Chapter 4

A Panorama of Stochastic-Process Limits

4.1. Introduction

In this chapter and Chapter 7 we give an overview of established stochastic-process limits. These stochastic-process limits are of interest in their own right, but they also can serve as initial stochastic-process limits in the continuous-mapping approach to establish new stochastic-process limits. Indeed, they all can be used to establish stochastic-process limits for queueing models. In fact, a queueing example was already given in Section 2.3. The FCLTs here, when applied to appropriate "cumulative-input" processes, translate into corresponding FCLTs for the workload process in that queueing model when we apply the continuous-mapping approach with the two-sided reflection map.

The fundamental stochastic-process limit is the convergence of a sequence of scaled random walks to Brownian motion in the function space D, provided by Donsker's (1951) theorem, which we already have discussed in Chapter 1. The other established stochastic-process limits mostly come from extensions of Donsker's theorem. In many cases, the limit process has continuous sample paths, in which case the topology on D can be the standard Skorohod J_1 topology. (The topologies on D are introduced in Section 3.3.) Even when the limit process has discontinuous sample paths, we often are able to use the standard J_1 topology, but there are cases in which the M_1 topology is needed. Even when the M_1 topology is not needed, it can be used. Thus, the M_1 topology can be used for all the stochastic-process limits here.

In this overview chapter we state results formally as theorems and give references, but we omit most proofs. We occasionally indicate how a secondary result follows from a primary result.

4.2. Self-Similar Processes

We start by looking at the lay of the land: In this section we consider the general form that functional central limit theorems can take, without imposing any stochastic assumptions such as independence, stationarity or the Markov property.

4.2.1. General CLT's and FCLT's

Consider a sequence $\{X_n : n \geq 1\}$ of $\mathbb{R}^k$-valued random vectors and form the associated *partial sums*

$$S_n \equiv X_1 + \cdots + X_n, \quad n \geq 1 ,$$

with $S_0 \equiv (0, \ldots, 0)$. We say that $\{X_n\}$ or $\{S_n\}$ obeys a *central limit theorem* (CLT) if there exist a sequence of constants $\{c_n : n \geq 1\}$, a vector m and a proper random vector S such that there is convergence in distribution (as $n \to \infty$)

$$c_n^{-1}(S_n - mn) \Rightarrow S \quad \text{in} \quad \mathbb{R}^k .$$

We call m the *translation scaling vector* (or constant if $\mathbb{R}^k = \mathbb{R}$) and $\{c_n\}$ the *space-scaling sequence.* (We might instead allow a sequence of translation vectors $\{m_n : n \geq 1\}$, but that is not the usual case and we do not consider that case here.)

Now form an associated sequence of *normalized partial-sum processes* in $D \equiv D([0, \infty), \mathbb{R}^k)$ by letting

$$\mathbf{S}_n(t) \equiv c_n^{-1}(S_{\lfloor nt \rfloor} - mnt), \quad t \geq 0 , \tag{2.1}$$

where $\lfloor t \rfloor$ is the greatest integer less than or equal to t. We say that $\{X_n\}$, $\{S_n\}$ or $\{\mathbf{S}_n\}$ obeys a *functional central limit theorem* (FCLT) if there exists a proper stochastic process $\mathbf{S} \equiv \{\mathbf{S}(t) : t \geq 0\}$ with sample paths in D such that

$$\mathbf{S}_n \Rightarrow \mathbf{S} \quad \text{in} \quad D , \tag{2.2}$$

for $\mathbf{S}_n$ in (2.1) and some (unspecified here) topology on D.

The classical setting for the CLT and the FCLT occurs when the basic sequence $\{X_n : n \geq 1\}$ is a sequence of *independent and identically distributed*

(IID) random vectors with finite second moments. However, we have not yet made those assumptions. Note that *any* sequence of $\mathbb{R}^k$-valued random vectors $\{S_n : n \geq 1\}$ can be viewed as a sequence of partial sums; just let

$$X_n \equiv S_n - S_{n-1}, \quad n \geq 1 ,$$

with $S_0 \equiv 0$. The partial sums of these new variables X_n obviously coincide with the given partial sums.

We also say a FCLT holds for a continuous-time $\mathbb{R}^k$-valued process $Y \equiv \{Y(t) : t \geq 0\}$ if (2.2) holds for $\mathbf{S}_n$ redefined as

$$\mathbf{S}_n(t) \equiv c_n^{-1}(Y(nt) - mnt), \quad t \geq 0 . \tag{2.3}$$

Here $\{Y(t) : t \geq 0\}$ is the continuous-time analog of the partial-sum sequence $\{S_n : n \geq 1\}$ used in (2.1). Note that the discrete-time process $\{S_n\}$ is a special case, obtained by letting $Y(t) = S_{\lfloor t \rfloor}$, $t \geq 0$.

