is given in Davis and Resnick in the weakly dependent case (4.88); however it applies to the long-memory situation as long as the conditions of Theorem 4.24 are fulfilled. The reason for this is that under the condition $\sum_j a_j^2 < \infty$, the quantity $\sum_j a_j a_{j+s}$ is also finite. We give a sketch of the proof for $\hat{\gamma}_n(q)$ only. Recall from Theorem 4.14 that

$$\sum_{t=1}^{n} \delta_{c_n^{-1}(X_t, \ldots, X_{t-K})} \Rightarrow \sum_{l=1}^{\infty} \sum_{r=0}^{\infty} \delta_{j_l(a_r, a_{r-1}, \ldots, a_{r-K})},$$

where j_l are points of the limiting Poisson process, c_n is such that $P(|\varepsilon_1| > c_n) \sim n^{-1}$, i.e. $c_n \sim A^{1/\alpha} n^{1/\alpha}$. The continuous mapping theorem yields

$$c_n^{-2} \sum_{t=1}^{n} X_t X_{t+q} 1\big\{|X_t| > c_n \gamma \text{ or } |X_{t+h}| > c_n \gamma\big\} \xrightarrow{d} \sum_{l=0}^{\infty} \sum_{t=0}^{\infty} a_j a_{j+q} j_l^2 1\big\{|j_l| > \min\big\{a_j^{-1}, a_{j+q}^{-1}\big\}\gamma\big\}.$$

As $\gamma \to 0$, the latter random variable converges to

$$\left(\sum_{j=0}^{\infty} a_j a_{j+q} \right) \sum_{l=0}^{\infty} j_l^2 \stackrel{d}{=} \left(\sum_{j=0}^{\infty} a_j a_{j+q} \right) S_{\alpha/2}\big(C_{\alpha/2}^{-2/\alpha}, 1, 1\big).$$

It remains to show that

$$\lim_{\gamma \to 0} \limsup_{n \to \infty} P\left(c_n^{-2} \left| \sum_{t=1}^{n} X_t X_{t+q} 1\big\{|X_t| < c_n \gamma, |X_{t+q}| < c_n \gamma\big\} \right| > \gamma \right) = 0.$$

This probability is bounded by

$$\frac{n}{c_n^2 \gamma} E\big[\big|X_1^2\big| 1\big\{|X_1| < \gamma c_n\big\}\big].$$

We conclude the proof by applying Karamata's theorem (Lemma 4.18) together with the tail estimates in Lemma 4.19. □

The situation is different for $\alpha \in (2, 4)$. We have a dichotomous behaviour, depending on the interplay between tails and memory.

Theorem 4.25 *Assume that X_t $(t \in \mathbb{N})$ is a linear process such that $a_j \sim c_a j^{d-1}$, $d \in (0, 1/2)$ (so that $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$, see (4.39)) and ε_t $(t \in \mathbb{Z})$ are i.i.d. random variables such that (4.121) holds with $\alpha \in (2, 4)$ and $E(\varepsilon_1) = 0$.*

- *If $\alpha \in (2, 4)$ and $0 < d < 1/\alpha$, then (4.122) holds.*
- *If $\alpha \in (2, 4)$ and $1/\alpha < d < 1/2$, then*

$$n^{1-2d} L_2^{-1/2}(n) \big(\hat{\gamma}_n(s) - \gamma_X(s)\big) \xrightarrow{d} Z_{2,H}(1),$$

where $Z_{2,H}(u)$ is a Hermite–Rosenblatt process, and $L_2(n) = 2! C_2 L_\gamma^2(n)$.

Proof Consider the decomposition $M_n + R_n$ from the proof of Theorem 4.23:

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\big(X_t X_{t+s} - E(X_t X_{t+s})\big) \\
&\quad = \frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{\infty} a_j a_{j+s}\big(\varepsilon_{t-j}^2 - \sigma_\varepsilon^2\big) + \frac{1}{n}\sum_{t=1}^{n}\sum_{j=0}^{\infty}\sum_{l=0;\ l\neq j+s}^{\infty} a_j a_l \varepsilon_{t-j}\varepsilon_{t-l} \\
&\quad =: M_n + R_n.
\end{aligned}$$

Since the random variables ε_t have a finite variance, we again have

$$n^{1-2d} L_2^{-1/2}(n) R_n \xrightarrow{d} Z_{2,H}(1)$$

if $d \in (1/4, 1/2)$ and $n^{-1/2} R_n = O_P(1)$ if $d \in (0, 1/4)$. The first part, M_n, is the partial sum of a linear process with summable coefficients and infinite variance, and hence we can conclude the stable limit for M_n. □

4.4.4 Stochastic Volatility Models

Some recent results were developed for stochastic volatility models (McElroy and Politis 2007, Kulik and Soulier 2012). In the latter paper, the authors show differences between LMSV and models with a leverage.

Consider a stochastic volatility model $X_t = \sigma_t \xi_t$ $(t \in \mathbb{N})$ such that the sequences σ_t $(t \in \mathbb{N})$ and ξ_t $(t \in \mathbb{N})$ are independent. Assume that $E(\xi_1) = 0$. We are interested in sample covariances of X_t and X_t^2. For the first one, we note that

$$\hat{\gamma}_n(s) = \frac{1}{n}\sum_{t=1}^{n-s} \xi_t \xi_{t+s} \sigma_t \sigma_{t+s}$$

is a martingale w.r.t. sigma field generated by (σ_j, ξ_j), $j \le t$. Therefore, if we assume additionally $E[\xi_1^2] < \infty$, then

$$\sqrt{n}\hat{\gamma}_n(s) \xrightarrow{d} N\big(0, v^2\big),$$

where $v^2 = E[\sigma_0^2\sigma_s^2]E^2[\xi_1^2]$. The more interesting situation happens in the second case of squares. Assume that $E[\xi_1^4] < \infty$. Then

$$\frac{1}{n}\sum_{t=1}^{n}\left(\xi_t^2\xi_{t+s}^2\sigma_t^2\sigma_{t+s}^2 - E\left[\xi_t^2\xi_{t+s}^2\right]E\left[\sigma_j^2\sigma_{t+s}^2\right]\right)$$
$$= \frac{1}{n}\sum_{t=1}^{n}\sigma_t^2\sigma_{t+s}^2\left(\xi_t^2\xi_{t+s}^2 - E\left[\xi_t^2\xi_{t+s}^2\right]\right) + E^2\left[\xi_1^2\right]\frac{1}{n}\sum_{t=1}^{n}\left(\sigma_t^2\sigma_{t+s}^2 - E\left[\sigma_t^2\sigma_{t+s}^2\right]\right)$$
$$=: M_n + R_n.$$

Again, the first part is a martingale, and therefore it is $O_P(n^{-1/2})$. The second part is a possible long-memory contribution of the bivariate sequence $\sigma_t\sigma_{t+s}$ $(t \in \mathbb{N})$. For example, if we consider $\sigma_t = \exp(p\zeta_t)$, where ζ_t $(t \in \mathbb{N})$ is the long-memory Gaussian process as in Example 4.23, then for $d \in (1/4, 1/2)$ (refer to Example 4.23 for the precise notation),

$$n^{-(d+1/2)}L_f^{-1/2}(n)R_n \xrightarrow{d} 2E^2\left[\xi_1^2\right]c_1\int D(\lambda)\,dW_\zeta(\lambda),$$

where W_ζ is the spectral random measure associated with ζ_t $(t \in \mathbb{N})$. Therefore, since the second part R_n dominates, the limiting distribution for

$$n^{1-(d+1/2)}L_2^{-1/2}\left(n^{-1}\right)\hat{\gamma}_n(s)$$

is the same as for R_n. If on the other hand $d \in (0, 1/4)$, then both terms M_n and R_n are of the same order.

This consideration can be extended to random variables ξ_t such that (4.121) holds with $\alpha \in (2, 4)$. Then, we have again a dichotomous behaviour: the limit can be either a stable random variable or a Hermite–Rosenblatt random variable. The situation becomes complicated though when one considers models with leverage. We refer to Davis and Mikosch (2001) and Kulik and Soulier (2012).

4.4.5 Summary of Limit Theorems for Sample Covariances

We consider a centred linear process $X_t = \sum_{j=0}^{\infty} a_j\varepsilon_{t-j}$ such that either $E(\varepsilon_1^4) < \infty$ or

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}$$

with $\alpha \in (1, 4)$ and appropriate regularity conditions (that assure existence of the process). In the table, $Z_{2,H}(\cdot)$ is a Hermite–Rosenblatt process on $[0, 1]$, and $\tilde{S}_{\alpha/2}$ is an $\alpha/2$-stable random variable. Furthermore, c is a generic constant. The main results are summarized in Table 4.3.

Table 4.3 Limits for sample covariances

	Sample covariances	
	Finite moments	Infinite moments
Linear processes	$n^{1/2}(\hat{\gamma}_n(s)-\gamma_X(s))\overset{d}{\to} cN(0,1)$ if $d\in(0,1/4)$ $n^{1-2d}(\hat{\gamma}_n(s)-\gamma_X(s))\overset{d}{\to} cZ_{2,H}(1)$ if $d\in(1/4,1/2)$ (Theorem 4.23)	$\alpha\in(1,2)$ $n^{1-2/\alpha}(\hat{\gamma}_n(s)-\gamma_X(s))\overset{d}{\to} c\tilde{S}_{\alpha/2}$ $\alpha\in(2,4)$ $n^{1-2/\alpha}(\hat{\gamma}_n(s)-\gamma_X(s))\overset{d}{\to} c\tilde{S}_{\alpha/2}$ if $d\in(0,1/\alpha)$ $n^{1-2d}(\hat{\gamma}_n(s)-\gamma_X(s))\overset{d}{\to} cZ_{2,H}(1)$ if $d\in(1/\alpha,1/2)$ (Theorems 4.25, 4.24)

4.5 Limit Theorems for Quadratic Forms

In this section we consider quadratic forms,

$$Q_n(u) := \sum_{t,s=1}^{[nu]} b_{t-s}\big\{G(X_t,X_s)-E\big[G(X_t,X_s)\big]\big\}, \qquad Q_n := Q_n(1), \quad (4.123)$$

where b_k ($k\in\mathbb{Z}$) is a sequence of constants, and $G:\mathbb{R}^2\to\mathbb{R}$. We are interested in asymptotic properties of $Q_n(u)$.

In the Gaussian case, such studies were conducted in Rosenblatt (1979), Fox and Taqqu (1985, 1987), Avram (1988), Terrin and Taqqu (1990), Beran and Terrin (1994), among others. For linear processes, classical limit theorems for weakly dependent sequences are given in Brillinger (1969) and Hannan (1970) (and references therein); also see Klüppelberg and Mikosch (1996). They follow directly from limit theorems for sample covariances, proven in Theorem 4.23. For long memory such studies were initiated by Giraitis and Surgailis (1990). The authors concluded a weakly dependent behaviour, using approximation of a quadratic form by another quadratic form with weakly dependent variables. Other results along these lines were proven for instance in Horváth and Shao (1999) and Bhansali et al. (1997). The case of the multivariate Appell polynomials is studied in Terrin and Taqqu (1991), Giraitis and Taqqu (1997, 1998, 1999a, 2001), Giraitis et al. (1998). Kokoszka and Taqqu (1997) discuss quadratic forms for infinite-variance processes. We also refer to Giraitis and Taqqu (1999b) for an overview.

There are two principal applications of quadratic forms. First, we can derive the limiting behaviour of the periodogram and the Whittle estimator (see Sect. 5.5 for results and references), or we can use quadratic forms to test for possible changes in the long-memory parameter (see e.g. Beran and Terrin 1996, Horváth and Shao 1999).

4.5.1 Gaussian Sequences

In this section we shall assume that X_t $(t \in \mathbb{Z})$ is a centred Gaussian sequence with autocovariance function $\gamma_X(k) = L_\gamma(k)k^{2d-1}$. First, we exploit the relation between sample covariances and quadratic forms. Using results obtained in Sect. 4.4, we obtain a *long-memory behaviour I* (i.e. of "type I") of $Q_n(u)$ for $d \in (1/4, 1/2)$ directly from limit theorems for sample covariances. The result was proven in Fox and Taqqu (1985) and is presented in Theorem 4.26. For $d \in (0, 1/4)$, we obtain convergence with rate $n^{-1/2}$, as proven in Fox and Taqqu (1985) as well. The result is presented in Theorem 4.27 and is referred to as *weakly dependent behaviour I*.

These results are very similar to those for partial sums $\sum_{t=1}^{[nu]}(X_t^2 - 1)$. These sums were studied in Sect. 4.2.3, and we recall the dichotomous behaviour: convergence to the Hermite–Rosenblatt process or Brownian motion for $d \in (1/4, 1/2)$ and $d \in (0, 1/4)$ respectively.

In Theorem 4.26 the limiting process will be degenerated if $\sum_l b_l = 0$, as it happens for Fourier coefficients. Another type of weakly dependent behaviour is obtained if in addition to $\sum_l b_l = 0$ the coefficients also decay to zero fast enough. Then, the coefficients b_l *compensate* for long memory, and $Q_n(\cdot)$ converges at rate $n^{1/2}$ for all $d \in (0, 1/2)$ (*weakly dependent behaviour II*). Such results were proven in Fox and Taqqu (1985, Theorem 3; 1987), Avram (1988), Beran and Terrin (1994) (also Beran 1986). The authors use the method of cumulants; see the proof of Theorem 4.28. On the other hand, if the coefficients b_l do not compensate for long memory, then Terrin and Taqqu (1990) prove that the limiting process is neither Gaussian nor Hermite–Rosenblatt (*long-memory behaviour II*). The authors use multiple Wiener–Itô integrals; see the proof of Theorem 4.29.

4.5.1.1 Long Memory Behaviour I

Recall that the sample covariances for the sequence X_t $(t \in \mathbb{Z})$ are defined by

$$\hat{\gamma}_n(s) = \frac{1}{n}\sum_{t=1}^{n-|s|} X_t X_{t+|s|}.$$

Reorganizing indices, we may write

$$Q_n(1) = \sum_{t,s=1}^{n} b_{t-s}\big(X_t X_s - E(X_t X_s)\big) = n \sum_{|l|\le n-1} b_l\big(\hat{\gamma}_n(l) - \gamma_X(l)\big).$$

Recall that for $d \in (1/4, 1/2)$ (see (4.116)),

$$n^{1-2d} L_2^{-1/2}(n)\big(\hat{\gamma}_n(1) - \gamma_X(1), \dots, \hat{\gamma}_n(K) - \gamma_X(K)\big) \xrightarrow{\mathrm{d}} \big(Z_{2,H}(1), \dots, Z_{2,H}(1)\big). \tag{4.124}$$

This, together with the continuous mapping theorem, implies that for any fixed integer $K > 0$,

$$n^{-2d} L_2^{-1/2}(n) Q_{n,K}(1) := n^{-2d} L_2^{-1/2}(n) n \sum_{|l| \le K} b_l \big(\hat{\gamma}_n(l) - \gamma_X(l)\big)$$
$$\xrightarrow{\mathrm{d}} \left(\sum_{l=-K}^{K} b_l\right) Z_{2,H}(1).$$

Clearly, $(\sum_{l=-K}^{K} b_l) Z_{2,H}(1) \xrightarrow{\mathrm{p}} (\sum_{l=-\infty}^{\infty} b_l) Z_{2,H}(1)$. Furthermore,

$$\lim_{K\to\infty} \limsup_{n\to\infty} P\big(n^{-2d} L_2^{-1/2}(n) \big| Q_{n,K}(1) - Q_n(1)\big| > \delta\big) = 0$$

for each $\delta > 0$. The reader is referred to Fox and Taqqu (1985, Theorem 1) for details on the latter approximation and tightness. This leads to the following result, which is formulated more generally in a functional form.

Theorem 4.26 *Assume that X_t $(t \in \mathbb{Z})$ is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$, $d \in (1/4, 1/2)$. If $\sum_{l=-\infty}^{\infty} |b_l| < \infty$, then*

$$n^{-2d} L_2^{-1/2}(n) Q_n(u) = n^{-2d} L_2^{-1/2}(n) \sum_{t,s=1}^{[nu]} b_{t-s}\big(X_t X_s - E(X_t X_s)\big)$$
$$\Rightarrow \left(\sum_{l=-\infty}^{\infty} b_l\right) Z_{2,H}(u),$$

where $L_2(n) = 2! C_2 L_\gamma^2(n)$ (cf. (4.22)), $H = d + \frac{1}{2}$, $\Rightarrow$ denotes weak convergence, and $Z_{2,H}(\cdot)$ is the Hermite–Rosenblatt process.

This result has been proven in fact in a more general setting Fox and Taqqu (1985). Consider

$$Q_n(u; H_m) := \sum_{t,s=1}^{[nu]} b_{t-s}\big\{H_m(X_t) H_m(X_s) - E\big[H_m(X_t) H_m(X_s)\big]\big\}.$$

The same methodology as above works, given that we use (4.118) instead of (4.124):

$$n^{1-2d} L_2^{-1/2}(n)\big(\hat{\gamma}_n(1; H_m) - m!\gamma_X^m(1), \ldots, \hat{\gamma}_n(K; H_m) - m!\gamma_X^m(K)\big)$$
$$\xrightarrow{\mathrm{d}} m! m\big(\gamma_X^{m-1}(1), \ldots, \gamma_X^{m-1}(K)\big) Z_{2,H}(1).$$

We conclude for $d \in (1/4, 1/2)$ and under the condition $\sum_{l=-\infty}^{\infty} |b_l| < \infty$,

$$n^{-2d} L_{A_2}^{-1/2}(n) Q_n(1; H_m) \xrightarrow{\mathrm{d}} m! m \left(\sum_{l=-\infty}^{\infty} b_l \gamma_X^{m-1}(l) \right) Z_{2,H}(1).$$

4.5.1.2 Weakly Dependent Behaviour I

Theorem 4.26 above requires $d \in (1/4, 1/2)$. What about $d \in (0, 1/4)$? As in the case of partial sums $\sum_{t=1}^{[nu]} (X_t^2 - 1)$, one obtains a weakly dependent behaviour, i.e. a central limit theorem with scaling $n^{-1/2}$ Fox and Taqqu (1985).

Theorem 4.27 *Assume that X_t ($t \in \mathbb{Z}$) is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$, $d \in (0, 1/4)$. Then*

$$n^{-1/2} Q_n(u) = n^{-1/2} \sum_{t,s=1}^{[nu]} b_{t-s} \big(X_t X_s - E(X_t X_s) \big) \Rightarrow \sigma_0 B(u),$$

where $B(\cdot)$ is a standard Brownian motion, and $\sigma_0 > 0$.

The constant σ_0 is given in a complicated form, and we refer to Fox and Taqqu (1985) for a precise formula.

4.5.1.3 Weakly Dependent Behaviour II

In Theorem 4.26 it may happen that $\sum_{l=-\infty}^{\infty} b_l = 0$ and hence the limit will be degenerated. This can happen when b_l are Fourier coefficients of a real-valued function g. Specifically, let

$$b_l = \int_{-\pi}^{\pi} e^{il\lambda} g(\lambda)\, d\lambda =: 2\pi \hat{g}_l, \qquad g(\lambda) \sim c_g |\lambda|^{-\gamma} \quad \text{as } |\lambda| \to 0. \tag{4.125}$$

To assure the existence of Fourier coefficients, we assume that $\gamma < 1$. Then, $b_l \sim c_b l^{\gamma-1}$, $c_b = 2 c_g \Gamma(1-\gamma) \sin(\pi \frac{\gamma}{2})$. The following result was proven in Fox and Taqqu (1987); see also Theorem 3 in Fox and Taqqu (1985) and Avram (1988).

Theorem 4.28 *Assume that X_t ($t \in \mathbb{Z}$) is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$, $d \in (0, 1/2)$. If*

$$2d + \gamma < 1/2, \tag{4.126}$$

then

$$n^{-1/2} Q_n(1) \xrightarrow{d} \sigma_Q Z, \tag{4.127}$$

where

$$\sigma_Q^2 := 16\pi^3 \int_{-\pi}^{\pi} \big(f(\lambda)g(\lambda)\big)^2\, d\lambda,$$

$f = f_X$ is the spectral density of X_t $(t \in \mathbb{Z})$, and Z is a standard normal random variable.

Let us comment on condition (4.126). First, it assures that σ_Q^2 is finite. Second, it means that the coefficients b_l decay appropriately fast, to compensate for long memory in X_t $(t \in \mathbb{Z})$.

Proof We present a modified version of the proof in Avram (1988). Let $\Sigma = [\gamma_X(j-l)]_{j,l=1}^{n}$ and $B = [b_{j-l}]_{j,l=0}^{n-1}$. Then,

$$Q_n(1) = (X_1, \ldots, X_n) B (X_1, \ldots, X_n)^T$$

has the pth cumulant equal to (see Grenander and Szegö 1958, p. 218)

$$\mathrm{cum}_p\big(Q_n(1)\big) = 2^{p-1}(p-1)!\mathrm{Trace}(\Sigma B)^p.$$

Note that

$$\gamma_X(j-l) = \int_{-\pi}^{\pi} e^{i(j-l)\lambda} f_X(\lambda)\, d\lambda =: 2\pi \hat{f}_{j-l},$$

where $\hat{f}_{j-l}$ is the Fourier coefficient of the spectral density $f = f_X$. Furthermore, $B = 2\pi[\hat{g}_{j-l}]_{j,l=0}^{n-1}$. Recall that the trace of a matrix is the sum of its diagonal elements. We have

$$\frac{1}{n}\mathrm{Trace}(\Sigma) = \frac{2\pi}{n}(\hat{f}_0 + \cdots + \hat{f}_0) = 2\pi \hat{f}_0 = \int_{-\pi}^{\pi} f_X(\lambda)\, d\lambda.$$

Of course, f_X is integrable given $d < 1/2$. Analogously, recall that the trace can be written as a Hadamard product: $\mathrm{Trace}(\Sigma B) = \sum_{j,l} \gamma_X(j-l) B_{j,l}$. Since $\hat{f}_l \hat{g}_l$ is summable, we then obtain

$$\frac{1}{n}\mathrm{Trace}(\Sigma B) = 4\pi^2 \frac{1}{n} \sum_{j,l=1}^{n} \hat{f}_{j-l}\hat{g}_{j-l} = 4\pi^2 \frac{1}{n} \sum_{l=-(n-1)}^{n-1} \big(n - |l|\big) \hat{f}_l \hat{g}_l$$

$$\approx 4\pi^2 \sum_{l=-(n-1)}^{n-1} \hat{f}_l \hat{g}_l \to 4\pi^2 \sum_{l=-\infty}^{\infty} \hat{f}_l \hat{g}_l$$

as $n \to \infty$. By the Parseval identity and since g is real,

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Trace}(\Sigma B) = 4\pi^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} f_X(\lambda)\bar{g}(\lambda)\, d\lambda = 2\pi \int_{-\pi}^{\pi} f_X(\lambda) g(\lambda)\, d\lambda.$$

On the other hand, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of ΣB, then we can write alternatively

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Trace}(\Sigma B) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \lambda_j \to \frac{4\pi^2}{2\pi}\int_{-\pi}^{\pi} f_X(\lambda)g(\lambda)\,d\lambda.$$

The matrix $(\Sigma B)^p$ has eigenvalues λ_j^p, $j = 1, \ldots, n$. One can then argue analogously that

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Trace}(\Sigma B)^p = \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \lambda_j^p = \frac{(4\pi^2)^p}{2\pi}\int_{-\pi}^{\pi} (f_X(\lambda)g(\lambda))^p\,d\lambda.$$

Thus,

$$\mathrm{cum}_p\big(n^{-1/2}Q_n(1)\big) = n^{-p/2}\mathrm{cum}_p\big(Q_n(1)\big) = \frac{2^{p-1}}{n^{p/2-1}}\frac{(p-1)!}{n}\mathrm{Trace}(\Sigma B)^p.$$

Consequently, $\lim_{n\to\infty}\mathrm{cum}_p(n^{-1/2}Q_n(1)) = 0$ if $p > 2$, and

$$\lim_{n\to\infty}\mathrm{cum}_2\big(n^{-1/2}Q_n(1)\big) = 16\pi^3\int_{-\pi}^{\pi}\big(f_X(\lambda)g(\lambda)\big)^2\,d\lambda,$$

which provides the limiting variance. Application of the method of cumulants (see Theorem 4.1) then yields the result. □

4.5.1.4 Long-Memory Behaviour II

In contrast to Theorem 4.28, if the coefficients b_l do not compensate for long memory (i.e., when (4.126) fails to hold), then we have the following result, due to Terrin and Taqqu (1990). Recall that $g(\lambda) \sim c_g|\lambda|^{-\gamma}$ as $\lambda \to 0$ (see (4.125)) and that $M_0(\cdot)$ is a random measure that appears in the spectral representation of the linear Gaussian sequence; see Sect. 4.1.3.

Theorem 4.29 *Assume that X_t $(t \in \mathbb{Z})$ is a stationary sequence of standard normal random variables such that $\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$, $d \in (0, 1/2)$. If*

$$1/2 < 2d + \gamma < 1, \tag{4.128}$$

then

$$n^{-(2d+\gamma)}L_f^{-1}\big(n^{-1}\big)Q_n(u) \Rightarrow c_g Z(u), \tag{4.129}$$

where

$$Z(u) = \iint \psi_u(\lambda_1,\lambda_2)\frac{1}{\lambda_1}\frac{1}{\lambda_2}\,dM_0(\lambda_1)\,dM_0(\lambda_2),$$

and

$$\psi_u(\lambda_1,\lambda_2)=\int_{\mathbb{R}}\frac{e^{iu(\lambda_1-\lambda)}-1}{i(\lambda_1+\lambda)}\frac{e^{iu(\lambda_2+\lambda)}-1}{i(\lambda_2-\lambda)}|\lambda|^{-\gamma}\,d\lambda.$$

The limiting process is self-similar with $H=2d+\gamma\in(\frac{1}{2},2)$, but neither Gaussian nor Hermite–Rosenblatt.

We note that for $\gamma=0$, we have $b_l=1$ for $l=0$ and 0 otherwise. In this case the result of Theorem 4.29 reduces to the asymptotic behaviour of $\sum_{t=1}^{[nu]}(X_t^2-1)$, see Theorem 4.3.

Proof The proof is sketched here. It follows the same idea as in the case of partial sums $\sum_{t=1}^{n}H_m(X_t)$. Recall that the multiple Wiener–Itô integral "removes" the diagonal (see Appendix A). We write

$$X_tX_s-E(X_tX_s)=\int_{[-\pi,\pi]^2\setminus\{\lambda_1=\lambda_2\}}e^{it\lambda_1}e^{is\lambda_2}a(\lambda_1)a(\lambda_2)\,dM_0(\lambda_1)\,dM_0(\lambda_2),$$

where $|a(\lambda)|^2=f_X(\lambda)$.