More generally, we can consider limits for continuous-time $\mathbb{R}^k$-valued stochastic processes indexed by a real variable s where $s \to \infty$. We then have

$$\mathbf{S}_s(t) \equiv c(s)^{-1}(Y(st) - mst), \quad t \geq 0 , \tag{2.4}$$

for $s \geq s_0$. We say that a FCLT holds for $\mathbf{S}_s$ as $s \to \infty$ if $\mathbf{S}_s \Rightarrow \mathbf{S}$ in D as $s \to \infty$ for some (unspecified here) topology on D.

4.2.2. Self-Similarity

Before imposing any stochastic assumptions, it is natural to consider what can be said about the possible translation vectors and space-scaling functions $c(s)$ in (2.4) and the possible limit processes $\mathbf{S}$. Lamperti (1962) showed that convergence of all finite-dimensional distributions has strong structural implications. The possible limit processes are the self-similar processes, which were called semi-stable processes by Lamperti (1962) and then self-similar processes by Mandelbrot (1977); see Chapter 7 of Samorodnitsky and Taqqu (1994).

We say that an $\mathbb{R}^k$-valued stochastic process $\{Z(t) : t \geq 0\}$ is *self-similar with index* $H > 0$ if, for all $a > 0$,

$$\{Z(at) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{a^H Z(t) : t \geq 0\} , \tag{2.5}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution; i.e., if the stochastic process $\{Z(at) : t \geq 0\}$ has the same finite-dimensional distributions as the stochastic process $\{a^H Z(t) : t \geq 0\}$ for all $a > 0$. Necessarily $Z(0) = 0$ w.p.1. The classic example is Brownian motion, which is H-self-similar with $H = 1/2$. The scaling exponent H is often called the *Hurst parameter* in recognition of the early work by Hurst (1951, 1955).

Indeed, we have already encountered self-similarity in Chapter 1. Self-similarity is the natural consequence of the plots looking identical for all sufficently large n. The limit process has identical plots for all time scalings. In the plots the appropriate space scaling is produced automatically.

The class of self-similar processes is very large; e.g., if $\{Y(t) : -\infty < t < \infty\}$ is any stationary process, then

$$Z(t) \equiv t^H Y(\log t), \quad t > 0\,,$$

is H-self-similar. Conversely, if $\{Z(t) : t \geq 0\}$ is H-self-similar, then

$$Y(t) \equiv e^{-tH} Z(e^t), \quad -\infty < t < \infty\,,$$

is stationary; see p. 64 of Lamperti (1962) and p. 312 of Samorodnitsky and Taqqu (1994). In general, the sample paths of self-similar stochastic-processes can be very complicated, see Vervaat (1985), but Lamperti showed that self-similar processes must be *continuous in probability*, i.e.,

$$\|Z(t+h) - Z(t)\| \Rightarrow 0 \quad \text{in} \quad \mathbb{R}^k \quad \text{as} \quad h \to 0$$

for all $t \geq 0$. That does not imply that the stochastic process Z necessarily has a version with sample paths in D, however.

Convergence of the finite-dimensional distributions also implies that the space-scaling function has a special form. The space-scaling function $c(s)$ in (2.4) must be *regularly varying with index* H; see Appendix A. A regularly varying function is a generalization of a simple power; the canonical case is the simple power, i.e., $c(s) = cs^H$ for some constant c.

Theorem 4.2.1. (Lamperti's theorem) *If, for some* $k \geq 1$,

$$(\mathbf{S}_s(t_1), \ldots, \mathbf{S}_s(t_l)) \Rightarrow (\mathbf{S}(t_1), \ldots, \mathbf{S}(t_l)) \quad \text{in} \quad R^{kl} \quad \text{as} \quad s \to \infty$$

for all positive integers l *and all* l*-tuples* $(t_1, \ldots, t_l)$ *with* $0 \leq t_1 < \cdots < l_l$, *where* $\mathbf{S}_s$ *is the scaled process in* (2.4), *then the limit process* $\mathbf{S}$ *is self-similar with index* H *for some* $H > 0$ *and continuous in probability. Then the space scaling function* $c(s)$ *must be regularly varying with the same index* H.