Thus,

$$\begin{aligned}
Q_n(1)&=\sum_{t,s=0}^{n-1}\int_{-\pi}^{\pi}e^{i(t-s)\lambda}g(\lambda)\,d\lambda\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{it\lambda_1}e^{is\lambda_2}a(\lambda_1)a(\lambda_2)\,dM_0(\lambda_1)\,dM_0(\lambda_2)\\
&=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}a(\lambda_1)a(\lambda_2)\\
&\quad\times\left(\int_{-\pi}^{\pi}\frac{e^{in(\lambda_1+\lambda)}-1}{e^{i(\lambda_1+\lambda)}}\frac{e^{in(\lambda_2-\lambda)}-1}{e^{i(\lambda_2-\lambda)}}g(\lambda)\,d\lambda\right)dM_0(\lambda_1)\,dM_0(\lambda_2)\\
&=c_gn^{\gamma}\int_{-n\pi}^{n\pi}\int_{-n\pi}^{n\pi}a\left(\frac{\lambda_1}{n}\right)a\left(\frac{\lambda_2}{n}\right)\\
&\quad\times\psi_1(\lambda_1,\lambda_2;n)n^{1/2}\,dM_0\big(n^{-1}\lambda_1\big)n^{1/2}\,dM_0\big(n^{-1}\lambda_2\big),
\end{aligned}$$

where

$$\psi_1(\lambda_1,\lambda_2;n)=\left(\int_{-n\pi}^{n\pi}D_n\left(\frac{\lambda_1+\lambda}{n}\right)D_n\left(\frac{\lambda_2-\lambda}{n}\right)g(\lambda)\,d\lambda\right),$$

$$D_n(\lambda)=\frac{e^{i\lambda n}-1}{n(e^{i\lambda}-1)}1\{|\lambda|\le\pi n\}.$$

Thus, $Q_n(1)$ equals in distribution to

$$\int_{-n\pi}^{n\pi}\int_{-n\pi}^{n\pi}a\left(\frac{\lambda_1}{n}\right)a\left(\frac{\lambda_2}{n}\right)\psi_1(\lambda_1,\lambda_2;n)\,dM_0(\lambda_1)\,dM_0(\lambda_2).$$

Clearly, $\lim_{n\to\infty}\psi_1(\lambda_1,\lambda_2;n)=\psi_1(\lambda_1,\lambda_2)$, and as in the alternative proof of Theorem 4.2, one can argue that the convergence is uniform. Therefore, the same method as in Theorem 4.2 applies, and the result (4.129) follows for $u=1$. A proof of functional convergence is omitted here. □

4.5.2 Linear Processes

As in the case of partial sums, the results on quadratic forms for Gaussian LRD sequences have a counterpart for general linear sequences

$$X_t=\sum_{j=0}^{\infty}a_j\varepsilon_{t-j}\quad (t\in\mathbb{Z}),\tag{4.130}$$

where $\sum_{j=0}^{\infty}a_j^2=1$, ε_t $(t\in\mathbb{Z})$ are i.i.d. zero mean random variables with $\mathrm{var}(\varepsilon_1)=\sigma_\varepsilon^2=1$. We will assume that either $\sum_{j=0}^{\infty}|a_j|<\infty$ or $a_j\sim L_a(j)j^{d-1}$ with $d\in(0,1/2)$.

Results for quadratic forms

$$Q_n(u)=\sum_{t,s=1}^{[nu]}b_{t-s}\bigl(X_tX_s-E(X_tX_s)\bigr)$$

based on weakly dependent linear processes are classical (see Brillinger 1969; Hannan 1970; also see Klüppelberg and Mikosch 1996) and follow directly from limit theorems for sample covariances, as proven before in Theorem 4.23.

For long memory, such studies had been initiated by Giraitis and Surgailis (1990). The authors concluded a weakly dependent behaviour, similar to that of Theorem 4.28, using an approximation of the quadratic form by another quadratic form with weakly dependent variables. Other results along this line can be found in Horváth and Shao (1999) and Bhansali et al. (1997).

When one replaces $Q_n(u)$ by

$$Q_n(u;P_{m_1,m_2})=\sum_{t,s=1}^{[nu]}b_{t-s}\bigl\{P_{m_1,m_2}(X_tX_s)-E\bigl[P_{m_1,m_2}(X_t,X_s)\bigr]\bigr\},$$

where P_{m_1,m_2} is a multivariate Appell polynomial, then limit theorems are very complicated; see Terrin and Taqqu (1991), Giraitis and Taqqu (1997, 1998, 1999a, 2001). We refer to Giraitis and Taqqu (1999b) for an overview.

4.5.2.1 Weakly Dependent Processes

Assume that $\sum_{j=0}^{\infty}|a_j| < \infty$. Recall Theorem 4.23 and the multivariate convergence (4.120):

$$n^{1/2}\big(\hat{\gamma}_n(0)-\gamma_X(0),\ldots,\hat{\gamma}_n(K)-\gamma_X(K)\big) \overset{\mathrm{d}}{\to} (G_0,\ldots,G_K),$$

where $(G_0,\ldots,G_K)$ is a Gaussian vector. We apply a similar method as in the proof of Theorem 4.26. There we concluded long-memory behaviour of quadratic forms from long-memory behaviour of sample covariances. Here, we will conclude short-memory behaviour of quadratic forms from short memory-behaviour of sample covariances.

We have

$$Q_n(1)=\sum_{t,s=1}^{n} b_{t-s}\big(X_tX_s - E(X_tX_s)\big) = n\sum_{|l|\le n-1} b_l\big(\hat{\gamma}_n(l)-\gamma_X(l)\big).$$

The continuous mapping theorem implies

$$n^{-1/2}Q_{n,K}(1) := n^{-1/2}n\sum_{|l|\le K} b_l\big(\hat{\gamma}_n(l)-\gamma_X(l)\big) \overset{\mathrm{d}}{\to} b_0G_0 + 2\sum_{l=1}^{K} b_lG_l.$$

To apply Proposition 4.1, we need to show that

$$\lim_{K\to\infty}\limsup_{n\to\infty} P\left(\sqrt{n}\left|\sum_{l=K+1}^{n-1} b_l\big(\hat{\gamma}_n(l)-\gamma_X(l)\big)\right| > \delta\right) = 0.$$

This is straightforward since the correlations between $\hat{\gamma}_n(l)$ $(l \ge 1)$ are absolutely summable. Therefore, we may apply Chebyshev inequality in a suitable way to finish the proof. □

4.5.2.2 Long-Memory Sequences

The following result is a counterpart to Theorem 4.28.

Theorem 4.30 *Assume that X_t $(t \in \mathbb{N})$ is a linear process with long-range dependence defined in (4.130), with spectral density $f_X(\lambda) \sim c_f|\lambda|^{-2d}$. Assume that the coefficients b_l are given by (4.125), i.e. $b_l \sim c_b l^{\gamma-1}$. Let κ_4 be the fourth cumulant of ε_1. If*

$$2d+\gamma < 1/2, \tag{4.131}$$

then

$$n^{-1/2}Q_n(1) = n^{-1/2}\sum_{t,s=1}^{n} b_{t-s}\big(X_tX_s - E(X_tX_s)\big) \overset{\mathrm{d}}{\to} \sigma_Q Z, \tag{4.132}$$

where Z is standard normal, and

$$\sigma_Q^2 := 16\pi^3 \int_{-\pi}^{\pi} \big(f_X(\lambda)g(\lambda)\big)^2 \, d\lambda + \kappa_4 \left(2\pi \int_{-\pi}^{\pi} f_X(\lambda)g(\lambda)\, d\lambda\right)^2.$$

Of course, if the innovations ε_t are normal, then $\kappa_4 = 0$, and the result reduces to Theorem 4.28.

Proof To prove this theorem, Giraitis and Surgailis (1990) do not use the method of cumulants. Instead, they approximate $Q_n = Q_n(1)$ by a weakly dependent sequence. A similar approach is also used in Bhansali et al. (1997), and we present a sketch of the method there.

Write $Q_{n,X} = \sum_{t,s=1}^{n} b_{t-s} X_t X_s$ and $Q_{n,\varepsilon} = \sum_{t,s=1}^{n} v_{t-s} \varepsilon_t \varepsilon_s$, where

$$v_l = 2\pi \int_{-\pi}^{\pi} g(\lambda) f_X(\lambda) e^{il\lambda}\, d\lambda.$$

Since $Q_{n,\varepsilon}$ is a quadratic form of independent random variables, it is much easier to derive its asymptotic distribution, namely (see Bhansali et al. 1997, Theorem 4.1):

$$\frac{1}{\sqrt{\operatorname{var}(Q_{n,\varepsilon})}} \big(Q_{n,\varepsilon} - E(Q_{n,\varepsilon})\big) \xrightarrow{\mathrm{d}} N(0,1),$$

where

$$\operatorname{var}(Q_{n,\varepsilon}) = v_0^2 n \cdot \sigma_\varepsilon^2 + 2 \sum_{j,l=1;\ j \neq l}^{n} v_{j-l}^2$$

and $\sigma_\varepsilon^2 = \operatorname{var}(\varepsilon_t)$. Under our assumptions,

$$g(\lambda) f_X(\lambda) \sim c_g |\lambda|^{-\gamma} c_f |\lambda|^{-2d}$$

as $\lambda \to 0$. Therefore, the coefficients v_l satisfy

$$v_l \sim c_v l^{2d+\gamma-1}, \quad c_v = 2 c_f c_g \Gamma\big(1 - (2d+\gamma)\big) \sin\left(\pi \frac{2d+\gamma}{2}\right).$$

Furthermore, $Q_{n,X} - Q_{n,\varepsilon} = o_P(1)$. Evaluation of this is quite challenging, and the reader is referred to Giraitis and Surgailis (1990). Once this is verified, the convergence of $Q_{n,X}$ follows from the convergence of $Q_{n,\varepsilon} - E(Q_{n,\varepsilon})$. □

The limiting behaviour of quadratic forms becomes more involved if one considers nonlinear functionals. Recall the definition of bivariate Appell polynomials. Redefine Q_n as

$$Q_n(u) = Q_n(u; P_{m_1,m_2}) = \sum_{t,s=1}^{[nu]} b_{t-s} \big\{P_{m_1,m_2}(X_t, X_s) - E\big[P_{m_1,m_2}(X_t, X_s)\big]\big\}.$$

Table 4.4 Panorama of limits for quadratic forms of Gaussian sequences

Quadratic forms—Gaussian sequences (notation: $g(\lambda)=\frac{1}{2\pi}\sum b_l e^{-il\lambda}$)		
$g(0)=\sum_l b_l \neq 0$ $\sum \lvert b_l\rvert < \infty$	$d\in(0,1/2)$ $n^{-1/2}Q_n(u)\Rightarrow cB(u)$ Theorem 4.27	$d\in(1/4,1/2)$ $n^{-2d}Q_n(u)\Rightarrow c\,Z_{2,H}(u)$ Theorem 4.26
$g(\lambda)\sim c_g\lvert\lambda\rvert^{-\gamma}$ $(\lambda\to 0)$	$d\in(0,1/2)$ and $2d+\gamma<1/2$ $n^{-1/2}Q_n(1)\overset{\mathrm{d}}{\to} c\,B(1)$ Theorem 4.28	
$g(\lambda)\sim c_g\lvert\lambda\rvert^{-\gamma}$ $(\lambda\to 0)$	$d\in(0,1/2)$ and $1/2<2d+\gamma<1$ $n^{-(2d+\gamma)}Q_n(u)\Rightarrow cZ(u)$ Theorem 4.29	

Let $B=[b_{j-l}]_{j,l=1}^n$ and $\Sigma^{(m)}=[\gamma_X^m(j-l)]_{j,l=1}^n$. Also, let h^{*m} be the m-fold convolution of a function h. Giraitis and Taqqu (1997) showed that if

$$\lim_{n\to\infty}\frac{\mathrm{Trace}(\Sigma^{(m_1)}B\Sigma^{(m_2)}B)}{n}=\int_{-\pi}^{\pi} f_X^{*m_1}(\lambda)f_X^{*m_2}(\lambda)g^2(\lambda)\,d\lambda<\infty, \qquad (4.133)$$

then $n^{-1/2}Q_n$ converges in distribution to a normal random variable; however the formula for the limiting variance is quite complicated. Condition (4.133) holds if

$$\max\bigl(1-m_1(1-2d),0\bigr)/2+\max\bigl(1-m_2(1-2d),0\bigr)/2+\gamma<1/2. \qquad (4.134)$$

In particular, if $m_1=m_2=1$, then this is equivalent to $2d+\gamma<1/2$, so that we recover (4.131). On the other hand, if $m_1=1$, $m_2=2$, then the condition reads: $3d-1+\gamma<1/2$ if $d\in(1/4,1/2)$; $d+\gamma<3/2$ if $d\in(0,1/4)$.

If (4.134) does not hold, then there is a variety of different possible limits, as presented in Giraitis and Taqqu (1999b). The proofs involve the familiar method based on the multiple Wiener–Itô integrals.

4.5.3 Summary of Limit Theorems for Quadratic Forms

We summarize the main results for quadratic forms of Gaussian sequences in Table 4.4. We assume that X_t $(t\in\mathbb{Z})$ is a centred Gaussian sequence with covariance $\gamma_X(k)\sim c_\gamma k^{2d-1}$, $d\in(0,1/2)$, so that a slowly varying function can be omitted. In what follows, $B(\cdot)$ is a Brownian motion on $[0,1]$, $Z_{2,H}(\cdot)$ is a Hermite–Rosenblatt process on $[0,1]$, and $Z(\cdot)$ is the self-similar process with Hurst parameter $H=2d+\gamma$, as in Theorem 4.29. Furthermore, c is a generic constant.

4.6 Limit Theorems for Fourier Transforms and the Periodogram

In this section we present some basic properties of the Discrete Fourier Transform (DFT) and the periodogram. We analyse their second-order properties showing a remarkable difference between weakly dependent and long-memory linear processes. In particular, the DFT and the periodogram computed at Fourier frequencies are asymptotically independent under short memory but asymptotically dependent under long memory. To achieve asymptotic independence in the latter case, one has to consider the DFT at appropriately high frequencies. The asymptotic dependence of the DFT and the periodogram ordinates implies a different limiting behaviour of the DFT under short and long memory respectively.

4.6.1 Periodogram and Discrete Fourier Transform (DFT)

For an observed second-order stationary time series $X_1, \ldots, X_n$, let $\bar{x} = \bar{x}_n = n^{-1} \sum_{t=1}^n X_t$ and define by

$$\hat{\gamma}_X(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{x})(X_{t+|k|} - \bar{x}) \quad \bigl(|k| \le n-1\bigr),$$

$$\hat{\gamma}_X(k) = 0 \quad \bigl(|k| \ge n\bigr),$$

the sample autocovariances. Also, define the (centred) periodogram by

$$I_{n,X}^{\text{centred}}(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \hat{\gamma}_X(k) e^{-ik\lambda} = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \hat{\gamma}_X(k) e^{-ik\lambda}$$
$$= \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (X_t - \bar{x}) e^{-ik\lambda} \right|^2.$$

If $E[X_1] = \mu = 0$, then $I_{n,X}^{\text{centred}}(\lambda)$ can be approximated by

$$I_{n,X}(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-it\lambda} \right|^2.$$

For Fourier frequencies $\lambda_j = 2\pi j/n$ ($j = 1, \ldots, N_n$; $N_n = [(n-1)/2]$), we have the exact identity $I_{n,X}^{\text{centred}}(\lambda_j) = I_{n,X}(\lambda_j)$ since $\sum_{t=1}^n e^{-it\lambda_j} = 0$. Therefore, in most applications the non-centred periodogram $I_{n,X}$ is used. The non-centred periodogram can be written in terms of the discrete Fourier transform (DFT). Let

$$d_{n,X}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda}.$$

Then clearly $I_{n,X}(\lambda) = |d_{n,X}(\lambda)|^2$.

4.6.2 Second-Order Properties of the Fourier Transform and the Periodogram

4.6.2.1 Mean and Covariance of the DFT and the Periodogram

We are interested in a general expression for the expected value and covariance of the DFT and the periodogram ordinates $I_{n,X}(\lambda_j)$, where λ_j are Fourier frequencies.

Lemma 4.22 *Assume that X_t $(t \in \mathbb{Z})$ is a second-order stationary sequence with mean 0, covariance function γ_X and spectral density f_X. Then $E[d_{n,X}(\lambda_j)] = 0$,*

$$E\left(\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}\right) = \frac{1}{f_X(\lambda_j)} \int_{-\pi}^{\pi} K_n(\lambda_j - \lambda) f_X(\lambda)\, d\lambda$$

and

$$E\left[d_{n,X}(\lambda_j)\, \overline{d_{n,X}(\lambda_j)}\right] = \int_{-\pi}^{\pi} K_n(\lambda - \lambda_j) f_X(\lambda)\, d\lambda, \tag{4.135}$$

where

$$K_n(\lambda) = \frac{1}{2\pi n}\left(\frac{\sin(n\lambda/2)}{\sin(\lambda/2)}\right)^2$$

is the Féjer kernel.

Proof The formula is classical (see Priestley 1981 p. 419), but we give a proof for completeness. We have

$$\begin{aligned} E\left[I_{n,X}(\lambda_j)\right] &= \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} e^{-i(t-s)\lambda_j} E(X_t X_s) \\ &= \frac{1}{2\pi n} \sum_{k=-(n-1)}^{n-1} \big(n - |k|\big) e^{-ik\lambda_j} \gamma_X(k) \\ &= \frac{1}{2\pi n} \int_{-\pi}^{\pi} \left(\sum_{k=-(n-1)}^{n-1} \big(n - |k|\big) e^{-ik(\lambda - \lambda_j)} \right) f_X(\lambda)\, d\lambda. \end{aligned}$$

Furthermore,

$$\begin{aligned} \sum_{t=1}^{n} \sum_{s=1}^{n} e^{-i(t-s)u} &= \sum_{k=-(n-1)}^{n-1} \big(n - |k|\big) e^{iku} \\ &= \frac{1}{2\pi n}\left(\frac{\sin(nu/2)}{\sin(u/2)}\right)^2 = K_n(u). \end{aligned}$$

Similarly, (4.135) follows from

$$E\big[d_{n,X}(\lambda_j)\overline{d_{n,X}(\lambda_j)}\big] = \frac{1}{2\pi n}\sum_{t,s=1}^{n} e^{-i(t-s)\lambda_j}\gamma_X(t-s)$$

$$= \frac{1}{2\pi n}\sum_{t,s=1}^{n} e^{-i(t-s)\lambda_j}\int_{-\pi}^{\pi} e^{i(t-s)\lambda} f_X(\lambda)\,d\lambda$$

$$= \int_{-\pi}^{\pi} K_n(\lambda-\lambda_j) f_X(\lambda)\,d\lambda. \qquad \square$$

Note that the Féjer kernel is also defined by

$$K_n(\lambda) = \frac{1}{2\pi n}\sum_{t,s=1}^{n} e^{-i(t-s)\lambda} = \frac{1}{2\pi n}\big|D_n(\lambda)\big|^2,$$

where

$$D_n(\lambda) = \sum_{t=1}^{n} e^{it\lambda} = \frac{e^{i(n+1)\lambda} - e^{i\lambda}}{e^{i\lambda} - 1}$$

is (a version of) the Dirichlet kernel.

4.6.2.2 Weakly Dependent Sequences

Assume that X_t $(t \in \mathbb{Z})$ is a second-order stationary weakly dependent time series with mean 0. Then (see e.g. Brockwell and Davis 1991) the following holds:

- The periodogram is an asymptotically unbiased estimator of the spectral density:

$$E\big[I_{n,X}(\lambda_j) - f_X(\lambda_j)\big] = O\big(n^{-1}\big) \tag{4.136}$$

  uniformly in $j = 1, \ldots, [n/2]$.
- The periodogram ordinates at Fourier frequencies are asymptotically uncorrelated with correlations converging to zero uniformly:

$$\big|cov\big(I_{n,X}(\lambda_j), I_{n,X}(\lambda_l)\big)\big| \le C_1 n^{-1} \tag{4.137}$$

  with some finite constant C_1.
-

$$\left(\frac{I_{n,X}(\lambda_{j_1})}{f_X(\lambda_{j_1})}, \ldots, \frac{I_{n,X}(\lambda_{j_k})}{f_X(\lambda_{j_k})}\right) \underset{d}{\to} (Z_1, \ldots, Z_k), \tag{4.138}$$

  where $Z_1, \ldots, Z_k$ are i.i.d. standard exponential random variables, and λ_{j_1}, $\ldots, \lambda_{j_k}$ are distinct Fourier frequencies.

On the other hand, it will be shown in a subsequent section that these properties are no longer valid for linear time series with long memory.

Of course, the main tool to establish (4.137) and (4.138) is Lemma 4.22. Note that (cf. Gradshteyn and Rhyzhik 1965, p. 414) $\int_{-\pi}^{\pi} K_n(\lambda_j - \lambda)\, d\lambda = 1$. Thus, if $X_t = \varepsilon_t$ is a centred i.i.d. sequence, then $f_\varepsilon(\lambda) = \sigma_\varepsilon^2/(2\pi)$, and hence,

$$E\left(\frac{I_{n,\varepsilon}(\lambda_j)}{f_\varepsilon(\lambda_j)}\right) = 1 \quad \bigl(j = 1, \dots, [n/2]\bigr), \tag{4.139}$$

independently of the chosen Fourier frequency λ_j. This justifies (4.137) for an i.i.d. sequence. It should be mentioned, though, that this equality is valid at Fourier frequencies only. Furthermore, if ε_t $(t \in \mathbb{Z})$ are i.i.d. with mean zero and variance σ_ε^2, then we have, for distinct Fourier frequencies λ_k, λ_l $(k \neq l)$,

$$E\bigl[d_{n,\varepsilon}(\lambda_k)\,\overline{d_{n,\varepsilon}(\lambda_l)}\,\bigr] = \frac{\sigma_\varepsilon^2}{2\pi} \sum_{t=1}^{n} e^{it(\lambda_k - \lambda_l)} = 0. \tag{4.140}$$

If in addition the random variables ε_t are standard Gaussian, then the discrete Fourier transform at different Fourier frequencies is also jointly Gaussian and hence independent. Consequently, the periodogram ordinates $I_{n,\varepsilon}(\lambda_j) = |d_{n,\varepsilon}(\lambda_j)|^2$ computed at distinct Fourier frequencies are independent. Moreover, $2\pi I_{n,\varepsilon}(\lambda_j)$ $(j = 1, \dots, N_n;\ N_n = [(n-1)/2])$ have a standard exponential distribution. In particular,

$$E\bigl[2\pi I_{n,\varepsilon}(\lambda_j)\bigr] = 1, \qquad \operatorname{var}\bigl(2\pi I_{n,\varepsilon}(\lambda_j)\bigr) = 1. \tag{4.141}$$

If the random variables ε_t are not Gaussian, then $d_{n,\varepsilon}(\lambda_k)$, $d_{n,\varepsilon}(\lambda_l)$ are uncorrelated (i.e. (4.140) still holds), but they are no longer independent. For the periodogram, we have

$$cov\bigl(I_{n,\varepsilon}(\lambda_k), I_{n,\varepsilon}(\lambda_l)\bigr) = \frac{\kappa_4}{4\pi^2 n}, \tag{4.142}$$

where κ_4 is the fourth cumulant. Note that in the Gaussian case $\kappa_4 = 0$. Nevertheless, the periodogram ordinates are *asymptotically* independent and have the standard exponential distribution. This way one obtains (4.138).

4.6.2.3 Linear Long-Memory Sequences

Properties (4.136), (4.137) and (4.138) are not valid in the case of linear process with long memory. The behaviour of the periodogram at frequencies converging to zero can be formulated as follows (Künsch 1986; Hurvich and Beltrao 1993, 1994a, 1994b; Robinson 1995a):

Theorem 4.31 *Let $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ be a second-order stationary linear process and assume that $f_X(\lambda) \sim c_f |\lambda|^{-2d}$ as $|\lambda| \to 0$ with $d \in (0, 1/2)$. Define*

$$\mu(j; d) = |2\pi j|^{2d} \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2(\lambda/2)}{(2\pi j - \lambda)^2} |\lambda|^{-2d} \, d\lambda.$$

Then for any fixed positive integer j,

$$\lim_{n\to\infty} E\left[\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}\right] = \mu(j; d).$$

Proof We use Lemma 4.22. Using the assumption $f_X(\lambda) \sim c_f |\lambda|^{-2d}$, we have

$$\begin{aligned} E\left(\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}\right) &= \frac{1}{n} \int_{-n\pi}^{n\pi} K_n\left(\frac{2\pi j}{n} - \frac{\lambda}{n}\right) \frac{f_X(\lambda/n)}{f_X(2\pi j/n)} \, d\lambda \\ &\approx \left(\frac{2\pi j}{n}\right)^{2d} \frac{1}{n} \int_{-n\pi}^{n\pi} K_n\left(\frac{2\pi j - \lambda}{n}\right) \left|\frac{\lambda}{n}\right|^{-2d} d\lambda \\ &= \frac{1}{n} \int_{-n\pi}^{n\pi} K_n\left(\frac{2\pi j - \lambda}{n}\right) \left|\frac{2\pi j}{\lambda}\right|^{2d} d\lambda. \end{aligned} \tag{4.143}$$

It is easy to see that, as $n \to \infty$, the functions

$$\begin{aligned} g_n(\lambda) &:= \frac{1}{n} K_n\left(\frac{2\pi j - \lambda}{n}\right) \left|\frac{2\pi j}{\lambda}\right|^{2d} \\ &= \frac{1}{2\pi n^2} \frac{\sin^2(\frac{2\pi j - \lambda}{2})}{\sin^2(\frac{2\pi j - \lambda}{2n})} \left|\frac{2\pi j}{\lambda}\right|^{2d} \end{aligned}$$

converge pointwise to

$$\left|\frac{2\pi j}{\lambda}\right|^{2d} \frac{2}{\pi} \frac{\sin^2(\lambda/2)}{(2\pi j - \lambda)^2}.$$

Thus,

$$\lim_{n\to\infty} E\left(\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}\right) = |2\pi j|^{2d} \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2(\lambda/2)}{(2\pi j - \lambda)^2} |\lambda|^{-2d} \, d\lambda,$$

given that we can exchange limit with integration (which follows from Lebesgue dominated convergence) and that integration over $(-\infty, -n\pi) \cup (n\pi, \infty)$ is negligible. □

Detailed calculations can be found in Hurvich and Beltrao (1993). The authors considered a more general spectral density $f_X(\lambda) = |\lambda|^{-2d} f_*(\lambda)$ with a smooth function f_*. In fact, this computation is valid for $d \in (-0.5, 1.5)$; however, if $d > 0.5$, f_X is not a spectral density since the model is not stationary (Hurvich and Ray 1995). What is important here is that the normalized periodogram at Fourier frequencies depends on both j and d, as opposed to the i.i.d. case described in (4.139).

Furthermore, using the same argument as for the mean, Hurvich and Beltrao (1993) argue that for any two integers $l \neq k$,

$$\lim_{n\to\infty} E\left[\frac{d_{n,X}(\lambda_k)\,\overline{d_{n,X}(\lambda_l)}}{\sqrt{f_X(\lambda_k)f_X(\lambda_l)}}\right] =: \gamma_w(l,k;d),$$

where

$$\gamma_w(l,k;d) = (-1)^{l+k+1}|2\pi k|^d |2\pi l|^d \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{\sin^2(\lambda/2)}{(2\pi k-\lambda)(2\pi l+\lambda)}|\lambda|^{-2d}\,d\lambda.$$

Furthermore, if the random variables X_t are Gaussian, then

$$\lim_{n\to\infty} cov\left(\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}, \frac{I_{n,X}(\lambda_k)}{f(\lambda_k)}\right) = \gamma_w^2(j,k;d) + \gamma_w^2(j,-k;d) \quad (j \neq k),$$

$$\lim_{n\to\infty} \mathrm{var}\left(\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)}\right) = 2\gamma_w^2(j,j;d).$$

Thus, unlike the i.i.d. case, the DFTs and the normalized periodogram ordinates are not asymptotically independent.

4.6.2.4 Refined Covariance Bounds for Long-Memory Sequences

One can obtain the following asymptotic independence of the DFT and periodogram ordinates if the Fourier frequencies λ_j are not too close to zero.

Recall that $f_X(\lambda) \sim c_f|\lambda|^{-2d}$ and let

$$d_{n,X}^0(\lambda) = \frac{d_{n,X}(\lambda)}{\sqrt{c_f\lambda^{-2d}}}$$

and $\gamma_X(k) = cov(X_t, X_{t+k})$. Then the following holds.

Theorem 4.32 *Let $X_t = \sum_{j=0}^{\infty} a_j\varepsilon_{t-j}$ be a second-order stationary linear process with*

$$f_X(\lambda) = \left|1-\exp(-i\lambda)\right|^{-2d} f_*(\lambda) \approx |\lambda|^{-2d} f_*(\lambda) \approx c_f|\lambda|^{-2d} \qquad (4.144)$$

and such that

$$f_X(\lambda) = c_f|\lambda|^{-2d} + O\left(\lambda^{\rho-2d}\right) \qquad (4.145)$$

for some $0 < \rho \le 2$ *and* $-\frac{1}{2} < d < \frac{1}{2}$. *Let* j_n, k_n *be positive integer-valued sequences such that* $j_n/n \to 0$ *and* $j_n > k_n$. *Then,*

$$\begin{aligned} \operatorname{var}\big(d_{n,X}^0(\lambda_{j_n})\big) &= E\big[d_{n,X}^0(\lambda_{j_n})\,\overline{d_{n,X}^0(\lambda_{j_n})}\,\big] \\ &= 1 + O\left(\frac{\log j_n}{j_n}\right) + O\left(\left(\frac{j_n}{n}\right)^{\rho}\right) \end{aligned} \tag{4.146}$$

and

$$cov\big(d_{n,X}^0(\lambda_{j_n}), d_{n,X}^0(\lambda_{k_n})\big) = O\left(\frac{\log j_n}{k_n}\right). \tag{4.147}$$

Before we proceed with the proof, we comment on assumption (4.145). This is a smoothness condition for f_*. For example, if $\rho = 2$, then f_* is twice differentiable in the neighbourhood of the origin. This type of condition is crucial in studying for example semiparametric estimators of d.

Proof The essential arguments can be seen by considering (4.146). Condition (4.145) implies

$$\begin{aligned} f_X(\lambda_j) - c_f\lambda_j^{-2d} &= f_X(\lambda_j)\left[1 - \left(\frac{f_X(\lambda_j)}{c_f\lambda_j^{-2d}}\right)^{-1}\right] \\ &= f_X(\lambda_j)\left[1 - \frac{1}{1+O(\lambda_j^{\rho})}\right] \\ &= c_f\lambda_j^{-2d}\big[1 + O\big(\lambda_j^{\rho}\big)\big] = O\big(\lambda_j^{\rho-2d}\big), \end{aligned}$$

so that

$$\frac{f_X(\lambda_j)}{c_f\lambda_j^{-2d}} = 1 + O\left(\left(\frac{j}{n}\right)^{\rho}\right).$$

In a second step, one shows

$$E\big[d_{n,X}(\lambda_j)\,\overline{d_{n,X}(\lambda_j)}\,\big] = f_X(\lambda_j) + O\left(\lambda_j^{-2d}\frac{\log j}{j}\right), \tag{4.148}$$

so that

$$E\left[\frac{d_{n,X}(\lambda_j)\,\overline{d_{n,X}(\lambda_j)}}{f_X(\lambda_j)}\right] = 1 + O\left(\frac{\log j}{j}\right).$$

To show (4.148), we use the general formula for the covariance of DFT; see (4.135). Since K_n is 2π-periodic with $\int_{-\pi}^{\pi} K_n(u)\,du = 1$, we obtain

$$E\big[d_{n,X}(\lambda_j)\,\overline{d_{n,X}(\lambda_j)}\,\big] - f_X(\lambda_j) = \int_{-\pi}^{\pi}\big[f_X(\lambda) - f_X(\lambda_j)\big]K_n(\lambda - \lambda_j)\,d\lambda. \tag{4.149}$$

Now, for n large enough, λ_j is smaller than $\delta/2$, so that

$$f_X(\lambda_j) \le c_\delta \lambda_j^{-2d}, \qquad \left|f_X'(\lambda_j)\right| \le c_\delta \lambda_j^{-2d-1}$$

for a suitable finite constant c_δ. Noting that $K_n(u) = O(n^{-1})$ for $\delta/2 < u \le \pi$, we obtain

$$\begin{aligned}\int_{|\lambda|\ge\delta} \left|f_X(\lambda) - f_X(\lambda_j)\right| K_n(\lambda - \lambda_j)\, d\lambda &\le O\bigl(n^{-1}\bigr) \cdot \left[\int_{-\pi}^{\pi} f_X(\lambda)\, d\lambda + 2\pi c_\delta \lambda_j^{-2d}\right] \\ &= O\bigl(n^{-1}\bigr) + O\bigl(n^{-1}\lambda_j^{-2d}\bigr).\end{aligned}$$

For $0 < d < \frac{1}{2}$, this is of order $O((j/n)^{1-2d} \cdot j^{-1}) = o(j^{-1}\log j)$. Similarly, for $-\frac{1}{2} < d < \frac{1}{2}$, the overall order is $O(n^{-1}) = O((j/n)j^{-1}) = o(j^{-1}\log j)$. Therefore, the only relevant range of integration in (4.143) is $-\delta \le \lambda \le \delta$. There are two asymptotic poles that are approached asymptotically on the right-hand side of (4.149): a pole in f_X for $\lambda_j \to 0$ and an asymptotic singularity in $K_n(\lambda - \lambda_j)$ for $\lambda = \lambda_j$. The largest order is obtained for the integral over $\Delta_n = [\frac{1}{2}\lambda_j, 2\lambda_j]$. There, we have

$$\begin{aligned}&\int_{\lambda\in\Delta_n} \left|f_X(\lambda) - f_X(\lambda_j)\right| K_n(\lambda - \lambda_j)\, d\lambda \\ &\quad\le \max_{\lambda_j/2\le\lambda\le 2\lambda_j} \left|f_X'(\lambda)\right| \underbrace{\int_{\lambda_j/2}^{2\lambda_j} |\lambda - \lambda_j| K(\lambda - \lambda_j)\, d\lambda}_{J(\lambda_j)} = O\bigl(\lambda_j^{-1-2d}\bigr) \cdot J(\lambda_j).\end{aligned}$$

Since $|D_n(u)| \le 2|u|^{-1}$ $(0 < |u| < \pi)$, we have

$$\int_{-c\lambda_j}^{c\lambda_j} \left|D_n(\lambda)\right| d\lambda = O(\log j)$$

for any fixed $c > 0$. Moreover, $\lim_{\lambda\to\lambda_j} |\lambda - \lambda_j| K(\lambda - \lambda_j) = 0$, and we obtain

$$\begin{aligned}|\lambda - \lambda_j| K(\lambda - \lambda_j) &\le (2\pi n)^{-1} |\lambda - \lambda_j| \cdot 2|\lambda - \lambda_j|^{-1} \cdot \left|D_n(\lambda - \lambda_j)\right| \\ &= \pi^{-1} n^{-1} \left|D_n(\lambda - \lambda_j)\right|,\end{aligned}$$

and thus,

$$J(\lambda_j) = O\bigl(n^{-1}\log j\bigr).$$

Putting the orders together, we have

$$\begin{aligned}\int_{\lambda\in\Delta_n} \left|f_X(\lambda) - f_X(\lambda_j)\right| K_n(\lambda - \lambda_j)\, d\lambda &= O\bigl(\lambda_j^{-1-2d} \cdot n^{-1}\log j\bigr) \\ &= O\left(\lambda_j^{-2d} \cdot \frac{\log j}{j}\right),\end{aligned}$$

as required in (4.148). □

4.6.3 Limiting Distribution

4.6.3.1 Fourier Transform and Periodogram for Long-Memory Sequences

Now, we will describe the limiting distribution for the DFT and the periodogram ordinates. Let us write $d_{n,X}(\lambda_j) = A(\lambda_j) + iB(\lambda_j)$, where

$$A(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \cos(t\lambda), \qquad B(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t \sin(t\lambda).$$

Then $I_{n,X}(\lambda_j) = A^2(\lambda_j) + B^2(\lambda_j)$. Assume for simplicity that X_t is a Gaussian process. It follows from (4.147) that for each fixed K,

$$\left(\frac{d_{n,X}(\lambda_j)}{\sqrt{f_X(\lambda_j)}}, j = 1, \ldots, K \right)$$

converges to a multivariate Gaussian distribution with *dependent* components and covariance matrix $[\gamma_w(l,k;d)]_{k,l=1,\ldots,K}$. Furthermore, for each fixed j, the cosine and the sine parts $A(\lambda_j)$ and $B(\lambda_j)$ are uncorrelated with *different variances*. Therefore,

$$\frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j)} = \frac{A^2(\lambda_j)}{f_X(\lambda_j)} + \frac{B^2(\lambda_j)}{f_X(\lambda_j)} \underset{d}{\to} a\chi_1^2(1) + b\chi_1^2(2), \tag{4.150}$$

where a, b are constants, and $\chi_1^2(j)$, $j = 1, 2$, are independent χ^2 random variables with one degree of freedom. Thus, in contrast to the i.i.d. case, the normalized periodogram ordinates have a different asymptotic distribution at each frequency. Moreover, the limiting distribution has dependent components.

4.6.3.2 Sum of Periodogram Ordinates

Let ϕ be a deterministic, real-valued function and consider the partial sum

$$S_{n,X}(\phi) = \sum_{j=1}^{N_n} \phi\big(I_{n,X}(\lambda_j)\big),$$

where $N_n = [(n-1)/2]$. If $X_t = \varepsilon_t$ are i.i.d., then (cf. (4.141))

$$\mathrm{var}\left(\sum_{j=1}^{N_n} 2\pi I_{n,\varepsilon}(\lambda_j) \right) \approx n(1 + \kappa_4/2).$$

Also,

$$n^{-1/2} \sum_{j=1}^{N_n} 2\pi I_{n,\varepsilon}(\lambda_j) \underset{d}{\to} N(0, 1 + \kappa_4/2).$$

These asymptotic results are obvious when ε_t are Gaussian since the periodogram ordinates are independent. If $\phi = \log$ and ε_t are Gaussian, then

$$\operatorname{var}\bigl(\log\bigl(2\pi I_{n,\varepsilon}(\lambda_j)\bigr)\bigr) = \operatorname{var}\bigl(\log\bigl(I_{n,\varepsilon}(\lambda_j)/f_\varepsilon(\lambda_j)\bigr)\bigr) = \operatorname{var}\bigl(\log(Z)\bigr),$$

where Z is standard exponential. We compute

$$\begin{aligned}\operatorname{var}(\log Z) &= \int_0^\infty e^{-x}(\log x)^2\,dx - \left[\int_0^\infty e^{-x}(\log x)\,dx\right]^2 \\ &= \left(\frac{\pi^2}{6} + \eta^2\right) - (-\eta)^2 = \frac{\pi^2}{6}.\end{aligned} \tag{4.151}$$

Therefore, in the Gaussian i.i.d. case,

$$n^{-1/2}\sum_{j=1}^{N_n}\log\bigl(2\pi I_{n,\varepsilon}(\lambda_j)\bigr) \underset{d}{\to} N\bigl(0,\pi^2/6\bigr).$$

In the long-memory case, the periodogram ordinates are asymptotically dependent, so that these convergence results are not valid. However, for a proper choice of asymptotically negligible constants $c_{n,k}$, it is possible to obtain asymptotic normality of $\sum c_{n,k}\phi(I_{n,X}(\lambda_k))$ regardless whether X_t is weakly or strongly dependent. We will illustrate this in the context of semiparametric estimation of the long-memory parameter d.

4.7 Limit Theorems for Wavelets

4.7.1 Introduction

In this section we discuss limit theorems for the discrete wavelet transform of long-memory stochastic processes. We refer to Sect. 3.5 for basic definitions of wavelets. At this point we recall that for a scaling function ϕ and a wavelet function ψ, dilated and translated functions are defined as

$$\phi_{j,k}(x) = 2^{j/2}\phi\bigl(2^j x - k\bigr), \qquad \psi_{j,k}(x) = 2^{j/2}\psi\bigl(2^j x - k\bigr).$$

However, it is not necessary that the wavelet functions are constructed using the multiresolution analysis, nor that they are orthogonal.

4.7.2 Discrete Wavelet Transform of Stochastic Processes

Assume first that $Y(u)$ ($u \in \mathbb{R}$) is a continuous-time stochastic process. Define

$$d_{j,k}^Y = \int_{\mathbb{R}} Y(u)\psi_{j,k}(u)\,du, \qquad a_{j,k}^Y = \int_{\mathbb{R}} Y(u)\phi_{j,k}(u)\,du \quad (j,k \in \mathbb{Z}).$$

In other words, $d^Y_{j,k}$ and $a^Y_{j,k}$ are (random) wavelet coefficients of the continuous-time process $Y(u)$ $(u \in \mathbb{R})$. If the continuous-time process has mean zero, then clearly $E(d_{j,k}) = 0$ for each j, k. For simplicity, we write in the following $a_{j,k}$, $d_{j,k}$ instead of $a^Y_{j,k}$, $d^Y_{j,k}$.

Assume further that $Y(u)$ $(u \in \mathbb{R})$ has stationary increments. For each fixed resolution level j, the process $d_{j,k}$ $(k \in \mathbb{Z})$ is stationary. Indeed, we may verify, for instance, that the marginal distributions are invariant under translation: the random coefficient

$$\begin{aligned} d_{j,k+l} &= \int Y(u)\psi_{j,k+l}(u)\,du = \int Y(u+l)\psi_{j,k}(u)\,du \\ &= \int \big(Y(u+l) - Y(l)\big)\psi_{j,k}(u)\,du \end{aligned}$$

is equal in distribution to

$$\int \big(Y(u) - Y(0)\big)\psi_{j,k}(u)\,du = \int Y(u)\psi_{j,k}(u)\,du = d_{j,k}.$$

The same applies to the scaling coefficients $a_{j,k} = \int Y(u)\phi_{j,k}(u)\,du$. A more rigorous proof of stationarity can be found in e.g. Houdré (1994). See also Masry (1993) and Cambanis and Houdré (1995) for the DWT of stochastic processes.

If moreover, the process $Y(u)$ is H-self-similar, then for each j, k,

$$d_{j,k} \stackrel{\mathrm{d}}{=} 2^{-j(H+1/2)} d_{0,k}.$$

Indeed, heuristically,

$$\begin{aligned} d_{j,k} &= \int Y(u)\psi_{j,k}(u) = 2^{j/2} \int Y(u)\psi\big(2^j u - k\big)\,du \\ &= 2^{-j/2} \int Y\big(2^{-j}u\big)\psi(u-k)\,du \stackrel{\mathrm{d}}{=} 2^{-j/2} 2^{-jH} \int Y(u)\psi_{j,k}(u) \\ &= 2^{-j(H+1/2)} d_{0,k}. \end{aligned}$$

Hence, if the continuous-time process $Y(u)$ $(u \in \mathbb{R})$ is self-similar with stationary increments (H-SSSI), then

$$E\big[d^2_{j,k+l}\big] = 2^{-j(2H+1)} E\big[d^2_{0,k}\big] = 2^{-j(2H+1)} E\big[d^2_{0,0}\big].$$

This applies, in particular, to fractional Brownian motion. As we will see later, these formulas can be used to define a wavelet-based estimator of the self-similarity parameter H.

4.7.3 Second-Order Properties of Wavelet Coefficients

Now, we turn our attention to stationary processes $X(u)$ ($u \in \mathbb{R}$). For example, $X(u) = Y(u) - Y(u-1)$ ($u \in \mathbb{R}$) can be defined as increments of the H-SSSI process considered above. Define analogously wavelet and scaling coefficients:

$$d_{j,k} = d_{j,k}^X = \int_{\mathbb{R}} X(u)\psi_{j,k}(u)\,du,$$
$$a_{j,k} = a_{j,k}^X = \int_{\mathbb{R}} X(u)\phi_{j,k}(u)\,du \quad (j,k \in \mathbb{Z}).$$

Then $d_{j,k}$ and $a_{j,k}$ ($k \in \mathbb{Z}$) form stationary sequences. We verify for instance that the marginal distributions are shift-invariant: for $l \in \mathbb{Z}$, we have

$$\begin{aligned} d_{j,k+l} &= \int_{-\infty}^{\infty} X(u)\psi_{j,k+l}(u)\,du = 2^{j/2}\int_{-\infty}^{\infty} X(u)\psi\big(2^j u - (k+l)\big)\,du \\ &= 2^{j/2}\int_{-\infty}^{\infty} X\big(v + 2^{-j}l\big)\psi\big(2^j v - k\big)\,dv \overset{\mathrm{d}}{=} 2^{j/2}\int_{-\infty}^{\infty} X(v)\psi\big(2^j v - k\big)\,dv \\ &= d_{j,k}. \end{aligned}$$

Hence, we can analyse the covariance structure of the stationary sequence $d_{j,k}$ ($k \in \mathbb{Z}$). Assume that the process $X(u)$ ($u \in \mathbb{R}$) is centred, has the covariance function $\gamma_X(s)$ ($s \in \mathbb{R}$) and the spectral density

$$f_X(\lambda) = \int_{-\infty}^{\infty} \gamma_X(s)e^{-i\lambda s}\,ds.$$

Assume further that

$$f_X(\lambda) = \lambda^{-2d} f_*(\lambda), \quad \lambda \to 0,$$

where $\lim_{\lambda\to 0} f_*(\lambda) = c_f \in (0,\infty)$ and $d \in [0, 1/2)$. For example, $X(u)$ could be fractional Gaussian noise, i.e. increments of fractional Brownian motion with Hurst parameter $H = d + \frac{1}{2}$.

One of the most intriguing properties of DWT is the *decorrelation* (*whitening*) *property*. Specifically, if the wavelet ψ has M vanishing moments, then we will argue below that

$$cov(d_{j,0}, d_{j,k}) = O\big(k^{-2M+2d-1}\big) \quad (k \to \infty).$$

That is, the stationary sequence $d_{j,k}$ ($k \in \mathbb{Z}$) is *weakly* dependent (i.e. has summable covariances) if $M \geq 1$. For example, the *whitening property* applies to fractional Gaussian noise $X(u) = B_H(u) - B_H(u-1)$, where $B_H(u)$ is a fractional Brownian motion with Hurst parameter $H \in (1/2, 1)$. This phenomenon is discussed for instance in Flandrin (1992), Tewfik and Kim (1992), Abry et al. (1998) or Mielniczuk and Wojdyłło (2007a).

To justify the whitening property, recall that

$$\hat{\psi}(\lambda) = \int_{-\infty}^{\infty} \psi(x) e^{-i\lambda x}\, dx$$

is the Fourier transform of ψ. Hence,

$$\begin{aligned}\hat{\psi}_{j,k}(\lambda) &= \int_{-\infty}^{\infty} e^{-i\lambda x} \psi_{j,k}(x)\, dx = 2^{j/2} \int_{-\infty}^{\infty} e^{-i\lambda x} \psi\big(2^j x - k\big)\, dx \\ &= 2^{-j/2} e^{-i2^{-j}\lambda k} \int_{-\infty}^{\infty} e^{-i\lambda 2^{-j} x} \psi(x)\, dx = 2^{-j/2} e^{-i2^{-j}\lambda k} \hat{\psi}\big(2^{-j}\lambda\big).\end{aligned}$$

We can then evaluate covariance structure of the wavelet coefficients of the process $X(\cdot)$ as

$$\begin{aligned} &cov(d_{j,k}, d_{j',k'}) \\ &\quad = \int\int \gamma_X(v-u) \psi_{j,k}(v) \psi_{j',k'}(u)\, du\, dv \\ &\quad = \int_{-\infty}^{\infty} f_X(\lambda) \hat{\psi}_{j,k}(\lambda) \hat{\psi}_{j',k'}(\lambda)\, d\lambda \\ &\quad = 2^{-j/2} 2^{-j'/2} \int_{-\infty}^{\infty} f_X(\lambda) \hat{\psi}\big(2^{-j}\lambda\big) \overline{\hat{\psi}\big(2^{-j'}\lambda\big)} e^{-i2^{-j}\lambda k} e^{i2^{j'}\lambda k'}\, d\lambda. \end{aligned} \tag{4.152}$$

This formula is crucial to evaluate the variance and covariance structure of the wavelet coefficients for stochastic processes with long memory. A change of variables $\omega = 2^{-(j+j')/2}\lambda$,

$$\lambda = 2^{(j+j')/2}\omega,$$

and the form $f(\lambda) = \lambda^{-2d} f_*(\lambda)$ of the spectral density yield

$$\begin{aligned} &cov(d_{j,k}, d_{j\prime,k\prime}) \\ &\quad = \int_{-\infty}^{\infty} f_X\big(2^{(j+j')/2}\omega\big) \hat{\psi}\big(2^{(j'-j)/2}\omega\big) \overline{\hat{\psi}\big(2^{(j-j')/2}\omega\big)} e^{-i2^{(j-j')/2}\omega k} e^{i2^{(j'-j)/2}\omega k'}\, d\omega \\ &\quad = 2^{-(j+j')d} \int_{-\infty}^{\infty} \omega^{-2d} f_*\big(2^{(j+j')/2}\omega\big) \hat{\psi}\big(2^{(j'-j)/2}\omega\big) \overline{\hat{\psi}\big(2^{(j-j')/2}\omega\big)} e^{-ir\omega}\, d\omega, \end{aligned}$$

where

$$r = \big|2^{(j-j')/2}k - 2^{(j'-j)/2}\omega k'\big|.$$

When $j, j' \to -\infty$ (i.e. we are considering coarse resolution levels or "low frequencies"), then $2^{(j+j')/2}\omega \to 0$, so that

$$f_*\big(2^{(j+j')/2}\omega\big) \sim f_*(0) = c_f.$$

This motivates the following definition:

$$\Psi_{j,j'}(k,k') := \int_{-\infty}^{\infty} \omega^{-2d} \hat{\psi}\big(2^{(j'-j)/2}\omega\big) \overline{\hat{\psi}\big(2^{(j-j')/2}\omega\big)} e^{-ir\omega}\, d\omega. \tag{4.153}$$

We note that if $j \neq j'$, $k \neq k'$ and $d = 0$, then, due to orthogonality, the covariances vanish *if the wavelet family* $\psi_{j,k}$ *is constructed using the MRA*. As we will see below, in the case of long memory, orthogonality of wavelets is not crucial at all. The most important property is the number M of vanishing moments of the wavelet function ψ.

To see this, let $d > 0$ and consider $j = j'$ and $k' = 0$. Then

$$cov(d_{j,0}, d_{j,k}) = 2^{-2jd} \int \omega^{-2d} f_*\big(2^j\omega\big) \big|\hat{\psi}(\omega)\big|^2 e^{-ik\omega}\, d\omega.$$

Again, as $j \to -\infty$, we approximate this integral as

$$cov(d_{j,0}, d_{j,k}) = 2^{-2jd} f_*(0) \int \omega^{-2d} \big|\hat{\psi}(\omega)\big|^2 e^{-ik\omega}\, d\omega.$$

Next, recall now from Sect. 3.5 that if the wavelet function ψ has M vanishing moments, then

$$\big|\hat{\psi}(\lambda)\big| = \big|\hat{\psi}^{(M)}(0)\big| |\lambda|^M + o\big(|\lambda|^M\big) \quad (\lambda \to 0).$$

Thus, if k is large enough, then we have to analyse the following integral in a neighbourhood $(-\varepsilon/k, \varepsilon/k)$ of the origin:

$$2^{-2jd} c_f \big\{\hat{\psi}^{(M)}(0)\big\}^2 \int_{\varepsilon/k}^{\varepsilon/k} \omega^{-2d} \omega^{2M} e^{-ik\omega}\, d\omega.$$

The change of variables $\lambda = k\omega$ yields the approximation

$$2^{-2jd} c_f \big\{\hat{\psi}^{(M)}(0)\big\}^2 k^{-2M+2d-1} \int_{-\varepsilon}^{\varepsilon} \lambda^{2M-2d} e^{-i\lambda}\, d\lambda.$$

The integral is finite as long as $2M - 2d > -1$. Of course, in these computations several simplifications and informal approximations are used. Nevertheless, we have obtained heuristically the following *decorrelation property*.