Remark 4.2.1. *Self-similarity in network traffic.* Ever since the seminal work on network traffic measurements by Leland et al. (1994), there has been interest and controversy about the reported self-similarity observed in network traffic. From Lamperti's theorem, we see that some form of self-similarity tends to be an inevitable consequence of any macroscopic view of uncertainty. Since network traffic data sets are very large, they naturally lead to a macroscopic view. *The engineering significance of the reported self-similarity lies in the self-similarity index H.* With centering about a finite mean, the observed index H with $H > 1/2$ indicates that there is extra variability beyond what is captured by the standard central limit theorem. As we saw in Section 2.3, and as we will show in later chapters, that extra traffic burstiness affects the performance in a queue to which it is offered. From the performance-analysis perspective, these are important new ideas, but the various forms of variability have been studied for a long time, as can be seen from Mandelbrot (1977, 1982), Taqqu (1986), Beran (1994), Samorodnitsky and Taqqu (1994) and Willinger, Taqqu and Erramilli (1996). ■

From the point of view of generating simple parsimonious approximations, we are primarily interested in the special case in which the scaling function takes the relatively simple form $c(s) = cs^H$ for some constant c. (We will usually be considering the discrete case in which $c_n = cn^H$.) Then the approximation provided by the stochastic-process limit is characterized by the parameter triple (m, H, c) in addition to any parameters of the limit process. The parameter m is the centering constant, which usually is the mean; the parameter H is the self-similarity index and the space-scaling exponent; and the parameter c is the scale parameter, which appears as a constant multiplier in the space-scaling function. For example, there are no extra parameters beyond the triple (m, H, c) in the case of convergence to standard Brownian motion.

In most applications the underlying sequence $\{X_n\}$ of summands in the partial sums is stationary or asymptotically stationary, so that the limit process $\mathbf{S}$ in (2.2) must also have *stationary increments*, i.e.,

$$\{\mathbf{S}(t+u) - \mathbf{S}(u) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{\mathbf{S}(t) - \mathbf{S}(0) : t \geq 0\}$$

for all $u > 0$. Thus, the prospective limit processes of primary interest are the *H-self-similar processes with stationary increments*, which we denote by H-sssi.

By far, the most frequently occurring H-sssi process is Brownian motion. Brownian motion also has independent increments, continuous sample paths and $H = 1/2$. Indeed, we saw plenty of Brownian motion sample paths in

Chapter 1. The classical FCLT is covered by Donsker's theorem, which we review in Section 4.3. Then the basic sequence $\{X_n\}$ is IID with finite second moments. In Section 4.4 here and Section 2.3 of the Internet Supplement we show that essentially the same limit also occurs when the independence is replaced with weak dependence. The set of processes for which scaled versions converge to Brownian motion is very large. Thus these FCLTs describe remarkably persistent statistical regularity.

After Brownian motion, the most prominent H-sssi processes are the stable Lévy motion processes. The *stable Lévy motion processes* are the H-sssi processes with independent increments. The marginal probability distributions of stable Lévy motions are the stable laws. The Gaussian distribution is a special case of a stable law and Brownian motion is a special stable Lévy motion. The non-Gaussian stable Lévy motion processes are the possible limit processes in (2.2) when $\{X_n\}$ is a sequence of IID random variables with infinite variance, as we will see in Section 4.5. For stable Lévy motions, the self-similarity index H can assume any value greater than or equal to $1/2$. Brownian motion is the special case of a stable Lévy motion with $H = 1/2$. For $H > 1/2$, the stable marginal distributions have power tails and infinite variance. Non-Gaussian stable Lévy motions have discontinuous sample paths, so jumps enter the picture, as we saw in Chapter 1.

4.2.3. The Noah and Joseph Effects

It is possible to have stochastic-process limits with self-similarity index H assuming any positive value. Values of H greater than $1/2$ tend to occur because of either exceptionally large values – the *Noah effect* – or exceptionally strong positive dependence – the *Joseph effect.* The Noah effect refers to the biblical figure Noah who experienced an extreme flood; the Joseph effect refers to the biblical figure Joseph who experienced long periods of plenty followed by long periods of famine; Genesis 41, 29-30: *Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them*; see Mandelbrot and Wallis (1968) and Mandelbrot (1977, 1982).

The Joseph effect occurs when there is strong positive dependence. With the Joseph effect, but without heavy heavy tails, the canonical limit process is *fractional Brownian motion* (FBM). Like Brownian motion, FBM has normal marginal distributions and continuous sample paths. However, unlike Brownian motion, FBM has dependent increments. FBM is a natural H-sssi process exhibiting the Joseph effect, but unlike the stable Lévy motions arising with the Noah effect, FBM is by no means the only H-sssi limit process that can arise with strong dependence and finite second moments; see Vervaat (1985), O'Brien and Vervaat (1985) and Chapter 7 of Samorodnitsky and Taqqu (1994).

We will present FCLTs capturing the Noah effect in Sections 4.5 and 4.7, and the Joseph effect in Section 4.6. It is also possible to have FCLTs exhibiting *both* the Noah and Joseph effects, but unfortunately the theory is not yet so well developed in that area. However, many H-sssi stochastic processes exhibiting both the Noah and Joseph effects have been identified. A prominent example is the *linear fractional stable motion* (LFSM). Like FBM, but unlike stable Lévy motion, LFSM's with positive dependence have continuous sample paths.

The Noah and Joseph effects can be roughly quantified for the H-sssi processes that are also stable processes. A stochastic process $\{Z(t) : t \geq 0\}$ is said to be a *stable process* if all its finite-dimensional distributions are stable laws, all of which have a stable index α, $0 < \alpha \leq 2$; see Section 4.5 below and Chapters 1–3 of Samorodnitsky and Taqqu (1994). The normal or Gaussian distribution is the stable law with stable index $\alpha = 2$. A real-valued random variables X with a stable law with index α, $0 < \alpha < 2$, has a distribution with a power tail with exponent $-\alpha$, satisfying

$$P(|X| > t) \sim Ct^{-\alpha} \quad \text{as} \quad t \to \infty\,,$$

so that X necessarily has infinite variance. Thus, all the one-dimensional marginal distributions of a non-Gaussian stable process necessarily have infinite variance. The class of H-sssi stable processes includes all the specific H-sssi processes mentioned so far: Brownian motion, stable Lévy motion, fractional Brownian motion and linear fractional stable motion.

For H-sssi stable processes with independent increments, $H = \alpha^{-1}$. We are thus led to say that we have *only the Noah effect* when $H = \alpha^{-1} > 2^{-1}$, and we quantify it by the difference $\alpha^{-1} - 2^{-1}$. Similarly, we say that we have *only the Joseph effect* when $H > 1/\alpha = 1/2$, and we quantify it by the difference $H - \alpha^{-1}$. We say that we have both the Noah and Joseph effects when $H > \alpha^{-1} > 2^{-1}$, and quantify the Noah and Joseph effects, respectively, by the differences $\alpha^{-1} - 2^{-1}$ and $H - \alpha^{-1}$, as before. Of course, we have neither the Noah effect nor the Joseph effect when $H = \alpha^{-1} = 2^{-1}$.

It is natural to ask which effect is more powerful. From the perspective of the indices, we see that the Noah effect can be more dramatic: The Noah effect $\alpha^{-1} - 2^{-1}$ can assume any positive value, whereas the Joseph effect $H - \alpha^{-1}$ can be at most $1/2$.

Obviously there should be no Joseph effect when the H-sssi stable process has independent increments. However, it is possible to have 0 Joseph effect without having independent increments; that occurs with the log-fractional stable motion in Section 7.6 of Samorodnitsky and Taqqu (1994). It is also possible to have convergence to non-stable H-sssi limit processes.

It is important to recognize that, even for H-sssi stable processes, we are not constrained to have $H \geq \alpha^{-1} \geq 2^{-1}$; the full range of possibilities is much greater. In this chapter we emphasize positive dependence, which makes the self-similarity index H larger than it would be with independent increments, but in general the possible forms of dependence are more complicated. For example, fractional Brownian motion can have any self-similarity index H with $0 < H \leq 1$; we need not have $H \geq 1/2$. The FBMs with $H < 1/2$ represent *negative dependence* instead of positive dependence. Such negative independence often arises when there is conscious human effort to smooth a stochastic process. For example, an arrival process to a queue may be smoothed by scheduling arrivals, as at a doctor's office. Then the actual arrival process may correspond to some random perturbation of the scheduled arrivals. Then the long-run variability tends to be substantially less than might be guessed from considering only the distribution of the interarrival times between successive customers.

In fact, for H-sssi stable processes with $\alpha < 1$ it is only possible to have negative dependence, because it is possible to have any H with $0 < H \leq \alpha^{-1}$, but *not* any H with $H > \alpha^{-1}$; see p. 316 of Samorodnitsky and Taqqu (1994). However, it is possible to have $1 < \alpha \leq 2$ and $\alpha^{-1} < H < 1$. Properties of H-sssi α-stable processes are described in Samorodnitsky and Taqqu (1994).

4.3. Donsker's Theorem

In this section we consider the classical case in which $\{X_n\}$ is a sequence of IID random variables with finite second moments. The FCLT is Donsker's theorem, which we now describe, expanding upon the discussion in Chapter 1.