Lemma 4.23 *Assume that* $X(u)$ $(u \in \mathbb{R})$ *is a stationary centred process such that its spectral density is given by* $f_X(\lambda) = |\lambda|^{-2d} f_*(\lambda)$, $\lambda \in \mathbb{R}$, $d \in (0, 1/2)$ *and* $\lim_{\lambda \to 0} f_*(\lambda) = c_f \in (0, \infty)$. *Then for each* $j \in \mathbb{Z}$,

$$cov(d_{j,0}, d_{j,k}) = O\big(k^{-2M+2d-1}\big) \quad (k \to \infty).$$

The same result carried over to series X_t $(t \in \mathbb{Z})$ in discrete time, when transformed into their continuous-time versions as discussed in the introduction to wavelets. In particular, the restrictions $d < \frac{1}{2}$ and $M \geq 1$ imply that we always have $cov(d_{j,0}, d_{j,k}) = o(k^{-2})$. This means that

$$\sum_{k=-\infty}^{\infty} \left|cov(d_{j,0}, d_{j,k})\right| < \infty$$

and the wavelet coefficients $d_{j,k}$ ($k \in \mathbb{Z}$) are weakly dependent. Moreover, if the process $X(u)$ ($u \in \mathbb{R}$) is Gaussian, then the wavelet coefficients are Gaussian as well. Also, in the Gaussian case we have

$$cov\left(d_{j,0}^2, d_{j,k}^2\right) = 2cov^2(d_{j,0}, d_{j,k}),$$

so that these autocovariances converge as well.

As indicated above, a very useful property is also (4.153) because for large enough scales, i.e. for $j, j' \to -\infty$,

$$cov(d_{j,k}, d_{j',k'}) \approx 2^{-(j+j')d} f_*(0) \Psi_{j,j'}\left(k, k'\right).$$

Thus, the weak dependence extends to the wavelet coefficients at different resolution levels $j \neq j'$.

To evaluate the variance of $d_{j,k}$, set $j = j'$, $k = k'$ in (4.152). Then

$$\begin{aligned}\sigma_j^2 := \mathrm{var}(d_{j,k}) &= 2^{-j} \int f_X(\lambda) \left|\hat{\psi}\left(2^{-j}\lambda\right)\right|^2 d\lambda \\ &= 2^{-2jd} \int |\lambda|^{-2d} f_*\left(2^j\lambda\right) \left|\hat{\psi}(\lambda)\right|^2 d\lambda.\end{aligned}$$

Again, we approximate $f_*(2^j\lambda) \approx f_*(0) = c_f$ (for $j \to -\infty$) and hence

$$\mathrm{var}(d_{j,k}) \approx 2^{-2jd} c_f \int |\lambda|^{-2d} \left|\hat{\psi}(\lambda)\right|^2 d\lambda =: 2^{-2jd} c_f \Psi(2d), \tag{4.154}$$

where

$$\Psi(\gamma) = \int \lambda^{-\gamma} \left|\hat{\psi}(\lambda)\right|^2 d\lambda.$$

This heuristic approximation has been derived in Abry et al. (1998). More precise bounds have been obtained in Lemma 1 in Bardet et al. (2000) or Theorem 1 in Moulines et al. (2007a). A bound that requires a semiparametric assumption on the spectral density similar to the one used for the DFT is for instance:

Lemma 4.24 *Assume that for some* $d \in (0, 1/2)$,

$$f_X(\lambda) = \lambda^{-2d}\left(f_*(0) + O\left(|\lambda|^\rho\right)\right).$$

Under appropriate regularity conditions, we have, as $j \to -\infty$,

$$\left|\mathrm{var}(d_{j,k}) - 2^{-2jd} c_f \Psi(2d)\right| \leq 2^{-2jd} 2^{j\rho} \Psi(2d - \rho).$$

Proof In the proof, we omit several details, referring to the papers mentioned above. We note that

$$\left|\operatorname{var}(d_{j,k}) - 2^{-2jd} c_f \Psi(2d)\right| \le 2^{-2jd} \int |\lambda|^{-2d} \left|\left\{f_*\left(2^j \lambda\right) - f_*(0)\right\}\right| \left|\hat{\psi}(\lambda)\right|^2 d\lambda.$$

Under the assumption

$$f_*(\lambda) = |\lambda|^{-2d}\left(f_*(0) + O\left(|\lambda|^\rho\right)\right),$$

the bound is

$$2^{-2jd} \int |\lambda|^{-2d} \left\{2^j \lambda\right\}^\rho \left|\hat{\psi}(\lambda)\right|^2 d\lambda = 2^{-2jd} 2^{j\rho} \Psi(2d - \rho).$$ □

4.8 Limit Theorems for Empirical and Quantile Processes

4.8.1 Linear Processes with Finite Moments

The empirical distribution function plays an essential role in statistical inference. Many statistics that are concerned with inference for the marginal distribution of a process can be written as functionals of the (marginal) empirical distribution function $F_n(x)$. Therefore, in principle, their distribution follows "automatically", once the empirical distribution function is characterized asymptotically. Sometimes, the functionals are quite involved however so that the derivation requires some additional work. Relatively simple functionals occur for instance in goodness-of-fit tests, and even more directly in quantile estimation. For obvious reasons, limiting results for quantile processes follow directly from those for the empirical distribution function.

Recall that for a stationary process X_t $(t \in \mathbb{Z})$ with marginal distribution function $F_X(x) = P(X \le x)$, a simple nonparametric estimator of F_X is the (marginal) empirical distribution function

$$F_{n,X}(x) = \frac{1}{n} \sum_{t=1}^{n} 1\{X_t \le x\} \quad (x \in \mathbb{R}). \tag{4.155}$$

Under very general assumptions (for example ergodicity of the sequence), $F_{n,X}$ is a uniformly consistent estimator of F_X, which means that, as $n \to \infty$,

$$\sup_{x \in \mathbb{R}} \left|F_{n,X}(x) - F_X(x)\right| \underset{p}{\to} 0. \tag{4.156}$$

Furthermore, if X_t $(t \in \mathbb{Z})$ are i.i.d., then the classical Donsker invariance principle states

$$\sqrt{n} E_{n,X}(x) := \sqrt{n}\left[F_{n,X}(x) - F_X(x)\right] \Rightarrow \tilde{B}\left(F_X(x)\right), \tag{4.157}$$

where $\Rightarrow$ denotes weak convergence in $D[0,\infty)$, and $\tilde{B}(u)$ ($u \in [0,1]$) is a Brownian bridge, i.e. $\tilde{B}(u) = B(u) - uB(u)$ where $B(u)$ is standard Brownian motion. In other words, the appropriately normalized empirical processes $E_{n,X}(x)$ converge weakly to the time-changed Brownian bridge. An analogous result, with the same normalizing rate but a different limiting process, holds for weakly dependent processes under very general conditions. The situation is quite different, however, under long memory. This can be seen as follows. The indicator function is a very specific transformation of X, i.e. we consider

$$G(X; x) = 1\{X \le x\} - F_X(x).$$

Let $p_X = F_X'$ be the density of X. With the function $y \to G(y; x)$ we can associate the Appell coefficients $a_{\mathrm{app},j}$ ($j \ge 1$):

$$\begin{aligned} a_{\mathrm{app},j} &= (-1)^j \int G(y; x) p_X^{(j)}(y)\, dy \\ &= (-1)^j \left[\int_{-\infty}^{x} p_X^{(j)}(y)\, dy - F_X(x) \int_{-\infty}^{\infty} p_X^{(j)}(y)\, dy \right] \\ &= (-1)^j \int_{-\infty}^{x} p_X^{(j)}(y)\, dy = (-1)^j p_X^{(j-1)}(x). \end{aligned}$$

Furthermore, recall also (see Definition 4.1) that $G_\infty(y) = E[G(X+y)]$. Applying this to $G(y; x) = 1\{y \le x\}$, we obtain $G_\infty(y) = P(X \le x - y)$, and hence,

$$G_\infty^{(1)}(0) = -p_X(x-y)|_{y=0} = -p_X(x).$$

Therefore, the theory for partial sums of subordinated long-memory processes (considered e.g. in Sects. 4.2, 4.3) will imply the limiting behaviour for the empirical distribution $F_{n,X}(x)$ function when x is fixed.

The asymptotic behaviour of the empirical process based on long-memory linear processes with finite variance was studied in Dehling and Taqqu (1989b), Giraitis and Surgailis (1999), Ho and Hsing (1996), Giraitis et al. (1997), Wu (2003) and Csörgő et al. (2006), Csörgő and Kulik (2008a, 2008b). Here, we state the result under the assumptions that are needed to apply the martingale expansion technique of Ho and Hsing (1996) and Wu (2003), as considered in Theorem 4.9. When dealing with linear processes, this technique seems to be superior to the Appell expansion.

Theorem 4.33 *Let X_t ($t \in \mathbb{Z}$) be a linear process $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ with coefficients satisfying assumption* (B1), *i.e. $a_j \sim L_a(j) j^{d-1}$, $d \in (0, 1/2)$ (so that $\gamma_X(k) \sim L_\gamma(k) k^{2d-1}$). Also, assume that $E(|\varepsilon_1|^{4+\gamma}) < \infty$ for some $\gamma > 0$ and that p_ε, the density of the innovations, is such that*

$$\sup_{x \in \mathbb{R}} \left| p_\varepsilon^{(r)}(x) \right| + \int \left| p_\varepsilon^{(r)}(x) \right|^2 dx < \infty \quad (r = 0, 1, 2). \tag{4.158}$$

Then we have the uniform reduction principle

$$n^{\frac{1}{2}-d}L_1^{-\frac{1}{2}}(n)\sup_{x\in\mathbb{R}}\left|F_{n,X}(x)-F_X(x)+p_X(x)\bar{x}\right|\to_p 0. \tag{4.159}$$

Consequently,

$$n^{\frac{1}{2}-d}L_1^{-\frac{1}{2}}(n)\left[F_{n,X}(x)-F_X(x)\right]\Rightarrow p_X(x)Z, \tag{4.160}$$

where $L_1(n)=(d(2d+1))^{-1}L_\gamma(n)$, $\Rightarrow$ *denotes weak convergence in* $D(-\infty,\infty)$, *and* Z *is a standard normal random variable.*

Remark 4.4 Condition (4.158) implies that the same holds for the density p_X. In particular, the conditions on $p_X^{(1)}(x)$ and $p_X^{(2)}(x)$ are required to control a remainder term in the second-order expansion leading to (4.159). Note also that the assumptions of the theorem can be modified to $E(|\varepsilon_1|^{2+\gamma})<\infty$ and

$$\left|E\left[\exp(is\varepsilon_1)\right]\right|\le C\left(1+|s|\right)^{\delta} \tag{4.161}$$

for some $\delta>0$, $0<C<\infty$. Condition (4.161) means in principle that p_X is infinitely often differentiable. These assumptions were used in Giraitis and Surgailis (1999). The authors were also able to deal with double-sided linear processes, however, at the cost of additional moment assumptions.

Remark 4.5 Under the conditions of Theorem 4.33, the finite-dimensional convergence in (4.160) follows directly from Theorem 4.9 and Corollary 4.3. Tightness is usually not proven directly, but rather follows from the reduction principle (4.159). For the latter, we refer to Dehling and Taqqu (1989b) or Csörgö, Szyszkowicz and Wang in the Gaussian case and to Ho and Hsing (1996) and Wu (2003) in the linear case.

Proof We repeat the martingale approximation argument presented before Theorem 4.9, adapting it to the indicator function $G(y;x)=1\{y\le x\}$. Recall that $\mathcal{F}_K=\sigma(\varepsilon_j, j\le K)$ is the σ-algebra generated by ε_j $(j\le K)$. We start with an orthogonal expansion of the indicator function,

$$1\{X_t\le x\}-F_X(x)\underset{L_X^2(\Omega)}{=}\sum_{j=0}^{\infty}\zeta_t(j),$$

where

$$\zeta_t(j)=P(X_t\le x|\mathcal{F}_{t-j})-P(X_t\le x|\mathcal{F}_{t-j-1}).$$

Note that $\zeta_t(0)=1\{X_t\le x\}-P(X_t\le x|\mathcal{F}_{t-1})$. As before, the nice feature of this expansion is that, for fixed t, $\zeta_t(j)$ $(j=0,1,2,\ldots)$ is a martingale difference, so that we indeed obtain orthogonality in the sense that for $j\ne j^*$,

$$\left\langle\zeta_t(j),\zeta_t\left(j^*\right)\right\rangle=cov\left(\zeta_t(j),\zeta_t\left(j^*\right)\right)=0.$$

In more concrete terms, we have

$$P(X_i \le x|\mathcal{F}_{t-j}) = P\left(\sum_{s=0}^{j-1} a_s\varepsilon_{t-s} \le x - \sum_{s=j}^{\infty} a_s\varepsilon_{t-s}\right) = F_j(u_j),$$

where, given $\mathcal{F}_{t-j}$, the argument

$$u_j = x - \sum_{s=j}^{\infty} a_s\varepsilon_{t-s}$$

is fixed (of course, u_j depends on t as well, but this dependence is omitted). Similarly,

$$F_{j+1}(u_{j+1}) = P(X_t \le x|\mathcal{F}_{t-j-1}) = P\left(\sum_{s=0}^{j} a_s\varepsilon_{t-s} \le x - \sum_{s=j+1}^{\infty} a_s\varepsilon_{t-s}\right).$$

Note that $u_{j+1} = u_j - a_j\varepsilon_{t-j}$ and

$$\zeta_t(j) = F_j(u_j) - F_{j+1}(u_{j+1}).$$

A heuristic argument leads to the idea how one may obtain a linearization. We will use the notation $p_j(u) = F_j'(u)$ for the probability density function of $\sum_{s=0}^{j-1} a_s\varepsilon_{t-s}$ and $F_\varepsilon(y) = P(\varepsilon \le y)$. For $F_{j+1}(u_{j+1})$, we can write

$$F_{j+1}(u_{j+1}) = \int p_j(y)F_\varepsilon\big(q_j(x,y)\big)\,dy$$

with

$$q_j(x,y) = \frac{u_{j+1}(x) - y}{a_j}.$$

For the sake of argument, assume that $a_j > 0$ for j large enough. Since $a_j \to 0$ (as $j \to \infty$), we have $q_j \to \infty$ and $F_\varepsilon(q_j(x,y)) \to 1$ if $y < u_{j+1}(x)$. On the other hand, $q_j \to -\infty$ and $F_\varepsilon(q_j(x,y)) \to 0$, if $y > u_{j+1}$. Therefore, as $j \to \infty$,

$$F_{j+1}(u_{j+1}) \approx \int_{-\infty}^{u_{j+1}} p_j(y)\,dy = F_j(u_{j+1}).$$

Furthermore, using $u_j = u_{j+1} - a_j\varepsilon_{t-j}$ with $a_j\varepsilon_{t-j} \to 0$ in probability as $j \to \infty$, we obtain in first approximation

$$F_j(u_j) \approx F_j(u_{j+1}) - p_j(u_{j+1})a_j\varepsilon_{t-j},$$

so that

$$\begin{aligned}\zeta_t(j) &= F_j(u_j) - F_{j+1}(u_{j+1}) \\ &\approx \left[F_j(u_{j+1}) - p_j(u_{j+1})a_j\varepsilon_{t-j}\right] - F_j(u_{j+1}) \\ &= -p_j(u_{j+1})a_j\varepsilon_{t-j}.\end{aligned}$$

Finally, as $j \to \infty$, F_j converges to F_X (and p_j to f_X) and u_{j+1} to x, so that we may hope to obtain the following approximation:

$$\begin{aligned}F_{n,X}(x) - F_X(x) &= \frac{1}{n}\sum_{t=1}^{n}\left[1\{X_t \le x\} - F_X(x)\right] \\ &\approx \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{j=0}^{\infty} -p_j(u_{j+1})a_j\varepsilon_{t-j}\right) \\ &\approx -p_X(x)\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{j=0}^{\infty} a_j\varepsilon_{t-j}\right) = -p_X(x)\bar{x}.\end{aligned}$$

A precise computation establishes the rate in (4.159). □

Taking into account higher-order terms in the Taylor expansions above, a complete orthogonal decomposition can be obtained:

$$F_{n,X}(x) - F_X(x) = \frac{1}{n}\sum_{t=1}^{n}\sum_{r=1}^{\infty}(-1)^k F_X^{(r)}(x)V_{t,r} \tag{4.162}$$

with

$$V_{t,r} = \sum_{0\le j_1<j_2<\cdots<j_r}^{\infty}\prod_{s=1}^{r} a_{j_s}\varepsilon_{t-j_s},$$

already defined in (4.51).

Theorem 4.33 is remarkable not only because of the slower rate of convergence under long memory, but also because the asymptotic process $p_X(x)Z$ (in x) is degenerate. The entire sample path is determined by one normal variable Z and a deterministic function $p_X(x)$. In other words, all sample paths have the shape of $p_X(x)$! This is in sharp contrast to the case of weak memory where the asymptotic process is proportional to a Brownian bridge (see (4.157) above).

The convergence (4.160) can be extended further. In addition to (4.158), assume that the condition holds with $r = 3$. Then the following holds:

- If $d \in (1/4, 1/2)$, then

$$n^{1-2d}L_2^{-1/2}(n)\left[F_{n,X}(x) - F_X(x) + p_X(x)\bar{x}\right] \Rightarrow p_X^{(1)}(x)Z_{2,H}(1), \tag{4.163}$$

  where $Z_{2,H}(1)$ is the Hermite–Rosenblatt random variable, and $H = d + 1/2$.

- If $d \in (0, 1/4)$, then

$$\sqrt{n}\big[F_{n,X}(x) - F_X(x) + p_X(x)\bar{x}\big] \Rightarrow Z(x), \tag{4.164}$$

where $Z(\cdot)$ is a Gaussian process.

Essentially, these convergence results are very similar to the case of nonlinear functionals. The asymptotic behaviour of

$$F_{n,X}(x) - F_X(x) + p_X(x)\bar{x}$$

is determined by $\frac{1}{2}p_X^{(1)}(x)n^{-1}U_{n,2}$, where

$$U_{n,2} = 2!\sum_{t=1}^{n}\sum_{0=j_1<j_2}^{\infty} a_{j_1}a_{j_2}\varepsilon_{t-j_1}\varepsilon_{t-j_2}$$

is defined in (4.51).

Furthermore, Theorem 4.33 can be extended to subordinated processes $Y_j = \tilde{G}(X_t)$. As expected from Theorem 4.4 (Gaussian case) or Theorem 4.8 (the linear case), the rate of convergence and the asymptotic distribution depends on the Appell (or, equivalently, the power) rank of

$$G(X;x) = 1\big\{\tilde{G}(X) \le x\big\} - F_Y(x).$$

The limiting process is a Hermite–Rosenblatt random variable multiplied by a deterministic function.

4.8.2 Applications and Extensions

4.8.2.1 Quantile Processes and Trimmed Sums

Weak convergence (4.160) for empirical processes based on LRD linear sequences has immediate implications for sample quantiles. For $y \in (0, 1)$, define the quantile function

$$Q_X(y) = F_X^{-1}(y) = \inf\big\{x : F_X(x) \ge y\big\}.$$

We will assume that F_X and Q_X are differentiable, so that

$$Q_X(y) = \inf\big\{x : F_X(x) = y\big\}.$$

In an analogous manner, the empirical quantile function is defined as $Q_{n,X}(y) = F_{n,X}^{-1}(y)$ with $F_{n,X}$ defined in (4.155). By definition, $Q_{n,X}$ is left-continuous. Noting that for $x = Q_X(y)$,

$$Q'_X(y) = \frac{1}{p_X(x)},$$

(4.160) implies

$$L_1^{-\frac{1}{2}}(n)n^{\frac{1}{2}-d}\left[Q_{n,X}(y) - Q_X(y)\right] \Rightarrow Z, \tag{4.165}$$

where Z is a standard normal random variable, and the convergence is in $D[a,b]$ equipped with the sup-norm for $0 < a < b < 1$. It is remarkable that the limiting variable does not depend on y (this is of course due to the degenerate structure of the limiting process in (4.160)). A detailed evaluation and further extensions can be found in Ho and Hsing (1996), Wu (2005), Csörgő et al. (2006), Youndjé and Vieu (2006), Csörgő and Kulik (2008a, 2008b) or Coeurjolly (2008a, 2008b).

The result for the quantile function can be extended to trimmed sums

$$T_{n,h} := \frac{1}{n - 2[nh]} \sum_{t=[nh]+1}^{n-[nh]} X_{t:n}, \tag{4.166}$$

where $h \in (0, 1/2)$, and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics. Then

$$L_1^{-\frac{1}{2}}(n)n^{\frac{1}{2}-d} T_{n,h} \to_d Z.$$

See Ho and Hsing (1996), Wu (2003) or Kulik and Ould Haye (2008).

Note, however, that the weak convergence (4.165) cannot be extended to $(0,1)$. Similarly, the result (4.166) does not hold for sums of extremes $\sum_{t=1}^{[nh]} X_{t:n}$ or $\sum_{t=n-[nh]}^{n} X_{t:n}$. There, the limiting behaviour depends on an interplay between the dependence parameter d and the heaviness of tails of the random variables X_t. We refer to Kulik (2008a) for details. Similar issues will be discussed in Sect. 4.8.5 in connection with tail empirical processes.

4.8.2.2 Goodness-of-Fit Test

An immediate consequence for statistical inference is for instance an unusual behaviour of the Kolmogorov–Smirnov statistic, namely

$$L_1^{-\frac{1}{2}}(n)n^{\frac{1}{2}-d} T_{\mathrm{KS},n} := L_1^{-\frac{1}{2}}(n)n^{\frac{1}{2}-d} \sup_{x\in\mathbb{R}} \left|F_{n,X}(x) - F_X(x)\right| \underset{d}{\to} |Z| \sup_{x\in\mathbb{R}} p_X(x), \tag{4.167}$$

given that $\sup_{x\in\mathbb{R}} p_X(x) < \infty$. Therefore, we may approximate p-values by

$$P(T_{KS,n} > u) \approx 2\bar{\Phi}\left(\frac{u}{\sup_{x\in\mathbb{R}} p_X(x)} L_1^{\frac{1}{2}}(n) n^{d-\frac{1}{2}}\right), \tag{4.168}$$

where $u \ge 0$, Φ is the cumulative standard normal distribution, and $\bar{\Phi} = 1 - \Phi$. Note in particular that for a given density, the value $\sup_{x\in\mathbb{R}} p_X(x)$ is known. Of course, in general one has to estimate the dependence parameter d.

In contrast, for weakly dependent processes, the supremum of the transformed Brownian bridge $\tilde{B} \circ F$ over the interval $[0,1]$ is required.

4.8.3 Empirical Processes with Estimated Parameters

Consider the assumptions of Theorem 4.33. As mentioned previously, a direct statistical application of the limiting behaviour of the empirical process is the Kolmogorov–Smirnov statistic, as established in (4.167). As explained in (4.168), this result can be used, in principle, to test whether the marginal distribution F_X of an observed series $X_1, \ldots, X_n$ is equal to a specific distribution F^0. Usually, however, one needs to test whether F_X belongs to a certain type of distributions, instead of one fixed F^0. For instance, we would like to test whether F_X is in a parametric family $\{F_X(\cdot, \theta), \theta \in \mathbb{R}\}$, without specifying the parameter θ a priori. The nuisance parameter θ has to be estimated from the observed series. Thus, instead of $T_{\mathrm{KS}}(\theta) = T_{\mathrm{KS},n}(\theta)$, one considers

$$T_{\mathrm{KS}}(\hat{\theta}) = \sup_{x \in \mathbb{R}} \left| F_{n,X}(x) - F_X(x; \hat{\theta}) \right|,$$

where $\hat{\theta}$ is a suitable estimate of θ. If the observations are i.i.d., then the rate of convergence for both, the original Kolmogorov–Smirnov statistics $T_{\mathrm{KS}} = T_{\mathrm{KS}}(\theta)$ and $T_{\mathrm{KS}}(\hat{\theta})$, is the same, though the variances of the limiting distributions are different.

To show what may happen in the long-memory case, let us consider a sequence $Y_t = X_t + \mu$ ($t \in \mathbb{N}$). Clearly, $F_Y(x) = F_X(x; \mu) = F_X(x - \mu)$. The empirical processes

$$E_{n,X}(x) = F_{n,X}(x) - F_X(x) = \frac{1}{n} \sum_{t=1}^{n} 1\{X_t \le x\} - F_X(x)$$

and

$$E_{n,Y}(x; \mu) := F_{n,Y}(x) - F_Y(x) = \frac{1}{n} \sum_{t=1}^{n} 1\{Y_t \le x\} - F_Y(x)$$

are related by

$$E_{n,Y}(x; \mu) = E_{n,X}(x - \mu). \tag{4.169}$$

On account of (4.160), $L_1^{-\frac{1}{2}}(n) n^{\frac{1}{2}-d} E_{n,Y}(x)$ converges weakly to $p_X(x - \mu) Z$. Now, consider instead

$$E_{n,Y}(x; \hat{\mu}) = F_{n,Y}(x) - F_X(x; \hat{\mu}).$$

We will use the estimate $\hat{\mu} = \bar{y}$, so that $\hat{\mu} - \mu = \bar{x}$. We then write

$$\begin{aligned} E_{n,Y}(x; \hat{\mu}) &= F_{n,Y}(x) - F_Y(x) + F_Y(x) - F_X(x; \hat{\mu}) \\ &= E_{n,X}(x - \mu) + F_X(x; \mu) - F_X(x; \hat{\mu}). \end{aligned}$$

Now, we apply Taylor's expansion to obtain

$$\begin{aligned} F_X(x; \mu) - F_X(x; \hat{\mu}) &= p_X(x - \mu)(\hat{\mu} - \mu) - \frac{1}{2} p_X^{(1)}(x - \mu)(\hat{\mu} - \mu)^2 + R_n \\ &= p_X(x - \mu)\bar{x} - \frac{1}{2} p_X^{(1)}(x - \mu)\bar{x}^2 + R_n, \end{aligned}$$

where R_n is of a smaller order than $\bar{x}^2$. Furthermore, the reduction principle (4.159) implies

$$n^{\frac{1}{2}-d}L_1^{-\frac{1}{2}}(n)\sup_{x\in\mathbb{R}}\left|E_{n,X}(x-\mu)+p_X(x-\mu)\bar{x}\right|\to_p 0.$$

Thus,

$$\begin{aligned}&n^{\frac{1}{2}-d}L_1^{-\frac{1}{2}}(n)E_{n,Y}(x;\hat{\mu})\\&\quad=o_P(1)-\frac{1}{2}n^{\frac{1}{2}-d}L_1^{-\frac{1}{2}}(n)\bigl(p_X^{(1)}(x-\mu)\bar{x}^2+R_n\bigr)=o_P(1),\end{aligned}$$

where the bound $o_P(1)$ is uniform in x given that $\sup_{x\in\mathbb{R}}|p_X^{(2)}(x)|<\infty$. In other words, the empirical processes $E_{n,Y}(\cdot\,;\mu)$ and $E_{n,Y}(\cdot\,;\hat{\mu})$ have different rates of convergence. Surprisingly, plugging in the parameter estimate improves the rate of convergence of the empirical process and therefore of goodness-of-fit tests such as the Kolmogorov–Smirnov or Anderson–Darling tests (Beran and Ghosh 1991; Ho 2002; Kulik 2009). The precise convergence rates are described in the following theorem.

Theorem 4.34 *Assume that the conditions of Theorem 4.33 are fulfilled. Additionally, assume that (4.158) holds with $r=3$.*

- *If $d\in(1/4,1/2)$ then*

$$n^{1-2d}L_1^{-1/2}(n)E_{n,Y}(x;\hat{\mu})\Rightarrow p_X^{(1)}(x-\mu)\left(Z_2-\frac{1}{2}Z_1^2\right),\tag{4.170}$$

  where Z_1 and Z_2 are uncorrelated random variables, $Z_1\sim N(0,1)$, and $Z_2=Z_{2,H}(1)$ is the Hermite–Rosenblatt variable.
- *If $d\in(0,1/4)$ then*

$$\sqrt{n}E_{n,Y}(x;\hat{\mu})\Rightarrow Z(x-\mu),\tag{4.171}$$

  where $Z(\cdot)$ is a Gaussian process.

Remark 4.6 The limiting Gaussian process has a rather complicated covariance structure. Nevertheless, the result (4.171) suggests that for $d\in(0,1/4)$, we can apply standard resampling techniques available for weakly dependent data, see Chap. 10.

To shed some light on the results of Theorem 4.34, consider the case $d\in(1/4,1/2)$. The expression for the limiting process follows essentially from the approximation

$$E_{n,Y}(x;\hat{\mu})\approx\bigl\{E_{n,X}(x-\mu)+p_X(x-\mu)\bar{x}\bigr\}+\frac{1}{2}p_X^{(1)}(x-\mu)\bar{x}^2.$$

Now, the result follows from (4.163) and the limiting behaviour of the sample mean.

Furthermore, the limiting behaviour may change if different estimators of the mean μ are considered or if one considers a location-scale family $Y = \mu + \sigma X$ (see Beran and Ghosh 1991; Ho 2002; Kulik 2009).

4.8.4 Linear Processes with Infinite Moments

As noticed above, finite-dimensional convergence of the appropriately scaled empirical process $E_{n,X} = F_{n,X} - F_X(x)$ follows from the result for partial sums of subordinated linear processes, by considering the function $y \to G(y;x) = 1\{y \le x\}$. We will apply the same idea to linear processes $X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ with i.i.d. symmetric infinite variance innovations, i.e.

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha} \tag{4.172}$$

with $\beta = 0$. The general result mimics Theorem 4.17. We established there that for $0 < d < 1 - 1/\alpha$, we have

$$n^{-H}\sum_{t=1}^{[nu]}\bigl\{G(X_t) - E\bigl[G(X_1)\bigr]\bigr\} \Rightarrow A^{1/\alpha}C_\alpha^{-1/\alpha}\frac{c_a}{d}G_\infty^{(1)}(0)\tilde{Z}_{H,\alpha}(u),$$

where $\tilde{Z}_{H,\alpha}(\cdot)$ is a linear fractional stable motion with $H = d + \alpha^{-1}$ and $G_\infty(y) = E[G(X+y)]$. Setting $u = 1$ and evaluating $G_\infty(y) = P(X \le x - y)$, $G_\infty^{(1)}(0) = -p_X(x)$, we may conclude that for a fixed $x \in \mathbb{R}$,

$$n^{-H}\sum_{t=1}^{n}\bigl(1\{X_t \le x\} - P(X_1 \le x)\bigr) \overset{\mathrm{d}}{\to} A^{1/\alpha}C_\alpha^{-1/\alpha}\frac{c_a}{d}p_X(x)\tilde{Z}_{H,\alpha}(1).$$

This can be extended to convergence of the process $E_{n,X}(x)$ ($x \in \mathbb{R}$), see Koul and Surgailis (2001).

Theorem 4.35 *Assume that X_t ($t \in \mathbb{Z}$) is a linear process with $a_j \sim c_a j^{d-1}$,*

$$0 < d < 1 - 1/\alpha,$$

and ε_t ($t \in \mathbb{Z}$) are i.i.d. symmetric random variables such that (4.89) holds with $\alpha \in (1,2)$ and $\beta = 0$:

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}.$$

Furthermore, assume that the distribution F_ε of ε_1 is such that

$$\bigl|F_\varepsilon^{(2)}(x)\bigr| \le C\bigl(1+|x|\bigr)^{-\alpha}, \qquad \bigl|F_\varepsilon^{(2)}(x) - F_\varepsilon^{(2)}(y)\bigr| \le C|x-y|\bigl(1+|x|\bigr)^{-\alpha},$$

where $|x-y|<1$, $x\in\mathbb{R}$. *Then*

$$n^{1-H}E_{n,X}(x)\Rightarrow A^{1/\alpha}C_{\alpha}^{-1/\alpha}\frac{c_a}{d}p_X(x)\tilde{Z}_{H,\alpha}(1), \tag{4.173}$$

where $\tilde{Z}_{H,\alpha}(1)$ *is a symmetric* α*-stable random variable with scale* η *given by*

$$\eta=\left(\int_{-\infty}^{1}\left\{(1-v)_+^d-(-v)_+^d\right\}^{\alpha}dv\right)^{1/\alpha}.$$

4.8.5 Tail Empirical Processes

Let X_t $(t\in\mathbb{Z})$ be a stationary sequence with marginal distribution F_X. More specifically, we shall assume that X_t is a stochastic volatility model considered in Sect. 4.3.4. Recall that the model is $X_t=\xi_t\sigma_t$ $(t\in\mathbb{Z})$, where

$$\sigma_t=\sigma(\zeta_t),\quad \zeta_t=\sum_{j=1}^{\infty}a_j\varepsilon_{t-j},$$

and $\sigma(\cdot)$ is a positive function. It is assumed that ξ_t $(t\in\mathbb{Z})$ is a sequence of i.i.d. random variables such that

$$P(\xi_1>x)\sim A\frac{1+\beta}{2}x^{-\alpha},\qquad P(\xi_1<-x)\sim A\frac{1-\beta}{2}x^{-\alpha}. \tag{4.174}$$

Also, we assume that the sequences ξ_t $(t\in\mathbb{Z})$ and ε_t $(t\in\mathbb{Z})$ are mutually independent. In particular (cf. Lemma 4.20), we have

$$P\big(|X_1|>x\big)\sim E\big(\sigma^{\alpha}(\zeta_1)\big)P\big(|\xi_1|>x\big),$$

provided that

$$E\big[\sigma^{\alpha+\delta}(\zeta_1)\big]<\infty \tag{4.175}$$

for some $\delta>0$. In Theorem 4.19 we saw that the limiting behaviour of partial sums depends on an interplay between the long-memory parameter d and the tail index α. Therefore, it is important to have reliable estimates of both parameters, d and α. With the help of the tail empirical process it is possible to prove asymptotic normality of the so-called Hill estimator of α.

We note first that the tail behaviour of X implies that, as $n\to\infty$,

$$T_n(x):=P\big(X_1>(1+x)u_n|X_1>u_n\big)=\frac{\bar{F}_X((1+x)u_n)}{\bar{F}_X(u_n)}\to T(x):=(1+x)^{-\alpha}$$

for any sequence of constants $u_n \to \infty$. The tail empirical distribution functions $\tilde{T}_n(s)$ and the tail empirical processes $e_n(s)$ are defined by

$$\tilde{T}_n(s) = \frac{1}{n\bar{F}_X(u_n)} \sum_{t=1}^{n} 1\{X_t > u_n(1+s)\}$$

and

$$e_n(s) = \tilde{T}_n(s) - T_n(s) \ \bigl(s \in [0,\infty)\bigr). \tag{4.176}$$

We note that for large values of u_n, only extreme observations are included in the sum. Hence the name “tail empirical”.

Drees (1998, 2000) and Rootzén (2009) show that for weakly dependent observations X_t, scaled processes $w_n e_n$ converge weakly in $D[0,\infty)$ to a Gaussian process $w = B \circ T$, where B is a standard Brownian motion, and $w_n^2 = n\bar{F}_X(u_n)$. The situation changes in the long-memory case. The limiting behaviour depends on an interplay between the memory parameter d and the behaviour of u_n. If u_n grows sufficiently fast (that means that very few extremes are included in the tail empirical distribution), then long memory does not influence the limit: $w_n e_n \Rightarrow w$ with, as before, $w_n = \sqrt{n\bar{F}_X(u_n)}$ and $w = B \circ T$. However, if u_n grows at an appropriately slow rate, then long memory starts to play a role: $w_n e_n$ converge weakly to a degenerate limiting process $w(s) = CT(s)Z_{m,H}(1)$ (where C is a constant), and the scaling factor is different, namely $w_n = n^{m(\frac{1}{2}-d)}L(n)$, where L is a slowly varying function. The corresponding result is stated in Theorem 4.36.

In order to state the result, let us define the function G_n on $(-\infty,\infty) \times [0,\infty)$ by

$$G_n(x,s) = \frac{P(\sigma(x)\xi_1 > (1+s)u_n)}{P(\xi_1 > u_n)}. \tag{4.177}$$

This function converges pointwise to $T(s)G(x) = T(s)\sigma^{\alpha}(x)$. Furthermore, the Hermite coefficients $J_n(m,s)$ of the function $x \to G_n(x,s)$ converge (as $n \to \infty$) to $J(m)T(s)$, uniformly with respect to $s \ge 0$, where $J(m)$ is the m-th Hermite coefficient of G. This implies that for large n, the Hermite rank $m_n(s)$ of $G_n(\cdot,s)$ is not greater than the Hermite rank m of G. To avoid further complications, we impose the assumption $\inf_{s\ge 0} m_n(s) = m$ for sufficiently large n.

Theorem 4.36 *Consider the stochastic volatility model $X_t = \xi_t\sigma_t$ ($t \in \mathbb{Z}$) and assume that* (4.174) *and* (4.175) *hold. Additionally, we assume that ζ_j ($t \in \mathbb{Z}$) is a Gaussian linear process with coefficients a_j satisfying* (B1), *i.e. $a_j = L_a(j)j^{d-1}$, $d \in (0,1/2)$ (so that $\gamma_X(k) \sim L_\gamma(k)k^{2d-1}$). Let $m \ge 1$ be the Hermite rank of the function $\sigma^{\alpha}(\cdot)$, and set $H = d + 1/2$. Assume that $E[\sigma^{2\alpha+\delta}(X_1)] < \infty$.*

(i) *If $n\bar{F}_X(u_n) \to \infty$ and $n^{1-m(1-2d)}L_m(n)\bar{F}_X(u_n) \to 0$ as $n \to \infty$, then $\sqrt{n\bar{F}_X(u_n)}\,e_n$ converges weakly in $D[0,\infty)$ to the Gaussian process $B \circ T$, where B is a standard Brownian motion.*

(ii) *If* $n\bar{F}_X(u_n) \to \infty$ *and* $n^{1-m(1-2d)}L_m(n)\bar{F}_X(u_n) \to \infty$ *as* $n \to \infty$, *then*

$$n^{m(\frac{1}{2}-d)}L_m^{-1/2}(n)e_n(s) \Rightarrow \frac{J(m)T(s)}{E[\sigma^\alpha(\zeta_1)]}Z_{m,H}(1),$$

where $\Rightarrow$ *denotes weak convergence in* $D[0,\infty)$, $Z_{m,H}(\cdot)$ *is a Hermite–Rosenblatt process, and* $L_m(n) = m!C_m L_\gamma^m(n)$.

The practical application of these limit theorems for $e_n(\cdot)$ is not quite straightforward. First of all, $\bar{F}_X(u_n)$ is unknown. The second problem is that we would like to center the tail empirical distribution function by $T(s)$, not $T_n(s)$. The second question can be addressed by introducing the assumption

$$\lim_{n\to\infty} w_n \|T_n - T\|_\infty = 0, \tag{4.178}$$

where

$$\|T_n - T\|_\infty = \sup_{t\geq 1}\left|\frac{P(X_1 > u_n t)}{P(X_1 > u_n)} - t^{-\alpha}\right|,$$

and the scaling w_n is either $\sqrt{n\bar{F}_X(u_n)}$ or $n^{m(\frac{1}{2}-d)}L_m^{-1/2}(n)$ in cases (i) and (ii) respectively. In other words, we impose a condition that makes the bias $T_n - T$ negligible. This is related to the so-called second-order regular variation (see Drees 1998; Kulik and Soulier 2011), but we omit details here. As an example, assume for instance that

$$P(\xi_1 > x) = cx^{-\alpha}\big(1 + O\big(x^{-\beta}\big)\big) \quad (x \to \infty)$$

for some constant $c > 0$. Then the second-order regular variation refers to the second-order term $x^{-\beta}$ in the expansion for the tail of ξ_1.

Now, suppose that the second-order assumption holds. Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be the order statistics of $X_1, \ldots, X_n$, define $k_n = n\bar{F}_X(u_n)$ and replace u_n by $X_{n-k:n}$ in the definition of the tail empirical distribution function. Implicitly, $k = k_n$ will become a user chosen number of extreme statistics such that $k_n \to \infty$ and $k_n = o(n)$. Thus, we define

$$\hat{T}_n(s) = \frac{1}{k}\sum_{t=1}^{n} 1\big\{X_t > X_{n-k:n}\cdot(1+s)\big\}$$

and the practically computable processes

$$\hat{e}_n^*(s) = \hat{T}_n(s) - T(s) \qquad \big(s \in [0,\infty)\big).$$

It follows from Rootzén (2009) and Kulik and Soulier (2011) that

$$w_n\,\hat{e}_n^*(s) \Rightarrow w^*(s) = w(s) - T(s)w(1).$$

In particular, if $w_n = \sqrt{n\bar{F}_X(u_n)} = \sqrt{k_n}$ and $w(s) = B(T(s))$, then $w^*(s) = \tilde{B}(T(s))$, where $\tilde{B}$ is a Brownian bridge. However, if $w_n = n^{m(\frac{1}{2}-d)}L(n)$ and

$w(s) = CT(s)Z_{m,H}(1)$, then $w^*(s) = 0$. This is a similar effect as for the standard empirical process with estimated parameters considered in Sect. 4.34. More surprisingly, we have the following result for the process $\hat{e}_n^*(s)$.

Theorem 4.37 *Assume that the conditions of Theorem 4.36 are fulfilled. Assume additionally that (4.178) holds. Then $\sqrt{k}\,\hat{e}_n^*(s)$ converges weakly in $D[0,\infty)$ to the Gaussian process $\tilde{B}(T(s))$, where $\tilde{B}$ is a standard Brownian bridge, regardless of the behaviour of $n^{1-m(1-2d)}L_m(n)\bar{F}_X(u_n)$.*

4.8.5.1 Application to Tail Index Estimation

One of the most important problems when dealing with heavy tails is to estimate the tail index α. The best known (though in many ways not always reliable) method is Hill's estimator. Using the notation $\gamma = \alpha^{-1}$, the Hill estimator of γ is defined by

$$\hat{\gamma}_n = \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{n-j+1:n}}{X_{n-k:n}}\right).$$

Noting that

$$\int_0^\infty \frac{\hat{T}_n(s)}{1+s}\,ds = \frac{1}{k}\sum_{t=1}^{n}\int_0^\infty \frac{1\{s < X_t/X_{n-k:n} - 1\}}{1+s}\,ds$$
$$= \frac{1}{k}\sum_{t=1}^{n}\log\left(1+\max\left\{\frac{X_t}{X_{n-k:n}} - 1, 0\right\}\right),$$

the estimator can also be written as

$$\hat{\gamma}_n = \int_0^\infty \frac{\hat{T}_n(s)}{1+s}\,ds.$$

Since $\gamma = \int_0^\infty (1+s)^{-1}T(s)\,ds$, we have

$$\hat{\gamma}_n - \gamma = \int_0^\infty \frac{\hat{e}_n^*(s)}{1+s}\,ds.$$

Thus we can apply Theorem 4.37 to obtain the asymptotic distribution of the Hill estimator. Heuristically,

$$\sqrt{k_n}(\hat{\gamma}_n - \gamma) \to_d \int_0^\infty \frac{\tilde{B}(T(s))}{1+s}\,ds.$$

This integral is a normal random variable with variance γ^2 (for details, see Kulik and Soulier 2011). In summary, we have the following result.

Corollary 4.5 *Under the assumptions of Theorem 4.37, $\sqrt{k}(\hat{\gamma}_n - \gamma)$ converges in distribution to a centred Gaussian distribution with variance γ^2.*

This result can be used to construct confidence intervals for γ. It is known that this result gives the best possible rate of convergence for the Hill estimator for i.i.d. data (see Drees 1998). The surprising result is that it is possible to achieve the same i.i.d. rates regardless of the dependence parameter d.

4.8.5.2 Proof of Theorem 4.36

Proof We follow a similar idea as in the proof of Theorem 4.19. Let $\mathcal{E}$ be the σ-field generated by the Gaussian process ζ_t ($t \in \mathbb{Z}$). Write

$$\begin{aligned} e_n(s) &= \frac{1}{n\bar{F}_X(u_n)} \sum_{t=1}^{n} \left\{ 1\left\{ X_t > (1+s)u_n \right\} - P\left(X_t > (1+s)u_n | \mathcal{E} \right) \right\} \\ &\quad + \frac{1}{n\bar{F}_X(u_n)} \sum_{t=1}^{n} \left\{ P\left(X_t > (1+s)u_n | \mathcal{E} \right) - \bar{F}_X(u_n) \right\} \\ &=: M_n(s) + R_n(s). \end{aligned} \tag{4.179}$$

The difference between (4.179) and the decomposition used in the proof of Theorem 4.19 is that here the first part is the sum of conditionally independent random variables, instead of being a martingale. The second part is a function of the Gaussian sequence ζ_t ($t \in \mathbb{N}$) and does not depend on the sequence ξ_t ($t \in \mathbb{N}$).

For the first part, it can be shown that, using the conditional independence,

$$\log E\left[\exp\left(it\sqrt{n\bar{F}_X(u_n)} M_n(0) \right) \middle| \mathcal{E} \right] \to_P -t^2/2.$$

The bounded convergence theorem implies

$$\sqrt{n\bar{F}_X(u_n)} M_n(0) \to_d T(0)Z,$$

where Z is standard normal. Using the Cramer–Wald device, it is extended to

$$\begin{aligned} &\sqrt{n\bar{F}_X(u_n)} \left(M_n(s_1), M_n(s_l) - R_n(s_{l-1}), l = 2, \ldots, K \right) \\ &\quad \to_d \left(N\left(0, T(s_1)\right), N\left(0, T(s_l) - T(s_{l-1})\right), l = 2, \ldots, K \right), \end{aligned} \tag{4.180}$$

where the normal random variables are independent. Computations are somewhat involved, but the idea is relatively easy. Since the random variables are conditionally independent, the characteristic function can be evaluated.

Recall that

$$G_n(x,s) = \frac{P(\sigma(x)\xi_1 > (1+s)u_n)}{P(\xi_1 > u_n)}$$

converges pointwise to $T(s)G(x) = T(s)\sigma^\alpha(x)$. Let us now write

$$\sum_{t=1}^{n}\big(G_n(\zeta_t,s) - E\big[G_n(\zeta_t,s)\big]\big)$$
$$= \sum_{t=1}^{n}\sum_{q=m}^{\infty}\frac{T(s)J(q)}{q!}H_q(\zeta_t) + \sum_{t=1}^{n}\sum_{q=m}^{\infty}\frac{J_n(q,s) - T(s)J(q)}{q!}H_q(\zeta_t)$$
$$=: T(s)R_n^* + \tilde{R}_n(s)$$

with $R_n^* = \sum_{t=1}^{n} G(\zeta_t)$. Convergence of $T(s)R_n^*$ is concluded in the very same way as in (4.102) and (4.103). For $m(1/2-d) < 1$ and $m(1/2-d) > 1$, we have, respectively,

$$n^{-(1-m(\frac{1}{2}-d))}L_m^{-1/2}(n)R_n^* \Rightarrow \frac{J(m)}{m!}Z_{m,H}(1)$$

and

$$n^{-1/2}R_n^* \Rightarrow vZ,$$

where v is a constant. The second part, $\tilde{R}_n(s)$ is of a smaller order than R_n^*, uniformly in $s \ge 0$. Since

$$R_n(s) = \frac{P(\xi_1 > u_n)}{n\bar{F}_X(u_n)}\sum_{t=1}^{n}\big(G_n(\zeta_t,s) - E\big[G_n(\zeta_t,s)\big]\big), \tag{4.181}$$

and $P(\xi_t > u_n)/\bar{F}_X(u_n) \to 1/E[\sigma^\alpha(\zeta_1)]$, we conclude that for $m(1/2-d) < 1$,

$$n^{m(\frac{1}{2}-d)}L_m^{-1/2}(n)R_n(s) \to_d \frac{J(m)T(s)}{E[\sigma^\alpha(\zeta_1)]}Z_{m,H}(1). \tag{4.182}$$

This convergence is easily extended to multivariate convergence. If $m(1/2-d) > 1$, then $R_n(s)$ is uniformly negligible w.r.t. the conditionally independent part $M_n(s)$. Therefore, (4.182) and (4.180) yield the finite-dimensional convergence. For details and proof of tightness, we refer to Kulik and Soulier (2011). □

4.8.5.3 Further Extensions

The results given above are extendable to stochastic volatility models with leverage. Instead of decomposing $e_n(s)$ into a conditionally i.i.d. part $M_n(s)$ and a long-memory part $R_n(s)$, we may apply the martingale decomposition as in the proof of Theorem 4.19. For details, see Luo (2011).

4.9 Limit Theorems for Counting Processes and Traffic Models

In this section we review limit theorems for counting processes and traffic models, such as renewal reward, ON–OFF, shot-noise and infinite source Poisson processes, considered in Sect. 2.2.4.

4.9.1 Counting Processes

Let X_j $(j \geq 1)$ be a stationary sequence of strictly positive random variables with distribution F and finite mean. Let τ_0 have the distribution $F^{(0)}$ and define

$$\tau_j = \tau_0 + \sum_{k=1}^{j} X_k \quad (j \geq 1)$$

and

$$S_n(t) = \sum_{j=1}^{[nt]} X_j.$$

Note that the notation X_j and $S_n(t)$ is different from what was used previously (which was $S(u) = \sum_{t=1}^{[nu]} X_t$). The reason is that here the natural time parameter is in the upper limit $[nt]$ of the sum.

Now, let $N(t)$ be the associated counting process. Since

$$N(t) = \max\{k \geq 0 : \tau_{k-1} \leq t\} = \min\{k \geq 0 : \tau_k > t\},$$

one can view $N(t)$ as the generalized inverse of the partial sums process $S_n(t)$. Consequently, if the limiting process for partial sums is Gaussian, Lemma 4.7 will imply the weak convergence of $N(t)$ from that of $S_n(t)$. In other words, we apply Lemma 4.7 to

- $y_n(t) = S_n(t)/(n\mu)$,
- $y_n^{-1}(t) = N_n(t)/n$, where $N_n(t) = N(n\mu t)$.

If $c_n^{-1}(S_n(t)/(n\mu) - t)$ converges to a process $S(t)$ with some constants c_n, then $c_n^{-1}(N(n\mu t)/n - t)$ converges to $-S(t)$. The same procedure applies to any stationary counting process associated with a stationary sequence X_j $(j \in \mathbb{N})$ with finite mean.

Example 4.24 Recall Theorem 4.5. There, X_j $(j \in \mathbb{N})$ is a linear process $X_j = \sum_{k=0}^{\infty} a_k \varepsilon_{j-k}$ with summable coefficients a_k and i.i.d. centred innovations ε_j $(j \in \mathbb{Z})$. We can reformulate Theorem 4.5 to accommodate $\mu = E(X_1) \neq 0$. We have

$$n^{-1/2} \sum_{j=1}^{[nt]} (X_j - \mu) \Rightarrow v B(t)$$

in $D[0,1]$, where $v^2 = \sigma_X^2 + 2\sum_{k=1}^{\infty}\gamma_X(k)$, and $B(t)$ ($t \in [0,1]$) is a standard Brownian motion. Equivalently,

$$\frac{S_n(t)/(n\mu) - t}{n^{-1/2}} \Rightarrow v\mu^{-1}B(t),$$

so that $S(t) = v\mu^{-1}B(t)$ and $c_n = n^{-1/2}$. Application of Lemma 4.7 yields

$$n^{-1/2}\bigl(N(n\mu t) - nt\bigr) \Rightarrow v\mu^{-1}B(t).$$

However, we cannot extend this to the situation of Theorem 4.6. The long-range dependent linear process must have zero mean and hence cannot be strictly positive.

Example 4.25 Recall Example 4.12. The model considered there is $X_j = \xi_j\sigma(\zeta_j)$, where ξ_j ($j \ge 1$) are strictly positive random variables with mean $E(\xi_1)$, and ζ_j is a centred Gaussian sequence with covariance $\gamma_\zeta(k) \sim L_\gamma(k)k^{2d-1}$, $d \in (0, 1/2)$. We established in Example 4.12 that for $G(x) = x$ and $\sigma(x) = \exp(x)$, we have

$$n^{-(d+1/2)}L_1^{-1/2}(n)\sum_{j=1}^{[nt]}\bigl(X_j - E(X_1)\bigr) \Rightarrow J(1)B_H(t)$$

weakly in $D[0,1]$, where $B_H(\cdot)$ is fractional Brownian motion with $H = d + 1/2$ and $J(1) = E(\zeta_1\exp(\zeta_1))E(\xi_1)$. Hence, for the inverse processes, we obtain

$$n^{-H}L_1^{-1/2}(n)\bigl(N(n\mu t) - nt\bigr) \Rightarrow J(1)\mu^{-1}B_H(t).$$

Thus, long memory in the interpoint distances generates long-memory-type behaviour in the functional central limit theorem for the counting process.