4.3.1. The Basic Theorems

Since Donsker's theorem is a generalization of the classical CLT, we start by reviewing the classical CLT. For that purpose, let $N(m, \sigma^2)$ denote a random variable with a *normal or Gaussian distribution* with mean m and variance σ^2. We call the special case of the normal distribution with $m = 0$

and $\sigma^2 = 1$ the *standard normal distribution.* Let Φ be the *cumulative distribution function* (cdf) and n the *probability density function* (pdf) of the standard normal distribution, i.e.,

$$\Phi(x) \equiv P(N(0,1) \le x) \equiv \int_{-\infty}^{x} n(y)dy \ ,$$

where

$$n(x) \equiv \frac{e^{-x^2}}{\sqrt{2\pi}}, \quad -\infty < x < \infty \ .$$

Recall that

$$N(m,\sigma^2) \stackrel{\mathrm{d}}{=} m + \sigma N(0,1)$$

for each $m \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$.

Theorem 4.3.1. (classical central limit theorem) *Suppose that* $\{X_n : n \ge 1\}$ *is a sequence of IID random variables with mean* $m \equiv EX_1$ *and finite variance* $\sigma^2 \equiv Var\, X_1$. *Let* $S_n \equiv X_1 + \cdots + X_n$, $n \ge 1$. *Then* (*as* $n \to \infty$)

$$n^{-1/2}(S_n - mn) \Rightarrow \sigma N(0,1) \quad in \quad \mathbb{R} \ .$$

Donsker's theorem is a FCLT generalizing the CLT above. It is a limit for the entire sequence of partial sums, instead of just the n^{th} partial sum. We express it via the normalized partial-sum process

$$\mathbf{S}_n(t) \equiv n^{-1/2}(S_{\lfloor nt \rfloor} - mnt), \quad t \ge 0 \ , \tag{3.1}$$

in $D \equiv D([0,\infty),\mathbb{R})$, i.e., as in (2.1) with $c_n = \sqrt{n}$.

Theorem 4.3.2. (Donsker's FCLT) *Under the conditions of the CLT in Theorem* 4.3.1,

$$\mathbf{S}_n \Rightarrow \sigma\mathbf{B} \quad in \quad (D, J_1) \ ,$$

where $\mathbf{S}_n$ *is the normalized partial-sum process in* (3.1) *and* $\mathbf{B} \equiv \{\mathbf{B}(t) : t \ge 0\}$ *is standard Brownian motion.*

The limiting Brownian motion in Donsker's FCLT is a Lévy process with continuous sample paths; a *Lévy process* is a stochastic process with stationary and independent increments; see Theorem 19.1 of Billingsley (1968). Those properties imply that an increment $B(s+t) - B(s)$ of Brownian motion $\{B(t) : t \ge 0\}$ is normally distributed with mean mt and variance $\sigma^2 t$ for some constants m and σ^2. *Standard Brownian motion* is Brownian motion with parameters $m = 0$ and $\sigma^2 = 1$.

The most important property of standard Brownian motion is that it exists. Existence is a consequence of Donsker's theorem; i.e., Brownian motion can be defined as the limit process once the limit for the normalized partial sums has been shown to exist.

In applications we often make use of the self-similarity scaling property

$$\{\mathbf{B}(ct) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{\sqrt{c}\mathbf{B}(t) : t \geq 0\} \quad \text{for any} \quad c > 0\ .$$

We can obtain Brownian motion with drift m, diffusion (or variance) coefficient σ^2 and initial position x for any m, $x \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$, denoted by $\{\mathbf{B}(t; m, \sigma^2, x) : t \geq 0\}$, by simply scaling standard Brownian motion:

$$\mathbf{B}(t; m, \sigma^2, x) \equiv x + mt + \sigma\mathbf{B}(t), \quad t \geq 0\,.$$

We have seen Donsker's theorem in action in Chapter 1. Plots of random walks with IID steps converging to Brownian motion are shown in Figures 1.2, 1.3, 1.4 and 1.17.

Donsker's FCLT is an *invariance principle* because the limit depends upon the distribution of X_1 only through its first two moments. By applying the continuous mapping theorem with various measurable real-valued functions on D that are continuous almost surely with respect to Brownian motion, we obtain many useful corollaries. For example, two useful functions are

$$f_1(x) \equiv \sup_{0 \leq t \leq 1} x(t) \tag{3.2}$$

and

$$f_2(x) \equiv \lambda(\{t \in [0,1] : x(t) > 0\}) \tag{3.3}$$

where λ is Lebesgue measure on $[0,1]$; see Section 11 of Billingsley (1968). The supremum function f_1 in (3.2) was discussed in Section 1.2 while establishing limits for the random-walk plots. The function f_2 in (3.3) is not continuous at all x, as can be seen by considering the constant functions $x_n(t) = n^{-1}$, $0 \leq t \leq 1$, $n \geq 1$, but f_2 is measurable and continuous almost surely with respect to Brownian motion. By applying the function f_2 in (3.3), we obtain the arc sine law. For general probability distributions, this result was first obtained directly by Erdös and Kac (1947). The distribution of $f_2(\mathbf{B})$ was found by Lévy (1939).