Let now X_j ($t \in \mathbb{N}$) be an i.i.d. sequence of strictly positive random variables such that

$$P(X_1 > x) \sim Ax^{-\alpha} \quad (A > 0, \alpha > 1).$$

In Sect. 4.3 we saw that the appropriately centred and normalized $S_n(t)$ converges to an α-stable Lévy process with independent increments (cf. (4.80)):

$$c_n^{-1}\sum_{j=1}^{[nt]}(X_j - \mu) \Rightarrow C_\alpha^{-1/\alpha}Z_\alpha(t),$$

where $c_n = \inf\{s : P(X > x) \le n^{-1}\}$, $c_n \sim A^{1/\alpha}n^{1/\alpha}$, and $Z_\alpha(t)$ is an α-stable Lévy motion such that $Z_\alpha(1) \overset{d}{=} S_\alpha(1,1,0)$. The limiting process has discontinuous sample paths, and hence Lemma 4.7 is not applicable. However (see Theorem 7.3.2 in Whitt 2002), one can generalize Vervaat's result to cover the case of limiting processes with discontinuous sample paths. One has to mention though that although $S_n(t)$ may converge in the standard Skorokhod topology, the same does not apply

Table 4.5 Limits for counting processes—tails vs. dependence

Counting processes	Weak dependence	Strong dependence
Interarrival times with finite variance	Brownian motion (Example 4.24)	fBm (Example 4.25)
Interarrival times with infinite variance	Lévy process (Example 4.26)	fBm or Lévy process (Example 4.27)

to the counting process. One has to consider a weaker M_1 topology (see comments on p. 235 as well as Sects. 13.6 and 13.7 in Whitt 2002). Here, we just illustrate finite-dimensional convergence.

Example 4.26 In the situation described above,

$$c_n^{-1}\big(N(n\mu t) - nt\big) \overset{\text{fidi}}{\to} -C_\alpha^{-1/\alpha}\mu^{-1}Z_\alpha(t). \tag{4.183}$$

Thus, a heavy-tailed distribution of interarrival times X_j generates Long-Range count Dependence (LRcD) in the counting process (see Example 2.5). On the other hand, the limiting process has independent increments. Furthermore, in Example 2.5 we found out that $\text{var}(N(t))$ is proportional to t^{2H} (as $t \to \infty$) with $H = (3-\alpha)/2$. On the other hand, $n^{-H}(N(n\mu t) - nt)$ converges to 0 in probability. Hence, $N(\cdot)$ is an example of a second-order stationary process where its standard deviation does not yield an appropriate scaling.

Example 4.27 Recall Example 4.17. If $d + 1/2 < 1/\alpha$, then by Whitt's approach

$$n^{-1/\alpha}\big(N(n\mu t) - nt\big) \overset{\text{fidi}}{\to} -A^{1/\alpha}C_\alpha^{-1/\alpha}\big\{E\big(\sigma_1^\alpha\big)\big\}^{1/\alpha}\mu^{-1}Z_\alpha(t). \tag{4.184}$$

If however $d + 1/2 > 1/\alpha$, we can use Vervaat's Lemma 4.7 to conclude

$$n^{-(d+1/2)}L_1^{-1/2}(n)\big(N(n\mu t) - nt\big) \Rightarrow J(1)E(\xi_1)\mu^{-1}B_H(t). \tag{4.185}$$

We summarize our findings in Table 4.5. It should be noted that in the case of strong dependence the results are just for the case in Examples 4.25, 4.27, not for all long-memory models.

4.9.2 *Superposition of Counting Processes*

Let $N^{(m)}(t)$ ($t \geq 0$, $m = 1, \ldots, M$) be independent copies of a stationary renewal process $N(t)$ associated with a renewal sequence X_j ($j \in \mathbb{N}$). We assume that, as $x \to \infty$,

$$\bar{F}(x) = P(X_1 > x) \sim x^{-\alpha}L(x) \quad (1 < \alpha < 2),$$

and that $P(\tilde{X}_0 > x) = \mu^{-1} \int_x^\infty \bar{F}(u)\,du$, where $\mu = E[X_1] = \lambda^{-1}$. Application of Lemma 4.6 yields

$$\lim_{M\to\infty} \frac{1}{M^{1/2}} \sum_{m=1}^{M} \big(N^{(m)}(t) - \lambda t\big) \Rightarrow G(t), \tag{4.186}$$

where $G(\cdot)$ is a Gaussian process with stationary increments and the same covariance structure as $N(t)$. In particular (see Example 2.5),

$$\operatorname{var}\big(G(t)\big) = \operatorname{var}\big(N(t)\big) \sim \frac{2\lambda}{(\alpha-1)(2-\alpha)(3-\alpha)} t^{3-\alpha} L(t) =: \sigma_0^2 t^{3-\alpha} L(t).$$

Indeed, to apply Lemma 4.6, we verify that for $t > s$,

$$\operatorname{var}\big(N(t) - N(s)\big) = \operatorname{var}\big(N(t-s)\big) \sim C(t-s)^{2H}$$

and $2H > 1$. Also, the second condition of Lemma 4.6 is easily verified.

We recognize that the limiting process has up to a constant the same variance as a fractional Brownian motion with the Hurst index $H = (3-\alpha)/2$. Now, let us consider the time scaled process $N^{(m)}(Tt)$. For a fixed $T > 0$, application of (4.186) yields

$$\lim_{M\to\infty} \frac{1}{M^{1/2}} \sum_{m=1}^{M} \big(N^{(m)}(Tt) - \lambda Tt\big) \Rightarrow G(Tt) = \sigma_0 B_H(Tt)$$

and $\operatorname{var}(G(Tt)) \sim \sigma_0^2 T^{2H} t^{2H} L(Tt) \sim \sigma_0^2 T^{2H} t^{2H} L(T)$ as $T \to \infty$. Thus, applying H-self-similarity of fractional Brownian motion, we have

$$\lim_{T\to\infty} \frac{1}{T^H} \lim_{M\to\infty} \frac{1}{M^{1/2}} \sum_{m=1}^{M} \big(N^{(m)}(Tt) - \lambda Tt\big) \Rightarrow \sigma_0 B_H(t).$$

On the other hand, (4.183) yields

$$\lim_{T\to\infty} a_T^{-1} \big(N^{(m)}(Tt) - \lambda Tt\big) \overset{\text{fidi}}{\to} -\mu^{-1} C_\alpha^{-1/\alpha} Z_\alpha^{(m)}(\lambda t) \quad (m = 1, \ldots, M),$$

where $Z^{(m)}(\cdot)$ $(m = 1, \ldots, M)$ are independent Lévy processes, and $a_T \sim T^{1/\alpha} \ell(T)$. Consequently, since the sum of independent Lévy processes yields a Lévy process, we obtain

$$\lim_{M\to\infty} \frac{1}{M^{1/\alpha}} \lim_{T\to\infty} a_T^{-1} \sum_{m=1}^{M} \big(N^{(m)}(Tt) - \lambda Tt\big) \overset{\text{fidi}}{\to} -\lambda^{1+1/\alpha} C_\alpha^{-1/\alpha} Z_\alpha(t),$$

where $Z_\alpha(\cdot)$ is an α-stable Lévy process. The limiting constants were obtained by replacing t with λt and using $Z_\alpha(\lambda t) \overset{\text{d}}{=} \lambda^{1/\alpha} Z_\alpha(t)$.

Table 4.6 Limits for superposition of counting processes—tails vs. dependence

Superposition of counting processes	Weak dependence	Strong dependence
Interarrival times with finite variance	$\lim_{M\to\infty}\lim_{T\to\infty} = \text{Bm}$ $\lim_{T\to\infty}\lim_{M\to\infty} = \text{Bm}$	$\lim_{M\to\infty}\lim_{T\to\infty} = \text{fBm}$ $\lim_{T\to\infty}\lim_{M\to\infty} = \text{fBm}$
Interarrival times with infinite variance	$\lim_{M\to\infty}\lim_{T\to\infty} = \text{Lévy}$ $\lim_{T\to\infty}\lim_{M\to\infty} = \text{fBm}$	

We observe that different limiting schemes yield different limiting processes. This feature will be also present in different traffic models.

In contrast, if the renewal sequence has a finite variance and short memory, then application of Example 4.24 yields that both procedures $\lim_{M\to\infty}\lim_{T\to\infty}$ and $\lim_{T\to\infty}\lim_{M\to\infty}$ produce the same limit, namely a Brownian motion. Likewise, in the case of strong dependence and a finite variance (as in Example 4.25), both procedures yield a fractional Brownian motion.

We summarize these observations in Table 4.6. We do not fill in the case of strong dependence and heavy tails (situation of Example 4.27). It is clear that there are four possible limits. If the counting process converges to fBm, then the limit for superpositions must be fBm as well. If the counting process converges to a Lévy process, then the superposition converges to either fBm or a Lévy process, depending on the order of taking these limits.

4.9.3 Traffic Models

Let $W(u)$ be a traffic model. It can be either a renewal reward, or ON–OFF, or infinite source Poisson or error duration process. In Sect. 2.2.4 we noted that the models have long memory in terms of non-integrable covariances or nonlinear growth of the variance of the integrated process. A very interesting feature is that long memory in a traffic process implies that the integrated process

$$W^*(t) = \int_0^t \left\{W(v) - E\left[W(v)\right]\right\} dv$$

converges in the sense of finite-dimensional distributions to an α-stable Lévy motion. The scaling factor has to be chosen as $T^{-1/\alpha}L(T)$, where L is a slowly varying function. In particular, this is another example of a second-order long-memory process where the variance grows at rate T^{2H}, but $T^{-H}W^*(Tt)$ converges to zero in probability as $T \to \infty$ (see e.g. Example 4.26). Furthermore, as in the case of counting processes, the convergence cannot hold in the $D[0,1]$ space equipped with the J_1-topology.With respect to J_1 the continuous process $W^*(Tt)$ must converge to a continuous limit, which is not the case here.

In the context of computer networks, these phenomena describe long memory of an individual source. However, they do not explain long memory at the level of teletraffic, which usually consists of a large number of sources. Assume now that we have M independent copies $W^{(m)}(\cdot)$ $(m = 1, \ldots, M)$ of the traffic process $W(t)$. Define

$$W^*_{T,M}(t) = \int_0^{Tt} \sum_{m=1}^{M} \{W^{(m)}(v) - E[W(v)]\} \, dv = \sum_{m=1}^{M} W^{(m)*}(Tt),$$

where $W^{(m)*}(u)$, $m = 1, \ldots, M$, are i.i.d. copies of the cumulated process $W^{(m)}(t)$. The process $W^*_{T,M}(t)$ can be interpreted as (centred) total workload of M workstations at time t or as cumulative packet counts in the network by time t. We are interested in the limiting behaviour of the properly normalized cumulative process $W^*_{T,M}(t)$.

We will consider two limiting scenarios. First, we will analyse what happens if we let first $M \to \infty$ and then $T \to \infty$. In this setup, we will proceed as follows.

Step 1: Use Lemma 4.6 to establish that with some sequence a_M,

$$\lim_{M\to\infty} a_M^{-1} \sum_{m=1}^{M} \{W^{(m)}(t) - E[W^{(m)}(t)]\}$$

converges to a process, say, $G(t)$. If the process is Gaussian, then its covariance structure is the same as that of $W(u)$.

Step 2: If the process $G(t)$ is Gaussian, then the integral $G^*(Tt) = \int_0^{Tt} G(u)\,du$ is Gaussian as well. We have

$$\operatorname{var}(G^*(Tt)) = \int_0^{Tt} \left(\int_0^{v} cov(W(0), W(s)) \, ds \right) dv.$$

From the form of the covariance function we will conclude either a Brownian motion or a fractional Brownian motion as limit.

Step 3: The sum of independent (fractional) Brownian motions yields (fractional) Brownian motion. We will conclude that

$$\lim_{t\to\infty} a_T^{-1} \lim_{M\to\infty} a_M^{-1} \int_0^{Tt} \sum_{m=1}^{M} (W^{(m)}(v) - E[W^{(m)}(v)]) \, dv$$

converges to a (fractional) Brownian motion, where a_T is proportional to $T^{1/2}$ or T^H $(H > 1/2)$, respectively.

As for the case $T \to \infty$ and then $M \to \infty$, we will proceed as follows.

Step 1: For each $m = 1, \ldots, M$, approximate

$$\lim_{T\to\infty} c_T^{-1} \int_0^{Tt} \{W^{(m)}(v) - E[W^{(m)}(v)]\} \, dv \approx c_T^{-1} \sum_{j=1}^{N(Tt)} U_j \quad (T \to \infty),$$

where $N(\cdot)$ is an appropriate counting process, and U_j $(j \in \mathbb{N})$ is an appropriate i.i.d. sequence. Note that both N and U_j depend on m. If the random variables U_j have a finite variance, then for each m, the limiting process is a Brownian motion, and $c_T = T^{1/2}$. If the random variables U_j are regularly varying with index α, then we obtain a Lévy process as a limit and $c_T = T^{1/\alpha}$.

Step 2: The sum of independent Brownian motions (Lévy processes) is a Brownian motion (Lévy process). We conclude the convergence for

$$\lim_{M\to\infty} d_M^{-1} \lim_{T\to\infty} c_T^{-1} \int_0^{Tt} \sum_{m=1}^{M} \left(W^{(m)}(v) - E\left[W^{(m)}(v)\right]\right) dv$$

with some sequence d_M.

One has to mention though that the proofs are sketched, without verifying some technical details.

4.9.4 Renewal Reward Processes

Recall from Example 2.12 the renewal reward process

$$W(t) = Y_0 1\{0 < t < \tau_0\} + \sum_{j=1}^{\infty} Y_j 1\{\tau_{j-1} \le t < \tau_j\},$$

$X_j = \tau_j - \tau_{j-1}$. We assume for simplicity that Y_j $(j \in \mathbb{N})$ is a centred i.i.d. sequence, independent of the renewal sequence τ_0, X_j $(j \ge 1)$, and also that $E[X_1] = \mu = \lambda^{-1}$ is finite. We are interested in the limiting behaviour of the cumulative process $W_{T,M}^*(t)$ defined above. For the purpose of the limiting regime $\lim_{M\to\infty} \lim_{T\to\infty}$, we represent the cumulative process as follows:

$$\int_0^{Tt} W(u)\, du = \min\{Tt, \tau_0\} Y_0 + \sum_{j=0}^{\infty} Y_{j+1} \left(\min\{Tt, \tau_{j+1}\} - \tau_j\right)_+. \tag{4.187}$$

Indeed, if $Tt < \tau_0$, then $\int_0^{Tt} W(u)\, du = Y_0 Tt$; if $\tau_0 < Tt < \tau_1$, then $\int_0^{Tt} W(u)\, du = Y_0 Tt + Y_1(Tt - \tau_0)$ etc.

An alternative representation will yield an approximation of the cumulative reward by a sum of i.i.d. random variables. For $Tt > \tau_0$, we may write

$$\int_0^{Tt} W(u)\, du = Y_0 \tau_0 + \sum_{j=1}^{N(Tt)} Y_j X_j - U, \tag{4.188}$$

where $N(t)$ is the renewal process associated with τ_j. The first two terms represent the renewal intervals that are at least partially included in $[0, Tt]$. For example, if

$\tau_0 < Tt < \tau_1$, then $N(Tt) = 1$, and the sum includes $Y_0\tau_0 + Y_1X_1$. However, not the entire renewal interval X_1 is included in $[0, Tt]$. We have to subtract a portion $(\tau_1 - Tt)Y_1$, and this is "hidden" in the variable U.

In most cases considered below, only $\sum_{j=1}^{N(Tt)} Y_jX_j$ contributes to the limiting behaviour of $\int_0^{Tt} W(u)\,du$.

We start with a standard limiting behaviour. Specifically, we assume first that $\mathrm{var}(X) = \sigma_X^2 < \infty$ and $\mathrm{var}(Y) = \sigma_Y^2 < \infty$. In particular, there is no LRcD in the counting process $N(t)$ and hence in the cumulative renewal reward process $\int_0^t W(u)\,du$.

Theorem 4.38 *Assume that*

- *Interarrival times have a finite variance*: $\mathrm{var}(X_1) = \sigma_X^2 < \infty$;
- *Rewards have a finite variance*: $\mathrm{var}(Y_1) = \sigma_Y^2 < \infty$.

Then,

$$\lim_{T\to\infty}\lim_{M\to\infty}\frac{W^*_{T,M}(t)}{T^{1/2}M^{1/2}} = \lim_{M\to\infty}\lim_{T\to\infty}\frac{W^*_{T,M}(t)}{T^{1/2}M^{1/2}} \overset{\mathrm{d}}{=} \sigma_{\mathrm{reward},1}B(t),$$

where $(B(t), t \in \mathbb{R})$ *is a standard Brownian motion*,

$$\sigma^2_{\mathrm{reward},1} = \frac{E[X_1^2]E[Y_1^2]}{E[X_1]},$$

and the convergence is to be understood as a finite-dimensional one.

Proof First, we consider the limit taken in the order $\lim_{M\to\infty}$ first, and then $\lim_{T\to\infty}$.

Step 1: Since $W^{(m)}$ $(m = 1, \ldots, M)$ are independent identically distributed processes with finite variance, application of Lemma 4.6 implies that for each T,

$$\lim_{M\to\infty}\frac{1}{M^{1/2}}\sum_{m=1}^{M} W^{(m)}(Tt) \Rightarrow G(Tt)$$

in $D[0,\infty)$, where $G(t)$ $(t \ge 0)$ is a centred stationary Gaussian process with covariance function $cov(W(0), W(u))$.

Step 2: The cumulative process $G^*(\cdot t) = \int_0^{\cdot t} G(t)\,du$ is still a Gaussian process with variance $\mathrm{var}(G^*(Tt)) = \mathrm{var}(\int_0^{Tt} W(u)\,du) = TtE[X_1^2]E[Y_1^2]/\mu$ (see Examples 2.5 and 2.12).

Step 3: The form of the covariance function yields that the process $T^{-1/2}G^*(Tt)$ $(t \ge 0)$ is a Brownian motion.

Now, we consider the reverse order of taking the limits.

Step 1: We use an approximation induced by representation (4.188).

$$\frac{1}{T^{1/2}}\sum_{j=1}^{N(Tt)} Y_jX_j = \left(\frac{N(Tt)}{T}\right)^{1/2}\frac{1}{\sqrt{N(Tt)}}\sum_{j=1}^{N(Tt)} Y_jX_j.$$

Recall that for a stationary renewal process, $N(Tt)/T \to EE[N(t)] = \lambda t = \mu^{-1}t$. Thus, as $T \to \infty$,

$$\frac{1}{T^{1/2}}\sum_{j=1}^{N(Tt)} Y_j X_j \approx \frac{t^{1/2}}{\mu^{1/2}}\frac{1}{(Tt)^{1/2}}\sum_{j=1}^{Tt} Y_j X_j \Rightarrow \frac{1}{\mu^{1/2}}\sqrt{\mathrm{var}(Y_1X_1)}B(t).$$

Since X_1 and Y_1 are independent and $E[Y_1] = 0$, we obtain $\mathrm{var}(Y_1X_1) = E[Y_1^2]E[X_1^2]$.

Step 2: Hence, for each fixed $m = 1, \dots, M$,

$$T^{-1/2}\int_0^{Tt} W^{(m)}(u)\,du \Rightarrow \sigma_{\mathrm{reward},1}B^{(m)}(t),$$

where $B^{(m)}(t)$ are independent standard Brownian motions. Hence, the superposition converges to a Brownian motion.

□

Next, we analyse what happens if the finite variance assumption on the rewards still holds, but the renewal process has intervals with an infinite variance. Recall that then the corresponding counting process $N(t)$ has the LRcD property (see Examples 2.5 and 2.12) since its variance grows faster than linear. Also (see Examples 2.5 and 2.12), the variance of the cumulative process $\int_0^{Tt} W(u)\,du$ grows faster than linear.

Theorem 4.39 *Assume that*

- *Interarrival times are regularly varying*: $P(X_1 > x) \sim C_X x^{-\alpha}$ $(\alpha \in (1,2))$ *as* $x \to \infty$;
- *Rewards have a finite variance* $\mathrm{var}(Y_1) = \sigma_Y^2 < \infty$, *and they are symmetric.*

Then,

$$\lim_{T\to\infty}\lim_{M\to\infty}\frac{W^*_{T,M}(t)}{T^H M^{1/2}} \stackrel{\mathrm{d}}{=} \sigma_{\mathrm{reward},2}B_H(t), \tag{4.189}$$

where $(B_H(t), t \in \mathbb{R})$ *is a standard fractional Brownian motion with Hurst index* $H = (3-\alpha)/2$, *and*

$$\sigma^2_{\mathrm{reward},2} = C_X\frac{2E[Y_1^2]}{E[X_1](\alpha-1)(2-\alpha)(3-\alpha)}.$$

On the other hand,

$$\lim_{M\to\infty}\lim_{T\to\infty}\frac{W^*_{T,M}(t)}{T^{1/\alpha}M^{1/\alpha}} \stackrel{\mathrm{d}}{=} C_{\mathrm{reward},1}Z_\alpha(t), \tag{4.190}$$

where $Z_\alpha(t) \stackrel{\mathrm{d}}{=} t^{1/\alpha}S_\alpha(1,0,0)$ *is a symmetric Lévy process, and*

$$C_{\mathrm{reward},1} = \mu^{-1/\alpha}E^{1/\alpha}\big[|Y_1|^\alpha\big]C_X^{1/\alpha}C_\alpha^{-1}.$$

Sketch of Proof First, we proceed with $\lim_{T\to\infty}\lim_{M\to\infty}$.

Step 1: As in the case of Theorem 4.38, Lemma 4.6 implies that for each T,

$$\lim_{M\to\infty}\frac{1}{M^{1/2}}\sum_{m=1}^{M}W^{(m)}(Tt)\Rightarrow G(Tt)$$

in $D[0,\infty)$, where $G(t)$ $(t\in\mathbb{R})$ is a centred stationary Gaussian process with covariance function $cov(W(0),W(t))$.

Step 2: The cumulative process $G^*(Tt)$ is Gaussian with variance $\sigma_{\text{reward},2}(Tt)^{2H}$, $H=(3-\alpha)/2$ (see Example 2.12).

Step 3: The form of the variance yields that the scaled process $T^{-H}G^*(Tt)$ is a fractional Brownian motion.

Next, we deal with the reversed order of limits.

Step 1: We have

$$\frac{1}{T^{1/\alpha}}\sum_{j=1}^{N(Tt)}Y_jX_j=\left(\frac{N(Tt)}{T}\right)^{1/\alpha}\frac{1}{(N(Tt))^{1/\alpha}}\sum_{j=1}^{N(Tt)}Y_jX_j\approx\frac{1}{\mu^{1/\alpha}}\frac{1}{T^{1/\alpha}}\sum_{j=1}^{Tt}Y_jX_j.$$

By applying Breiman lemma we note that

$$P(Y_1X_1>x)\sim E\big[Y_+^\alpha\big]P(X_1>x)\sim E\big[Y_+^\alpha\big]C_Xx^{-\alpha}$$

and

$$P(Y_1X_1<-x)\sim E\big[Y_-^\alpha\big]P(X_1>x)\sim E\big[Y_-^\alpha\big]C_Xx^{-\alpha}.$$

Thus, application of (4.80) yields

$$\frac{1}{T^{1/\alpha}}\sum_{j=1}^{N(Tt)}Y_jX_j\Rightarrow\mu^{-1/\alpha}E^{1/\alpha}\big[|Y_1|^\alpha\big]C_X^{1/\alpha}C_\alpha^{-1}Z_\alpha(t).$$

Step 2: The result follows by taking $d_M=M^{1/\alpha}$. □

Finally, we analyse the case where both interarrival times and rewards are heavy tailed. We separate both limiting regimes in two theorems below.

Theorem 4.40 *Assume that*

- *Interarrival times are regularly varying*: $P(X_1>x)\sim C_Xx^{-\alpha}$ $(\alpha\in(1,2))$ *as* $x\to\infty$;
- *Rewards are regularly varying*: $P(Y_1>x)\sim C_Yx^{-\beta}$ $(\beta\in(1,2))$ *as* $x\to\infty$; *and they are symmetric*.

We have the following limits as $\lim_{M\to\infty}\lim_{T\to\infty}$:

- *If* $\alpha<\beta<2$, *then* (4.190) *still holds.*

- *If* $\beta < \alpha < 2$, *then*

$$\lim_{M\to\infty}\lim_{T\to\infty}\frac{W^*_{T,M}(t)}{T^{1/\beta}M^{1/\beta}} \stackrel{d}{=} C_{\text{reward},2}Z_\beta(t), \tag{4.191}$$

where $Z_\beta(t) \stackrel{d}{=} t^{1/\beta}S_\beta(1,0,0)$ *is a symmetric Lévy process*, *and*

$$C_{\text{reward},2} = \mu^{-1/\beta}E^{1/\beta}\big[X_1^\beta\big]C_Y^{1/\beta}C_\beta^{-1}.$$

Proof The proof is very similar to that of Theorem 4.39. Recall that the limiting behaviour of $\int_0^{Tt} W(u)\,du$ is determined by $\sum_{j=1}^{N(Tt)} Y_jX_j$. If $\alpha < \beta$, we may proceed exactly in the same way as in Theorem 4.39. Otherwise, if $\beta < \alpha$, then

$$\frac{1}{T^{1/\beta}}\sum_{j=1}^{N(Tt)} Y_jX_j = \left(\frac{N(Tu)}{T}\right)^{1/\beta}\frac{1}{(N(Tt))^{1/\beta}}\sum_{j=1}^{N(Tt)} Y_jX_j \approx \frac{1}{\mu^{1/\beta}}\frac{1}{T^{1/\beta}}\sum_{j=1}^{Tt} Y_jX_j.$$

By applying Breiman lemma we have

$$P(Y_1X_1 > x) \sim E\big[X_1^\beta\big]P(Y_1 > x) \sim E\big[X_1^\beta\big]C_Yx^{-\beta}$$

and

$$P(Y_1X_1 < -x) \sim E\big[X_1^\beta\big]P(Y_1 < -x) \sim E\big[X_1^\beta\big]C_Yx^{-\beta}.$$

Thus, application of (4.80) yields

$$\frac{1}{T^{1/\beta}}\sum_{j=1}^{N(Tt)} Y_jX_j \Rightarrow \mu^{-1/\beta}E^{1/\beta}\big[X_1^\beta\big]C_Y^{1/\beta}C_\beta^{-1}Z_\beta(t).$$

□

We note also in passing that the case $\beta < \alpha$ above does not require that X_1 is regularly varying. Therefore, (4.191) holds also when $\beta < 2$ and $\mathrm{var}(X_1) < \infty$.

We consider now the case of the other limit.