Corollary 4.3.1. (arc sine law) *Under the assumptions of the CLT in Theorem* 4.3.1,

$$n^{-1}Z_n \Rightarrow f_2(\mathbf{B}) \quad in \quad \mathbb{R}\ ,$$

where Z_n is the number of the first n partial sums $S_1, \ldots, S_n$ that are positive, $\mathbf{B}$ *is standard Brownian motion and*

$$P(f_2(\mathbf{B}) \le x) = \frac{1}{\pi} \int_0^x \frac{dy}{\sqrt{y(1-y)}} = \frac{2}{\pi} arc\ sin(\sqrt{x}), \quad 0 < x < 1 .$$

As indicated in Chapter 1, Donsker's FCLT can be used to derive the limiting distributions in Corollary 4.3.1. Since the limit depends on the distribution of X_1 only through its first two moments, we can work with the special case of a *simple random walk* in which

$$P(X_1 = 1) = P(X_1 = -1) = 1/2 .$$

Combinatorial arguments can be used to calculate the limits for simple random walks; e.g., see Chapter 3 of Feller (1968).

It is interesting that the probability density function $f(y) = \pi^{-1}(y(1-y))^{-1/2}$ of $f_2(\mathbf{B})$ is U-shaped, having a minimum at 1/2. For large n, having 99% of the partial sums positive is about 5 times more likely than having 50% of the partial sums positive.

4.3.2. Multidimensional Versions

It is significant that Theorems 4.3.1 and 4.3.2 extend easily to k dimensions. A key for establishing this extension is the Cramér-Wold device; see p. 49 of Billingsley (1968).

Theorem 4.3.3. (Cramér-Wold device) *For arbitrary random vectors* $(X_{n,1}, \ldots, X_{n,k})$ *in* $\mathbb{R}^k$*, there is convergence in distribution*

$$(X_{n,1}, \ldots, X_{n,k}) \Rightarrow (X_1, \ldots, X_k) \quad in \quad \mathbb{R}^k$$

if and only if

$$\sum_{i=1}^k a_i X_{n,i} \Rightarrow \sum_{i=1}^k a_i X_i \quad in \quad \mathbb{R}$$

for all $(a_1, \ldots, a_k) \in \mathbb{R}^k$.

The multivariate (k-dimensional) CLT involves convergence of normalized partial sums of random vectors to the multivariate normal distribution. We first describe the multivariate normal distribution. A pdf in $\mathbb{R}^k$ of the form

$$f(x_1, \ldots, x_k) = \gamma^{-1} \exp\left(-(1/2) \sum_{i=1}^k \sum_{j=1}^k x_i Q_{i,j} x_j\right) , \qquad (3.4)$$

where $Q \equiv (Q_{i,j})$ is a symmetric $k \times k$ matrix (necessarily with positive diagonal elements) and γ is a positive constant, is a *nondegenerate k-dimensional normal or Gaussian pdf centered at the origin*; see Section III.6 of Feller (1971). The pdf $f(x_1 - m_1, \ldots, x_k - m_k)$ for f in (3.4) is a nondegenerate k-dimensional normal pdf centered at $(m_1, \ldots, m_k)$. A random vector $(X_1, \ldots, X_k)$ with a nondegenerate k-dimensional normal pdf centered at $(m_1, \ldots, m_k)$ has means $EX_i = m_i$, $1 \le i \le k$. Let the *covariance matrix* of a random vector $(X_1, \ldots, X_k)$ in $\mathbb{R}^k$ with means $(m_1, \ldots, m_k)$ be $\Sigma \equiv (\sigma_{i,j}^2)$, where

$$\sigma_{i,j}^2 \equiv E(X_i - m_i)(X_j - m_j) \ .$$

For a nondegenerate normal pdf, the matrices Q and Σ are nonsingular and related by

$$Q = \Sigma^{-1} \ ,$$

and the constant γ in (3.4) satisfies

$$\gamma^2 = (2\pi)^k \, |\Sigma| \ ,$$

where $|\Sigma|$ is the determinant of Σ. Let $N(m, \Sigma)$ denote a random (row) vector with a nondegenerate normal pdf in $\mathbb{R}^k$ centered at $m \equiv (m_1, \ldots, m_k)$ and covariance matrix Σ. Note that

$$N(m, \Sigma) \stackrel{\mathrm{d}}{=} m + N(0, \Sigma) \ .$$

If Σ is the $k \times k$ covariance matrix of a nondegenerate k-dimensional normal pdf, then there exists a nonsingular $k \times k$ matrix C, which is not unique, such that

$$N(0, \Sigma) \stackrel{\mathrm{d}}{=} N(0, I)C \ ,$$

where I is the identity matrix.

We can also allow degenerate k-dimensional normal distributions. We say that a $1 \times k$ row vector Y has a k-dimensional normal distribution with mean vector $m \equiv (m_1, \ldots, m_k)$ and $k \times k$ covariance matrix Σ if

$$Y \stackrel{\mathrm{d}}{=} m + XC \ ,$$

where X is a $1 \times j$ random vector for some $j \le k$ with a nondegenerate j-dimensional normal pdf centered at the origin with covariance matrix I and C is a $j \times k$ matrix with

$$C^t C = \Sigma \qquad (3.5)$$

where C^t is the transpose of C.

The following generalization of the CLT in Theorem 4.3.1 is obtained by applying the Cramér-Wold device in Theorem 4.3.3.

Theorem 4.3.4. (k-dimensional CLT) *Suppose that* $\{X_n : n \geq 1\} \equiv \{(X_{n,1}, \ldots, X_{n,k}) : n \geq 1\}$ *is a sequence of IID random vectors in* $\mathbb{R}^k$ *with* $EX_{1,i}^2 < \infty$ *for* $1 \leq i \leq k$. *Let* $m \equiv (m_1, \ldots, m_k)$ *be the mean vector with* $m_i \equiv EX_{1,i}$ *and* $\Sigma \equiv (\sigma_{i,j}^2)$ *the covariance matrix with*

$$\sigma_{i,j}^2 = E(X_{1,i} - m_i)(X_{1,j} - m_j) \tag{3.6}$$

for all i, j *with* $1 \leq i \leq k$ *and* $1 \leq j \leq k$. *Then*

$$n^{-1/2}(S_n - mn) \Rightarrow N(0, \Sigma) \quad in \quad \mathbb{R}^k$$

where $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$.

A *standard k-dimensional Brownian motion* is a vector-valued stochastic process

$$\mathbf{B} \equiv (\mathbf{B}_1, \ldots, \mathbf{B}_k) \equiv \{\mathbf{B}(t) : t \geq 0\} \equiv \{(\mathbf{B}_1(t), \ldots, \mathbf{B}_k(t)) : t \geq 0\} ,$$

where $\mathbf{B}_1, \ldots, \mathbf{B}_k$ are k IID standard one-dimensional BMs. A general k-dimensional Brownian motion with drift vector $m \equiv (m_1, \ldots, m_k)$, $k \times k$ covariance vector Σ and initial vector $x \equiv (x_1, \ldots, x_k)$, denoted by $\{\mathbf{B}(t; m, \Sigma, x) : t \geq 0\}$ can be constructed by letting

$$\mathbf{B}(t; m, \Sigma, x) = x + mt + \mathbf{B}(t)C , \tag{3.7}$$

where $\mathbf{B}$ is a standard j-dimensional Brownian motion and C is a $j \times k$ matrix satisfying (3.5). In (3.7) we understand that

$$\{\mathbf{B}(t) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{\mathbf{B}(t; 0, I, 0) : t \geq 0\} ,$$

where I is the $j \times j$ identity matrix and 0 is the j-dimensional zero vector.

We now state the k-dimensional version of Donsker's theorem. The limit holds in the space $D^k \equiv D([0, \infty), \mathbb{R}^k)$ with the SJ_1 topology.

Theorem 4.3.5. (k-dimensional Donsker FCLT) *Under the conditions of the k-dimensional CLT in Theorem* 4.3.4,

$$\mathbf{S}_n \Rightarrow \mathbf{B}C \quad in \quad (D^2, SJ_1) ,$$

where $\mathbf{S}_n$ *is the normalized partial-sum process in* (3.1), $\mathbf{B}$ *is a standard* j*-dimensional Brownian motion and* C *is a* $j \times k$ *matrix such that* (3.5) *holds, i.e.,*

$$\mathbf{B}C \stackrel{\mathrm{d}}{=} \{\mathbf{B}(t; 0, \Sigma) : t \geq 0\} \quad in \quad D^k ,$$

where $\Sigma \equiv (\sigma_{i,j}^2)$ *is the covariance matrix of* $(X_{1,1}, \ldots, X_{1,k})$ *in* (3.6).