Theorem 4.41 *Assume that*

- *Interarrival times consist of positive integers and are regularly varying*: $P(X_1 > x) \sim C_Xx^{-(\alpha+1)}$ $(\alpha \in (1,2))$ *as* $x \to \infty$;
- *Rewards are regularly varying and symmetric*: $P(Y_1 > x) \sim C_Y\beta x^{-\beta}$ $(\beta \in (1,2))$ *as* $x \to \infty$;

We have the following limits as $\lim_{T\to\infty}\lim_{M\to\infty}$:

- *If* $\beta < \alpha < 2$, *then* (4.191) *holds*.
- *If* $\alpha < \beta < 2$, *then*

$$\lim_{T\to\infty}\lim_{M\to\infty} T^{-(\beta-\alpha+1)/\beta}M^{-1/\beta}W^*_{T,M}(t) \stackrel{d}{=} C_X^{1/\beta}C_Y^{1/\beta}Z^*_\beta(t), \tag{4.192}$$

where $Z^*_\beta(t)$ is symmetric β-stable process with characteristic function

$$E\left[\exp\left(i\sum_{l=1}^{h}\theta_l Z^*_\beta(t_l)\right)\right] = \exp\bigl(-\sigma^\beta(\theta,\mathbf{t})\bigr),$$

where $\mathbf{t} = (t_1,\dots,t_h)^T$, $\theta = (\theta_1,\dots,\theta_h^T)$,

$$\sigma^\beta(\theta,\mathbf{t}) = C_\beta^{-1}\bigl(I(\theta,\mathbf{t}) + J(\theta,\mathbf{t})\bigr),$$

$$I(\theta,\mathbf{t}) = \mu^{-1}\int_0^\infty \left|\sum_{l=1}^{h}\theta_l(t_j\wedge x)\right|^\beta x^{-\alpha}\,dx,$$

$$J(\theta,\mathbf{t}) = \mu^{-1}\alpha\int_0^\infty\int_0^\infty \left|\sum_{l=1}^{h}\theta_l(t_j\wedge u - x)\right|^\beta (u-x)_+^{-\alpha-1}\,dx.$$

We observe that if $\beta < \alpha$, the order of taking limits does not matter. However, if $\alpha < \beta$, we obtain the new process $Z^*_\beta(t)$. This process has stationary increments and is self-similar with self-similarity parameter $H = (\beta - \alpha + 1)/\beta$. For details on this process, we refer to Levy and Taqqu (2000). Furthermore, note that the convergence to $Z^*_\beta(t)$ requires the additional technical assumption that the interarrival times assume positive integers only.

Sketch of Proof We note that the technique of the proofs of Theorems 4.38 or 4.39 does not work. We cannot apply Lemma 4.6 because the process does not have a finite variance. Instead, we present a simplified version of the proofs of Theorems 2.2 and 2.3 in Levy and Taqqu (2000).

We use representation (4.187). Assume for a moment that Y_k ($k \ge 0$) are symmetric β-stable, $Y_1 \stackrel{\mathrm{d}}{=} S_\beta(\eta,0,0)$, $\eta > 0$. Thus, its characteristic function is given by

$$\varphi_Y(\theta) = E\exp(i\theta Y_1) = \exp\bigl(-\eta^\beta|\theta|^\beta\bigr).$$

We compute the characteristic function of $R(Tu) = \int_0^{Tu} W(u)\,du$. Set $\tau_{-1} = 0$. Then, by conditioning on the entire sequence τ_j and using the fact that the random variables Y_j ($j \ge 0$) are i.i.d.,

$$E\left[\exp\left(i\sum_{l=1}^{h}\theta_l R(t_l)\right)\right]$$
$$= E\left[\exp\left(i\sum_{l=1}^{h}\theta_l\left(Y_0\bigl(\min\{t_l,\tau_0\}\bigr) + \sum_{j=0}^{\infty}Y_{j+1}\bigl(\min\{t_l,\tau_{j+1}\} - \tau_j\bigr)\right)\right)\right]$$

$$= \exp\left(-\eta^\beta E\left(\sum_{l=1}^{h}|\theta_l|\left(\min\{t_l,\tau_0\}+\sum_{j=0}^{\infty}\left(\min\{t_l,\tau_{j+1}\}-\tau_j\right)\right)\right)^\beta\right)$$
$$=: \exp\left(-\sigma^\beta(\theta,\mathbf{t};\eta)\right).$$

Since $W^*_{T,M}(t)$ is the sum of independent copies of the process $R(Tu)$, we have

$$E\left[\exp\left(i\sum_{l=1}^{h}\theta_l M^{-1/\beta}W^*_{1,M}(t_l)\right)\right] = \exp\left(-\sigma^\beta(\theta,\mathbf{t})\right).$$

An additional limiting argument applied to random variables Y_j that are regularly varying as in the theorem yields

$$\lim_{M\to\infty} M^{-1/\beta}W^*_{T,M}(t) \stackrel{\mathrm{d}}{=} Z^*_{\beta,T}(t),$$

where $Z^*_{\beta,T}(t)$ $(t\in[0,1])$ is a symmetric β-stable process with characteristic exponent $\sigma^\beta(\theta, T\mathbf{t}; C_Y/C_\beta)$. This process is neither self-similar, nor has it stationary increments.

More technical details are required to establish

$$T^{-(\beta-\alpha+1)/\beta}\sigma^\beta(\theta,T\mathbf{t};C_Y/C_\beta)\to\sigma^\beta(\theta,T\mathbf{t}).$$

This implies the finite-dimensional convergence of $T^{-(\beta-\alpha+1)/\beta}Z^*_{\beta,T}(t)$ to $Z^*_\beta(t)$. □

Several bibliographical notes are in place here. Theorem 4.38 was proven in Taqqu and Levy (1986, Theorem 5). Theorem 4.39 was proven in Taqqu and Levy (1986). Theorem 4.40 was proven in Levy and Taqqu (1987), whereas Theorem 4.41 can be found in Levy and Taqqu (2000) and Pipiras and Taqqu (2000b). In particular, in the latter paper, the authors showed that the limiting process $Z^*_\beta(t)$ is not a linear fractional stable motion. Also see Taqqu (2002) and Willinger et al. (2003) for an overview.

A summary of the results discussed here is given in Table 4.7.

4.9.5 Superposition of ON–OFF Processes

Assume now that we have M independent copies $W^{(m)}(\cdot)$ $(m=1,\ldots,M)$ of the ON–OFF process $W(t)$ defined in (2.77).

We shall assume that the ON and OFF periods in each model have the same distributions: $P(X_{j,\mathrm{on}}(m)>x)=\bar{F}_{\mathrm{on}}(x)$, $P(X_{j,\mathrm{off}}(m)>x)=\bar{F}_{\mathrm{off}}(x)$, where $X_{j,\mathrm{on}}(m)$, $X_{j,\mathrm{off}}(m)$ $(t\in\mathbb{Z})$ are the consecutive ON and OFF periods, respectively, in the mth ON–OFF process $(m=1,\ldots,M)$. Since $W^{(m)}(u)$ are stationary and have the same distribution for each m, we obtain

$$E\left[\int_0^{Tt}\sum_{m=1}^{M}W^{(m)}(u)\,du\right] = TME\left[W(0)\right]t = TM\frac{\mu_{\mathrm{on}}}{\mu_{\mathrm{on}}+\mu_{\mathrm{off}}}t = TM\frac{\mu_{\mathrm{on}}}{\mu}t.$$

Table 4.7 Limits for superposition of cumulative renewal reward processes—tails of interarrival times vs. tails of rewards. The tail parameters $\alpha \in (1,2)$, $\beta \in (1,2)$

Renewal reward processes	Rewards	
	$E[Y_1^2] < \infty$	$RV_{-\beta}, \beta \in (1,2)$
Interarrival times $E[X_1^2] < \infty$	$\lim_{M\to\infty} \lim_{T\to\infty} = \text{Bm}$ $\lim_{T\to\infty} \lim_{M\to\infty} = \text{Bm}$	$\lim_{M\to\infty} \lim_{T\to\infty} = Z_\beta$ $\lim_{T\to\infty} \lim_{M\to\infty} = Z_\beta$
Interarrival times $RV_{-\alpha}, \alpha \in (1,2)$	$\lim_{M\to\infty} \lim_{T\to\infty} = Z_\alpha$ $\lim_{T\to\infty} \lim_{M\to\infty} = \text{fBm}$	$\alpha < \beta$ $\lim_{M\to\infty} \lim_{T\to\infty} = Z_\alpha$ $\lim_{T\to\infty} \lim_{M\to\infty} = Z_\beta^*$ $\beta < \alpha$ $\lim_{M\to\infty} \lim_{T\to\infty} = Z_\beta$ $\lim_{T\to\infty} \lim_{M\to\infty} = Z_\beta$

Recall from Lemma 2.7 that the ON–OFF process has long memory (in the sense of Definition 1.4), or $\int_0^t W(u)du$ has long memory (in the sense of Definition 1.5) if the ON (or OFF) periods are heavy tailed. In this case we are interested in limit theorems for the superposition of ON–OFF processes. Such studies were conducted in Taqqu et al. (1997), Mikosch et al. (2002) or Dombry and Kaj (2011). Specifically, the following two theorems were proven in Taqqu et al. (1997).

Theorem 4.42 *Assume that ON and OFF periods satisfy* (2.78) *and* (2.79), *i.e.*

$$\bar{F}_{\text{on}}(x) = C_{\text{on}} x^{-\alpha_{\text{on}}}, \qquad \alpha_1 \in (1,2), \tag{4.193}$$

$$\bar{F}_{\text{off}}(x) = C_{\text{off}} x^{-\alpha_{\text{off}}}, \qquad \alpha_2 \in (1,2), \tag{4.194}$$

with $\alpha_{\text{on}} < \alpha_{\text{off}}$. *Then,*

$$\lim_{T\to\infty} \lim_{M\to\infty} \frac{W^*_{T,M}(t)}{T^H M^{1/2}} \overset{\text{d}}{=} C_{\text{on}}^{1/2} \sigma_{\text{on–off}} B_H(t),$$

where $(B_H(t), t \in (0,1))$ *is a fractional Brownian motion with Hurst parameter* $H = (3 - \alpha_{\text{on}})/2$, *and*

$$\sigma^2_{\text{on–off}} = \frac{\mu^2_{\text{on–off}}}{(\alpha_{\text{on}} - 1)\mu^3}.$$

Sketch of Proof

Step 1: Since $W^{(m)}(\cdot)$ $(m = 1, \ldots, M)$ are independent identically distributed bounded processes, application of Lemma 4.6 implies

$$\lim_{M\to\infty} \frac{1}{M^{1/2}} \sum_{m=1}^{M} \left\{ W^{(m)}(t) - E\left[W^{(m)}(t)\right] \right\} \Rightarrow G(t),$$

where $G(t)$ $(t \in [0,1])$ is a centred stationary Gaussian process with the covariance function $cov(W(0), W(t))$.

Step 2: Therefore, $\int_0^{Tt} G(t)\,du$ is still a Gaussian process with variance $\mathrm{var}(\int_0^{Tt} W(u)\,du)$. By Lemma 2.7, the variance grows at rate $C_{\text{on}}\sigma^2_{\text{on–off}}(Tt)^{2H}$ as $T \to \infty$, which is the same as for fractional Brownian motion. We conclude

$$\lim_{T\to\infty} \frac{1}{T^H} \int_0^{Tt} G(t)\,du \Rightarrow C_{\text{on}}^{1/2}\sigma_{\text{on–off}} B_H(t).$$

Step 3: Let

$$U(Tt) = \lim_{M\to\infty} \frac{W^*_{T,M}(t)}{T^H M^{1/2}}.$$

The tightness is verified by noting that as $T \to \infty$, for $t_1 < t_2$,

$$\begin{aligned} &E\big[\big(U(Tu_1) - U(Tu_2)\big)^2\big] \\ &= T^{-2H}\,\mathrm{var}\left(\int_0^{T(t_2-t_1)} W(u)\,du\right) \sim C_1\sigma^2_{\text{on–off}}(t_2-t_1)^{2H} \end{aligned}$$

and $2H > 1$. The tightness is verified by applying Lemma 4.5. □

However, similarly to the case of superposition of renewal processes, different orders of taking limits yield completely different limiting processes.

Theorem 4.43 *Assume that ON and OFF periods satisfy* (4.193) *and* (4.194) *with* $\alpha_{\text{on}} < \alpha_{\text{off}}$ *and* $\alpha_{\text{on}} \in (1,2)$. *Then*

$$\lim_{M\to\infty}\lim_{T\to\infty} (MT)^{-1/\alpha} \int_0^{Tt} \left(\sum_{m=1}^{M} \big(W^{(m)}(u) - E\big[W^{(m)}(u)\big]\big)\right) du \overset{\text{d}}{=} C_0 Z_\alpha(t), \tag{4.195}$$

where $Z_\alpha(t) \overset{\text{d}}{=} t^{1/\alpha} S_\alpha(1,1,0)$ *is a Lévy process, and* $C_0 = (\frac{\mu_{\text{off}}}{\mu^{1+1/\alpha}}) C_{\text{on}}^{1/\alpha} C_\alpha^{-1/\alpha}$.

Sketch of Proof

Step 1: First, we show that for each $m = 1, \ldots, M$,

$$\lim_{T\to\infty} T^{-1/\alpha} \int_0^{Tt} \big\{W^{(m)}(u) - E\big[W^{(m)}(u)\big]\big\}\,du \overset{\text{d}}{=} \left(\frac{\mu_{\text{off}}}{\mu^{1+1/\alpha}}\right) C_{\text{on}}^{1/\alpha} C_\alpha^{-1/\alpha} Z_\alpha^{(m)}(t), \tag{4.196}$$

where $Z_\alpha^{(m)}(t) \overset{\text{d}}{=} t^{1/\alpha} S_\alpha(1,1,0)$ are independent Lévy processes.

If $Tt \le \tau_0$, then there are three scenarios possible: either, at 0, the process is ON, then $\int_0^{Tt} W(u)\,du = \min(Tt, X_{0,\text{on}})$; or at 0, the process is OFF, and $X_{0,\text{off}} > Tt$, then $\int_0^{Tt} W(u)\,du = 0$; or at 0, the process is OFF, and $X_{0,\text{off}} < Tt$, then $\int_0^{Tt} W(u)\,du = Tt - X_{0,\text{off}} \le \tau_0 - X_{0,\text{off}} = X_{0,\text{on}}$ (this last situation is shown on Fig. 4.7). In either case, $\int_0^{Tt} W(u)\,du \le X_{0,\text{on}}$. Since $X_{0,\text{on}}$ is a random variable with a finite mean, we conclude that $X_{0,\text{on}}/T^{1/\alpha} \to 0$ in probability as $T \to \infty$.

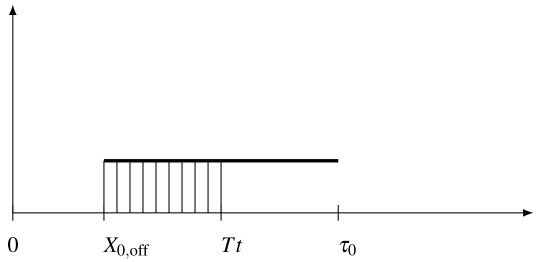

Fig. 4.7 ON–OFF process: The 0th interval starts with OFF period. The *marked area* shows $\int_0^{Tt} W(u)\,du$

If $Tt > \tau_0$, then

$$\int_0^{Tt} W(u)\,du = X_{0,\text{on}} + \sum_{j=1}^{N(Tt)} X_{j,\text{on}} - U,$$

where $U \le X_{N(Tt)+1,\text{on}}$. The first two terms represent the sum of all ON intervals that are at least partially included in $[0, Tt]$. For example, if $\tau_0 < Tt < \tau_1$, then $N(Tt) = 1$ and $\sum_{j=1}^{N(Tt)} X_{j,\text{on}} = X_{1,\text{on}}$; thus, both $X_{0,\text{on}}$ and $X_{1,\text{on}}$ are counted as fully included in $[0, Tt]$. Now, assume that the renewal intervals X_t start with ON periods. It may happen that either $\tau_0 + X_{1,\text{on}} = \tau_0 + X_{N(Tt),\text{on}} < Tt$, and then $U = 0$, or $\tau_0 + X_{1,\text{on}} > Tt$, and in the latter case we have to subtract a portion $(\tau_0 + X_{1,\text{on}} - Tt) \le X_{2,\text{on}}$ that is not included $[0, Tt]$. A similar consideration is valid if the renewal intervals X_t start with OFF periods.
We conclude that the only term that contributes to the limiting behaviour of $\int_0^{Tt} W(u)\,du$ is the sum $\sum_{j=1}^{N(Tt)} X_{j,\text{on}}$. In the same spirit,

$$Tt = X_{0,\text{on}} + X_{0,\text{off}} + \sum_{j=1}^{N(Tt)} X_{j,\text{on}} + \sum_{j=1}^{N(Tt)} X_{j,\text{off}} - Y,$$

where $Y \le X_{N(Tt)+1,\text{on}}$. Thus, informally,

$$\int_0^{Tt} E\big[W(u)\big]\,du = \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}} Tt \approx \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}} \left(\sum_{j=1}^{N(Tt)} X_{j,\text{on}} + \sum_{j=1}^{N(Tt)} X_{j,\text{off}} \right).$$

Consequently, the limiting behaviour of $T^{-1/\alpha} \int_0^{Tt} \{W(u) - E[W(u)]\}\,du$ is determined by

$$\frac{1}{T^{1/\alpha}} \sum_{j=1}^{N(Tt)} \big(J_j - E[J_j]\big),$$

where after some simple algebra

$$\begin{aligned} J_j &= X_{j,\text{on}} - \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}} (X_{j,\text{on}} + X_{j,\text{off}}) \\ &= \frac{\mu_{\text{off}}}{\mu_{\text{on}} + \mu_{\text{off}}} \big(X_{j,\text{on}} - E[X_{j,\text{on}}]\big) - \frac{\mu_{\text{on}}}{\mu_{\text{on}} + \mu_{\text{off}}} \big(X_{j,\text{off}} - E[X_{j,\text{off}}]\big). \end{aligned}$$

Table 4.8 Limits for superposition of ON–OFF processes

Superposition of ON–OFF processes	
ON times with finite variance	$\lim_{M\to\infty}\lim_{T\to\infty} = \text{Bm}$ $\lim_{T\to\infty}\lim_{M\to\infty} = \text{Bm}$
ON times with infinite variance	$\lim_{M\to\infty}\lim_{T\to\infty} = \text{Lévy}$ $\lim_{T\to\infty}\lim_{M\to\infty} = \text{fBm}$

We thus have

$$\frac{1}{T^{1/\alpha}}\sum_{j=1}^{N(Tt)}(J_j - E[J_j]) = \left(\frac{N(Tt)}{T}\right)^{1/\alpha}\frac{1}{(N(Tt))^{1/\alpha}}\sum_{j=1}^{N(Tt)}(J_j - E[J_j]).$$

Recall that for a stationary renewal process $N(Tt)/T \to E[N(t)] = (\mu_{\text{on}} + \mu_{\text{off}})^{-1}\mu^{-1}t$ as $T\to\infty$. Therefore, the limiting behaviour of sum is the same as that of

$$\frac{t^{1/\alpha}}{\mu^{1/\alpha}}\frac{1}{(Tt)^{1/\alpha}}\sum_{j=1}^{Tt}(J_j - E[J_j]).$$

We note that, as $x\to\infty$,

$$P(J_1 > x) \sim \left(\frac{\mu_{\text{off}}}{\mu}\right)^{\alpha_{\text{on}}} C_{\text{on}} x^{-\alpha_{\text{on}}}, \qquad P(J_1 < -x) \sim \left(\frac{\mu_{\text{on}}}{\mu}\right)^{\alpha_{\text{off}}} C_{\text{off}} x^{-\alpha_{\text{off}}}.$$

Since $\alpha = \alpha_{\text{on}} < \alpha_{\text{off}}$, application of (4.80) yields

$$T^{-1/\alpha}\sum_{j=1}^{Tt}(J_j - E[J_j]) \Rightarrow \left(\frac{\mu_{\text{off}}}{\mu}\right) C_{\text{on}}^{1/\alpha} C_\alpha^{-1/\alpha} Z_\alpha(u),$$

where $Z_\alpha(t) \stackrel{\text{d}}{=} t^{1/\alpha} S_\alpha(1,1,0)$ is a Lévy process. We conclude that (4.196) holds.
Step 2: Since the Lévy processes $Z^{(m)}(t)$ are independent, the result follows. □

If the ON and OFF times have a finite variance, similar arguments lead to a Brownian motion as a limit for both limiting regimes. We summarize our observations in Table 4.8.

Similar results as for renewal reward and ON–OFF hold for the Infinite Poisson source model, see Konstantopoulos and Lin (1998), Mikosch et al. (2002).

4.9.6 Simultaneous Limits and Further Extensions

What happens when T and M go to infinity simultaneously? The techniques described above fail. Following Mikosch et al. (2002), one can consider the parameter

M as an increasing function of T, i.e. $M = M(T)$. Alternatively, see Mikosch and Samorodnitsky (2007), one can consider the intensity of the point process τ_j to depend on a number of sources M. Consequently, following Mikosch and Samorodnitsky (2007), we consider the process

$$W^*_{\lambda_M,M}(t) = \sum_{m=1}^{M} W^{(m)*}(\lambda_M t) = \sum_{m=1}^{M} \int_0^{\lambda_M t} W^{(m)}(u)\,du,$$

where the $W(\cdot)$, $W^{(m)}(\cdot)$ ($m \geq 1$) are independent copies of either a renewal reward, an ON–OFF or an $M/G/\infty$ process. We observe that an increase in the intensity can be interpreted as an increase in time in our original cumulative process $W^*_{T,M}(t)$.

Define also a scaling sequence

$$a_M = \sqrt{M \operatorname{var}\left(\int_0^{\lambda_M} W(u)\,du\right)}.$$

In the examples considered above (i.e. renewal reward, ON–OFF, $M/G/\infty$) we have

$$\operatorname{var}\left(\int_0^{\lambda_M} W(u)\,du\right) \sim C\lambda_M^{3-\alpha} L(\lambda_M).$$

For fixed t, convergence of $a_M^{-1} W^*_{\lambda_M,M}(t)$ follows from a classical limit theorem for i.i.d. arrays. Indeed, for some $\delta > 0$, using Hölder's inequality and stationarity of $W(u)$,

$$E\left[\left|W^*_{\lambda_M,M}(t)\right|^{2+\delta}\right] \leq (\lambda_M t)^{1+\delta} \int_0^{\lambda_M t} E\left[\left|W(u) - E\left[W(u)\right]\right|^{2+\delta}\right] du \leq C(\lambda_M t)^{2+\delta}$$

as long as $E[|W(0)|^{2+\delta}] < \infty$. In particular, this is fulfilled for the ON–OFF model and both, renewal reward and $M/G/\infty$, as long as $E[Y_1^{2+\delta}] < \infty$.

If this is the case, we conclude that

$$M^{-\delta/2} \frac{E[|W^*_{\lambda_M,M}(t)|^{2+\delta}]}{(\operatorname{var}(\int_0^{\lambda_M} W(u)\,du))^{1+\delta/2}} \sim M^{-\delta/2} \frac{(\lambda_M t)^{2+\delta}}{\lambda_M^{(3-\alpha)(1+\delta/2)} L^{1+\delta/2}(\lambda_M)}.$$

For each t, the last expression converges to 0 as long as

$$\lambda_M = o\left(M^{1/(\alpha-1+\delta)}\right) \tag{4.197}$$

for some $\delta > 0$.

For each t, we conclude the convergence of $a_M^{-1} W^*_{\lambda_M,M}(t)$ to a normal distribution. The tightness follows clearly from

$$\operatorname{var}\left(a_M^{-1} W^*_{\lambda_M,M}(t-s)\right) = a_M^{-2} M \operatorname{var}\left(\int_0^{\lambda_M(t-s)} W(u)\,du\right) \leq C(t-s)^{3-\alpha}.$$

Therefore, under the fast growth condition (4.197), we conclude the convergence to an fBm. Of course, if we set $\lambda_M = T$, then, as $M \to \infty$, condition (4.197) is clearly fulfilled, and we may recover the convergence in the $\lim_{T\to\infty} \lim_{M\to\infty}$ scheme.

Condition (4.197) is called a *fast growth condition*. Indeed, it means that the number M of sources grows faster than the intensity λ_M, which as mentioned above, can be interpreted as time.

It should be mentioned that in the original paper, Mikosch et al. (2002), the *fast growth* for an $M/G/\infty$ process is defined as

$$\lim_{T\to\infty} \lambda_T T^{1-\alpha} = \infty. \tag{4.198}$$

On the other hand, the *slow growth* is defined as

$$\lim_{T\to\infty} \lambda_T T^{1-\alpha} = 0. \tag{4.199}$$

Similar conditions are imposed in the ON–OFF (Mikosch et al. 2002) or renewal reward context (Taqqu 2002, Pipiras et al. 2004). Roughly speaking, fast growth corresponds to convergence to an fBm, whereas slow growth is responsible for a stable convergence.

Furthermore, similar results to those presented here can be obtained for very general Poisson shot-noise and cluster processes; see Klüppelberg et al. (2003), Klüppelberg and Kühn (2004), Faÿ et al. (2006), Rolls (2010).

However, the picture may change if we consider more complicated models. In particular, we may obtain an fBm limit even in a slow growth regime (see Mikosch and Samorodnitsky 2007, Fasen and Samorodnitsky 2009).

Furthermore, if the limit in (4.199) is a finite, nonnegative constant, then the limiting process is a fractional Poisson process, see Dombry and Kaj (2011).

4.10 Limit Theorems for Extremes

In this section we study the limiting behaviour of partial maxima based on a stationary sequence X_t ($t \in \mathbb{Z}$). We start by recalling some basic results for i.i.d. sequences and illustrating Fréchet and Gumbel domains of attraction. Then, for long-memory sequences, we separate our discussion into the Gumbel and the Fréchet case. A primary example for the first situation is a stationary Gaussian sequence. We argue that there is no influence of dependence (in particular, of long memory) on the limiting behaviour of maxima (Berman 1964, 1971; Leadbetter et al. 1978, 1983; Buchmann and Klüppelberg 2005, 2006). On the other hand, there is no available theory for general linear processes with long memory in the Gumbel case. Furthermore, Breidt and Davis (1998) argue that maxima of Gaussian-based stochastic volatility models (with possible long memory) behave as if the random variables were independent.

Next, we turn our attention to the Fréchet domain of attraction. There, the main tool is point process convergence studied in Sect. 4.3. As we will see, the rate of convergence of maxima of linear processes (weakly or strongly dependent) is the same as for i.i.d. sequences, however, dependence implies that the so-called extremal index is smaller than one (Davis and Resnick 1985). On the other hand, extremes of heavy-tailed stochastic volatility models (with possible long memory) behave again like independent random variables (Davis and Mikosch 2001; Kulik and Soulier 2012, 2013).

These considerations in the Gumbel and Fréchet case may suggest that *long memory does not play any role in the limiting behaviour of maxima*. However, the picture is much more complicated. This will be illustrated by looking at the extremal behaviour of general stationary stable processes in Sect. 4.10.3. That theory was developed in Samorodnitsky (2004, 2006) and Resnick and Samorodnitsky (2004).

We start our discussion with a sequence X_t $(t \in \mathbb{Z})$ of i.i.d. random variables with common distribution function F. Define partial maxima by $M_n = \max\{X_1, \ldots, X_n\}$. The classical Fisher–Tippett theorem identifies three possible limits for M_n. We refer to Chap. 3 in Embrechts et al. (1997) for further details and examples.

Theorem 4.44 *Assume that X_t $(t \in \mathbb{Z})$ is a sequence of i.i.d. random variables. If there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ and a non-degenerate distribution function Λ such that*

$$c_n^{-1}\big(\max\{X_1, \ldots, X_n\} - d_n\big) \overset{d}{\to} \Lambda,$$

then Λ is one of the following distributions: Fréchet, Weibull or Gumbel, defined by the cumulative distribution functions

$$\Lambda_{\text{Frechet}}(x) = \exp\big(-x^{-\alpha}\big) \quad (x > 0, \alpha > 0),$$

$$\Lambda_{\text{Weibull}}(x) = \exp\big(-(-x)^{-\alpha}\big) \quad (x < 0, \alpha > 0),$$

$$\Lambda_{\text{Gumbel}}(x) = \exp\big(-\exp(-x)\big) \quad (x > 0).$$

Example 4.28 Assume that X_t $(t \in \mathbb{N})$ are standard normal. Choose $c_n = (2\ln n)^{-1/2}$ and

$$d_n = \frac{1}{2^{1/2}}\left\{2(\log n)^{1/2} - \frac{\log\log n + \log(4\pi)}{2\sqrt{\log n}}\right\}.$$

Then the limiting distribution is Gumbel.

Example 4.29 Assume that X_t $(t \in \mathbb{N})$ fulfill

$$P(X_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(X_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}. \tag{4.200}$$

(The left-tail behaviour is not needed here, however, we include it for completeness.) Let $A_\beta = A\frac{1+\beta}{2}$. Then,

$$P\big((A_\beta n)^{-1/\alpha}\max\{X_1, \ldots, X_n\} \le x\big) = F^n\big(A_\beta^{1/\alpha}xn^{1/\alpha}\big) = \big(1 - \bar{F}\big(A_\beta^{1/\alpha}xn^{1/\alpha}\big)\big)^n,$$

where $\bar{F}(x) = 1 - F(x)$. Hence, for n large enough,

$$P\left((A_\beta n)^{-1/\alpha} \max\{X_1, \dots, X_n\} \le x\right) = \left(1 - \frac{x^{-\alpha}}{n}\right)^n \to \exp\left(-x^{-\alpha}\right)$$

as $n \to \infty$. In this case $d_n = 0$, $c_n = (A_\beta n)^{1/\alpha}$, and the limiting law is Fréchet.

These examples identify two main classes of distributions and their corresponding extreme value behaviour: (a) the class of regularly varying distributions, that is $\bar{F}(x) = x^{-\alpha} L(x)$ as $x \to \infty$, where L is a slowly varying function; then the limit is Fréchet; and (b) a class of (informally speaking) light-tailed distributions with unbounded support, like normal, log-normal or Gamma; then the limit is Gumbel. The first class is called the *domain of attraction of the Fréchet law*, and the second one the *domain of attraction of the Gumbel law*. The third type, Weibull, appears when the distribution has a bounded support, with a regularly varying behaviour at a boundary. This case will not be discussed here.

In the context of the examples above, a natural question is what happens if we drop the i.i.d. assumption. We will discuss this problem separately for the Fréchet and Gumbel domains of attraction respectively.

4.10.1 Gumbel Domain of Attraction

It turns out that maxima of a (possibly LRD) Gaussian sequence X_t $(t \in \mathbb{N})$ behaves as if the random variables X_t $(t \in \mathbb{N})$ were independent.

Theorem 4.45 *Let X_t $(t \in \mathbb{N})$ be a stationary Gaussian process with covariance function $\gamma(k)$ such that Berman's condition holds*:

$$\lim_{k \to \infty} \log(k)\gamma(k) = 0. \tag{4.201}$$

Then

$$c_n^{-1}\left(\max(X_1, \dots, X_n) - d_n\right) \xrightarrow{d} \Lambda_{\text{Gumbel}},$$

where $c_n = (2 \log n)^{-1/2}$, and

$$d_n = \frac{1}{2^{1/2}}\left\{2(\log n)^{1/2} - \frac{\log\log n + \log(4\pi)}{2\sqrt{\log n}}\right\},$$

cf. Example 4.28.

Proof The proof is only sketched here; some additional technical details can be found in Berman (1964) or Leadbetter et al. (1978, 1983).

We start with the following special version of the normal comparison lemma (see Lemma 3.2 in Leadbetter et al. 1983). For each y,

$$\left| P\big(\max\{X_1, \ldots, X_n\} \le y\big) - \prod_{t=1}^{n} P(X_t \le y) \right|$$
$$\le Cn \sum_{k=1}^{n} \big|\gamma_X(k)\big| \exp\big(-y^2/\big(1 + \big|\gamma_X(k)\big|\big)\big).$$

Next, let us fix x and define $u_n = c_n x + d_n$. Then, since $c_n \to 0$ and $d_n \to \infty$, $u_n \sim d_n$ as $n \to \infty$. Furthermore,

$$d_n^2 = 2\log n + \frac{1}{8}\frac{(\log\log n + \log(4\pi))^2}{\log n} - \log\log n \sim 2\log n - \log\log n.$$

Hence,

$$\exp\big(-u_n^2/2\big) \sim \exp\big(-d_n^2/2\big) \sim n^{-1}\sqrt{\log n} \sim \frac{u_n}{\sqrt{2}n}.$$

We may write

$$n\big|\gamma_X(k)\big| \exp\left(-\frac{u_n^2}{(1 + |\gamma_X(k)|)}\right) = n\big|\gamma_X(k)\big| \exp\big(-u_n^2\big) \exp\left(-\frac{u_n^2|\gamma_X(k)|}{(1 + |\gamma_X(k)|)}\right).$$

Let $\beta > 0$ and $k > n^\beta$. Define $v_n = \sup_{k \ge n^\beta} |\gamma_X(k)|$. Note that

$$v_n u_n^2 \sim 2v_n \log(n) 2\frac{\log n}{\log n^\beta} v_n \log n^\beta = \frac{2}{\beta} v_n \log n^\beta \to 0$$

as $\gamma(n)\log(n) \to 0$. We note that this is exactly the place that Breiman's condition plays a role. Therefore,

$$n \sum_{k=n^\beta}^{n} \big|\gamma_X(k)\big| \exp\left(-\frac{u_n^2}{(1 + |\gamma_X(k)|)}\right)$$
$$\le n \exp\big(-u_n^2\big) v_n \sum_{k=n^\beta}^{n} \exp\left(\frac{u_n^2|\gamma_X(k)|}{(1 + |\gamma_X(k)|)}\right)$$
$$\le n^2 \exp\big(-u_n^2\big) v_n \exp\big(u_n^2 v_n\big) \le C v_n u_n^2 \exp\big(u_n^2 v_n\big) \to 0.$$

On the other hand, there exists $\delta > 0$ such that $1 + |\gamma_X(k)| < 2 - \delta$. Then

$$
\begin{aligned}
&n \sum_{k \le n^\beta} \left|\gamma_X(k)\right| \exp\Bigl(-\frac{u_n^2}{1+|\gamma_X(k)|}\Bigr) \\
&\quad \le n \sum_{k \le n^\beta} \left|\gamma_X(k)\right| \exp\Bigl(-\frac{u_n^2}{2-\delta}\Bigr) \\
&\quad \sim n n^{-2/(2-\delta)} (\log n)^{1/(2-\delta)} \sum_{k \le n^\beta} \left|\gamma_X(k)\right| \le C n^{1+\beta} n^{-2/(2-\delta)} (\log n)^{1/(2-\delta)}
\end{aligned}
$$

since we may assume without loss of generality that $|\gamma_X(k)| \le 1$. The bound converges to 0 when $\beta < \delta/(2+\delta)$. This finishes the proof. □

In Theorem 4.45 we considered a discrete-time process X_t $(t \in \mathbb{Z})$. The result can be extended to general continuous-time Gaussian processes, in particular to fractional Brownian motion $B_H(u)$; see Berman (1971). Furthermore, the result extends to stochastic differential equations driven by fBm. To illustrate this, we consider a continuous-time process $Y(u)$ $(u \in \mathbb{R})$ that solves

$$
Y(v) - Y(u) = \int_u^v \mu\bigl(Y(s)\bigr)\,ds + \int_u^v \sigma\bigl(Y(s)\bigr)\,dB_H(s) \quad (u < v), \tag{4.202}
$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are deterministic functions. We recall from Sect. 2.2.5.2 that if $\mu(x) = \mu < 0$, $\sigma(x) = \sigma$, then the solution is a fractional Ornstein–Uhlenbeck process

$$
Y(u) = \mathrm{FOU}(u) = \sigma \int_{-\infty}^u \exp\bigl(\mu(u-v)\bigr)\,dB_H(v).
$$

The general Berman theory applies and

$$
c_T^{-1}\Bigl(\max_{0 \le u \le T} \mathrm{FOU}(u) - d_T\Bigr) \xrightarrow{d} \Lambda_{\mathrm{Gumbel}},
$$

where

$$
\begin{aligned}
c_T &= \sigma(-\mu)^{-H}\sqrt{\Gamma(H+1/2)}(2\log T)^{-1/2}, \\
d_T &= \frac{(\Gamma(H+1/2))^{1/2}}{2^{1/2}(-\mu)^H}\Bigl\{2(\log n)^{1/2} + \frac{1-H}{2H}\frac{\log\log T}{(\log T)^{1/2}} + \frac{C_0}{(\log T)^{1/2}}\Bigr\}
\end{aligned}
$$

with a constant C_0. We note that the rate of convergence does not depend on the Hurst parameter H. This convergence can be treated as the counterpart to the discrete-time situation in Theorem 4.45.

More generally, Buchmann and Klüppelberg (2005, 2006) study processes of the form $Y_\psi(u) = \psi(\mathrm{FOU}(u))$, where $\mathrm{FOU}(u)$ is a fractional Ornstein–Uhlenbeck process, and ψ is a function. Under general conditions established in those papers,

$Y_\psi(u)$ solves (4.202), and the inverse function ψ^{-1} of ψ fulfills

$$\psi^{-1}(u) = \int_{\psi(0)}^{u} \frac{ds}{\sigma(s)}.$$

Furthermore, the authors give general conditions that guarantee

$$\left(c_T^*\right)^{-1}\left(\max_{0\le u\le T} Y_\psi(u) - \psi(d_T)\right) \overset{d}{\to} \Lambda_{\text{Gumbel}}, \tag{4.203}$$

where c_T^* is possibly different than c_T. The form of c_T^* depends on assumptions on ψ. For example, if

$$\lim_{y\to\infty} \frac{\psi(y + x/y) - \psi(y)}{\psi(y + 1/y) - \psi(y)} = x,$$

then

$$c_T^* = \frac{2^{1/2}(-\mu)^{2H}}{\Gamma(2H+1)}\left\{\psi\left(d_T + \frac{1}{d_T}\right) - \psi(d_T)\right\}.$$

In particular, we can choose $\psi(x) = \exp(x^q)$, $q \in (0, 2)$. Then (4.203) holds with c_T^* as above. We note further that this is not applicable when $q = 2$. Then the limiting distribution is Gumbel. Indeed, note that when Z is standard normal, then e^{Z^2} has a regularly varying tail and hence cannot belong to the Gumbel domain of attraction. We refer to Buchmann and Klüppelberg (2005, 2006) for further results.

A natural question arises. Can we generalize the theorem above to linear processes $X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}$, where ε_t ($t \in \mathbb{Z}$) belong to the domain of attraction of the Gumbel law? The answer is affirmative for weakly dependent sequences. Davis and Resnick (1988, p. 61; see also Rootzén 1986) show that if

$$P\left(c_n^{-1}\left(\max\{\varepsilon_1, \ldots, \varepsilon_n\} - d_n\right) < x\right) \to_d \Lambda(x),$$

then for the partial maxima of the linear process, we have

$$P\left(c_n^{-1}\left(\max\{X_1, \ldots, X_n\} - d_n\right) < x\right) \to_d \Lambda^\theta(x)$$

with some $\theta \in (0, 1)$. The parameter θ is called the *extremal index* and describes the contribution of dependence to the limiting law (see Embrechts et al. 1997 for more details). However, the authors assumed, in particular, that $\sum_{k=0}^{\infty} |a_k| < \infty$, so that long memory is excluded. At the moment there do not seem to be any results for linear processes in the case of long memory.

Breidt and Davis (1998) study stochastic volatility models

$$X_t = \xi_t \sigma_t = \xi_t \exp(\eta_t/2),$$

where ξ_t ($t \in \mathbb{N}$) is an i.i.d. standard normal sequence, independent of the stationary zero-mean Gaussian sequences η_t. After log-transformation, the sequence

$$Y_t := \log X_t^2 = \eta_t + \log \xi_t^2$$

is represented as the sum of a stationary Gaussian sequence and the log of a χ_1^2 random variables. The tail of Y_t has a complicated form, nevertheless it belongs to the domain of attraction of the Gumbel law. A modification of the normal comparison lemma allows us to prove the following result.

Theorem 4.46 *Let X_t $(t \in \mathbb{N})$ be a stochastic volatility model*

$$X_t = \xi_t \exp(\eta_t/2),$$

where ξ_t $(t \in \mathbb{N})$ is an i.i.d. standard normal sequence, independent of the stationary zero-mean Gaussian sequence η_t. Assume that the covariance function of η_t satisfies Berman's condition (4.201), and let $Y_t = \log X_t^2$. Then

$$c_n^{-1}\big(\max(Y_1, \ldots, Y_n) - d_n\big) \xrightarrow{d} \Lambda_{\text{Gumbel}},$$

where $c_n = (2\log n)^{-1/2}$,

$$d_n \sim 2\psi_1(\log n)^{1/2} + \psi_2 \log\big((2\log n)^{1/2}\big) - \psi_3(2\log n)^{-1/2}(\log\log n + \psi_4) + \psi_5,$$

where $\psi_1, \psi_2, \psi_3, \psi_4$ are positive constants, and $c_5 \in \mathbb{R}$.

We observe no influence of possible long memory in volatility on the limiting behaviour of maxima. As for Gaussian sequences considered in Theorem 4.45, the only difference appears in the form of the centering constants d_n.

4.10.2 Fréchet Domain of Attraction

Recall Example 4.29. If the random variables are i.i.d. such that (4.200) holds, then the limiting distribution is Fréchet. This result can also be obtained using point processes. We recall from Sect. 4.3, Theorem 4.13, that

$$N_n := \sum_{t=1}^{n} \delta_{\tilde{c}_n^{-1} X_t} \Rightarrow \sum_{l=1}^{\infty} \delta_{j_l} =: N,$$

where j_l are points of a Poisson process with intensity measure

$$d\lambda(x) = \alpha\left[\frac{1+\beta}{2} x^{-(\alpha+1)} 1\{0 < x < \infty\} + \frac{1-\beta}{2}(-x)^{-(\alpha+1)} 1\{-\infty < x < 0\}\right] dx, \tag{4.204}$$

and $\tilde{c}_n$ is such that $P(|X_1| > \tilde{c}_n) \sim n^{-1}$, that is $\tilde{c}_n \sim A^{1/\alpha} n^{1/\alpha}$. We note that the event $\{\max\{X_1, \ldots, X_n\} \le x\}$ is equivalent to $\{$no points of N_n in $(x, \infty)\}$. Hence, for $x > 0$,

$$P\big(\tilde{c}_n^{-1}\max\{X_1,\dots,X_n\} \le x\big) = P\big(N_n(x,\infty)=0\big) \to P\big(N(x,\infty)=0\big)$$
$$= \exp\Big(-\int_x^\infty d\lambda(u)\Big) = \exp\Big(-\frac{1+\beta}{2}x^{-\alpha}\Big).$$

Changing the scaling from $\tilde{c}_n$ to $c_n = (A_\beta n)^{1/\alpha}$, we immediately conclude

$$P\big(c_n^{-1}\max\{X_1,\dots,X_n\} \le x\big) \to \exp\big(-x^{-\alpha}\big) = \Lambda_{\text{Frechet}}(x).$$

This approach to extremes via point processes can be generalized to dependent sequences, including series with long memory.

We start with linear processes. As in Sect. 4.3, we assume that $X_t = \sum_{k=0}^\infty a_k \varepsilon_{t-k}$, where the random variables ε_t are i.i.d. with a regularly varying distribution, that is

$$P(\varepsilon_1 > x) \sim A\frac{1+\beta}{2}x^{-\alpha}, \qquad P(\varepsilon_1 < -x) \sim A\frac{1-\beta}{2}x^{-\alpha}. \tag{4.205}$$

If $\alpha \in (1,2)$, we assume also that $E(\varepsilon_1) = 0$. Of course, since ε_t are i.i.d.,

$$P\big(c_n^{-1}\max\{\varepsilon_1,\dots,\varepsilon_n\} \le x\big) \to \exp\big(-x^{-\alpha}\big) = \Lambda_{\text{Frechet}}(x),$$

where $c_n = (A_\beta n)^{1/\alpha}$.

We saw in Sect. 4.3 that

$$P(X_1 > x) \sim D_\alpha P(\varepsilon_1 > x), \qquad P(X_1 < -x) \sim D_\alpha P(\varepsilon_1 < -x),$$

where the constant $D_\alpha = \sum_{j=0}^\infty |a_j|^\alpha$ is assumed to be finite. Hence, if X_t^* $(t \in \mathbb{Z})$ is an i.i.d. sequence with the same marginal distribution as X_t, then with the same $c_n = (A_\beta n)^{1/\alpha}$,

$$P\big(c_n^{-1}\max\big\{X_1^*,\dots,X_n^*\big\} \le x\big) \to \exp\big(-D_\alpha x^{-\alpha}\big). \tag{4.206}$$

We note that the constant D_α does not play the role of the extremal index (for the definition see e.g. Embrechts et al. 1997) because the i.i.d. random variables X_t^* have the tail $P(X_1 > x) \sim D_\alpha P(\varepsilon_1 > x)$. The limiting distribution above will serve as a benchmark for comparison with dependent linear processes X_t that have the same marginal distribution as X_t^*. To do this, we will assume without loss of generality that $D_\alpha = 1$.

In Theorem 4.14 we showed, in particular, the following convergence of point processes:

$$\sum_{t=1}^n \delta_{\tilde{c}_n^{-1}X_t} \Rightarrow \sum_{l=1}^\infty \sum_{r=0}^\infty \delta_{j_l a_r},$$

where $\tilde{c}_n \sim A^{1/\alpha}n^{1/\alpha}$. Let us also assume for simplicity that all coefficients a_j are nonnegative. When restricted to $(0,\infty)$, the limiting Poisson process has the intensity measure (cf. Davis and Resnick 1985)

$$\alpha\frac{1+\beta}{2}a_{+}^{\alpha}x^{-(\alpha+1)}dx,$$

where $a_{+}=\max_j a_j$. The same argument as described above for the i.i.d. case leads to the following result on sample extremes for heavy-tailed processes with possible long memory. Limiting behaviour of extremes follows directly from Lemma 4.19 and Theorem 4.14, under the assumptions therein.

Theorem 4.47 *Let X_t ($t\in\mathbb{Z}$) be a linear process where the innovations ε_t ($t\in\mathbb{Z}$) are i.i.d. random variables such that* (4.205) *holds and $E(\varepsilon_1)=0$ if $\alpha\in(1,2)$. Suppose that either for some $\delta<\alpha$,*

$$\sum_{j=0}^{\infty}|a_j|+\sum_{j=0}^{\infty}|a_j|^{\delta}<\infty,$$

or $a_j\sim c_a j^{d-1}$, $d\in(0,1-1/\alpha)$, and ε_t ($t\in\mathbb{Z}$) are symmetric with $\alpha\in(1,2)$. Moreover, assume that $D_\alpha=1$ and $a_j\ge 0$. Then with $c_n=(A_\beta n)^{1/\alpha}$,

$$P\big(c_n^{-1}\max\{X_1,\dots,X_n\}\le x\big)\to\exp\big(-a_{+}x^{-\alpha}\big).$$

This result should be compared with the expression (4.206) for $X_1^*,\dots,X_n^*$ (with $D_\alpha=1$). The additional term $\theta:=a_{+}\in(0,1]$ in the limiting distribution in Theorem 4.47 is the extremal index and describes the effect of dependence on the limiting behaviour of extremes. Since the coefficients a_j are positive, extreme values of the sequence X_t are generated by large positive values of the sequence ε_t. If some of the coefficients are negative, large positive values of X_t are possibly due to large negative values of the innovations, and hence the extremal index will change:

$$\theta=a_{+}+a_{-}\frac{1-\beta}{1+\beta},$$

where $a_{-}=\max\{\max(-a_j),0\}$. We refer to Davis and Resnick (1985) and Embrechts et al. (1997) for more details.

We continue our discussion with heavy-tailed stochastic volatility models, as studied in Sect. 4.3.4. We assume that $X_t=\xi_t\sigma_t$, where ξ_t are i.i.d. such that

$$P(\xi_1>x)\sim A\frac{1+\beta}{2}x^{-\alpha},\qquad P(\xi_1<-x)\sim A\frac{1-\beta}{2}x^{-\alpha}.\qquad(4.207)$$

We will assume also for simplicity that the sequences σ_t and ξ_t are independent from each other. Then, $P(X_1>x)\sim AE(\sigma_1^{\alpha})\frac{1+\beta}{2}x^{-\alpha}$. Hence, if $X_1^*,\dots,X_n^*$ are independent copies of X_1, then with $c_n=(A_\beta n)^{1/\alpha}$,

$$P\big(c_n^{-1}\max\{X_1^*,\dots,X_n^*\}\le x\big)\to\exp\big(-E\big(\sigma_1^{\alpha}\big)x^{-\alpha}\big).$$

Again, the constant $E(\sigma_1^\alpha)$ is related the to marginal behaviour of X_t, not to the dependence structure. In Theorem 4.18 we concluded that the point process based on $X_1, \ldots, X_n$ has the same limit as for the corresponding i.i.d. copies $X_1^*, \ldots, X_n^*$. Directly from Theorem 4.18 we conclude that the limiting behaviour maxima associated with heavy-tailed stochastic volatility models is the same as in the i.i.d. case. There is no influence of any dependence in volatility.

Theorem 4.48 *Consider the LMSV model* $X_t = \xi_t \sigma_t$ $(t \in \mathbb{N})$ *such that* (4.207), *the Breiman condition* (4.94) *and* $E(\sigma_1^{\alpha+\varepsilon}) < \infty$ *with some* $\varepsilon > 0$ *hold. Also, assume that* σ_t $(t \in \mathbb{N})$ *is ergodic. Then*

$$P\big(c_n^{-1} \max\{X_1, \ldots, X_n\} \le x\big) \to \exp\big(-E\big(\sigma_1^\alpha\big)x^{-\alpha}\big).$$

4.10.3 Stationary Stable Processes

Samorodnitsky (2004, 2006) considers a general stationary symmetric α-stable (SαS) process X_t that can be represented by $X_t = \int g_t(s)\,dM(s)$, where M is an SαS random measure. As mentioned in Sect. 1.3.6.3, such processes can be decomposed into a dissipative and a conservative part. As we will indicate below, the dissipative part has no influence on the limiting behaviour of maxima, whereas the conservative part does.

Rosiński (1995) argues that the class of ergodic SαS processes that are generated by the dissipative flow coincides with the class of moving averages $X_t = \int g_t(s)\,dM(s) = \int g(t-s)\,dM(s)$. In particular, consider a Linear Fractional Stable Motion

$$Z_{H,\alpha}(u) = \int_{-\infty}^{\infty} Q_{u,1}(x; H, \alpha)\,dZ_\alpha(x), \tag{4.208}$$

where $Z_\alpha(\cdot)$ is a symmetric α-stable (SαS) Lévy process,

$$Q_{u,1}(x; H, \alpha) = c_1\big[(u-x)_+^{H-1/\alpha} - (-x)_+^{H-1/\alpha}\big] + c_2\big[(u-x)_-^{H-1/\alpha} - (-x)_-^{H-1/\alpha}\big], \tag{4.209}$$

and $H > 1/\alpha$. Let $X_t = Z_{H,\alpha}(t) - Z_{H,\alpha}(t-1)$. Samorodnitsky (2004) proves that in this case

$$P\big(n^{-1/\alpha} \max\{X_1, \ldots, X_n\} \le x\big) \to \exp\big(-Cx^{-\alpha}\big),$$

where C is a positive constant. Hence, the rate of growth of maxima is the same as in the i.i.d. case. We observed this already in the case of moving averages considered in Theorem 4.47.

In contrast, a simple (non-ergodic) example of an SαS process generated by the conservative flow is given by $X_t = Z^{1/\beta}\varepsilon_t$, $(t \in \mathbb{N})$, where Z is a strictly positive α/β-stable random variable, and ε_t is a sequence of i.i.d. symmetric $S_\beta(1, 0, 0)$ random variables, independent of Z, and $0 < \alpha < \beta < 2$. Then, marginally, the random variables X_t are α-stable.

We recall that the β-stability and symmetry of random variables ε_t yield

$$P(\varepsilon_1 > x) \sim \frac{1}{2} C_\beta x^{-\beta},$$

cf. (4.75). Choosing $c_n = (C_\beta/2)^{1/\beta} n^{1/\beta}$, we have

$$\begin{aligned} P\big(c_n^{-1} \max\{X_1, \dots, X_n\} \le x\big) &= E\big[P\big(c_n^{-1} \max\{\varepsilon_1, \dots, \varepsilon_n\} \le Z^{-1/\beta} x | Z\big)\big] \\ &\to E\big[\exp\big(-x^{-\alpha} Z^{\alpha/\beta}\big)\big]. \end{aligned}$$

Hence, even though the random variables X_t are α-stable, the scaling involves β, not α. In other words, maxima grow slower than in the i.i.d. case. This is a general pattern for stable processes generated by a dissipative flow. We refer to Samorodnitsky (2004, 2006) and Resnick and Samorodnitsky (2004) for further details.