Proof. The one-dimensional marginals converge by Donsker's theorem, Theorem 4.3.2. That convergence implies that the marginal processes are tight by Prohorov's theorem, Theorem 11.6.1. Tightness of the marginal processes implies tightness of the overall processes by Theorem 11.6.7. Convergence of all the finite-dimensional distributions follows from the CLT in Theorem 4.3.4 and the Cramér-Wold device in Theorem 4.3.3. Finally, tightness plus convergence of the finite-dimensional distributions implies weak convergence in D by Corollary 11.6.1. ∎

It follows from either the k-dimensional Donsker FCLT or the one-dimensional Donsker FCLT that linear functions of the coordinate of the partial-sum process converge to a one-dimensional Brownian motion.

Corollary 4.3.2. *Under the conditions of Theorem* 4.3.5*,*

$$\sum_{i=1}^{k} a_i \mathbf{S}_{n,i} \Rightarrow \sigma \mathbf{B} \quad in \quad D$$

where $\mathbf{S}_n \equiv (\mathbf{S}_{n,1}, \ldots, \mathbf{S}_{n,k})$ *is the normalized partial-sum process in* (3.1)*,* $\mathbf{B}$ *is a standard one-dimensional Brownian motion and*

$$\sigma^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \sigma_{i,j}^2 \ .$$

Donsker's FCLT was stated (as it was originally established) in the framework of a single sequence $\{X_n : n \geq 1\}$. There are extensions of Donsker's FCLT in the framework of a double sequence $\{X_{n,k} : n \geq 1, k \geq 1\}$, paralleling the extensions of the CLT. Indeed, a natural one is a special case of a martingale FCLT, Theorem 2.3.9 in the Internet Supplement.

It can be useful to go beyond the CLT and FCLT to establish bounds on the rate of convergence; see Section 2.2 of the Internet Supplement. For the FCLT, strong approximations can be exploited to produce bounds on the Prohorov distance.

4.4. Brownian Limits with Weak Dependence

For applications, it is significant that there are many generalizations of Donsker's theorem in which the IID assumption is relaxed. Many FCLTs establishing convergence to Brownian motion have been proved with independence replaced by weak dependence. In these theorems, only the space-scaling constant σ^2 in Donsker's theorem needs to be changed. Consequences of Donsker's theorem such as Corollary 4.3.1 thus still hold in these more general settings.

Suppose that we have a sequence of real-valued random variables $\{X_n : n \geq 1\}$. Let $S_n \equiv X_1 + \cdots + X_n$ be the n^{th} partial sum and let $\mathbf{S}_n$ be the normalized partial-sum process

$$\mathbf{S}_n(t) = n^{-1/2}(S_{\lfloor nt \rfloor} - mnt), \quad t \geq 0 \ . \tag{4.1}$$

in D, just as in (3.1). We want to conclude that there is convergence in distribution

$$\mathbf{S}_n \Rightarrow \sigma \mathbf{B} \quad \text{in} \quad (D, J_1) \ , \tag{4.2}$$

where $\mathbf{B}$ is standard Brownian motion and identify the scaling parameters m and σ, without assuming that $\{X_n\}$ is necessarily a sequence of IID random variables.

In this section and in Section 2.3 of the Internet Supplement we review some of the sufficient conditions for (4.2) to hold with the IID condition relaxed. We give only a brief account, referring to Billingsley (1968, 1999), Jacod and Shiryaev (1987) and Philipp and Stout (1975) for more. First, assume that $\{X_n : -\infty < n < \infty\}$ is a two-sided *stationary sequence*, i.e., that $\{X_{k+n} : -\infty < n < \infty\}$ has a distribution (on $\mathbb{R}^\infty$) that is independent of k. (It is always possible to construct a two-sided stationary sequence starting from a one-sided stationary sequence $\{X_n : n \geq 1\}$, where the two sequences with positive indices have the same distribution; e.g., see p. 105 of Breiman (1968).) Moreover, assume that $EX_n^2 < \infty$. The obvious parameter values now are

$$m \equiv EX_n \quad \text{and} \quad \sigma^2 \equiv \lim_{n\to\infty} \frac{Var(S_n)}{n} \ ; \tag{4.3}$$

i.e., m should be the mean and σ^2 should be the *asymptotic variance*, where

$$\sigma^2 = Var\, X_n + 2\sum_{k=1}^{\infty} Cov(X_1, X_{1+k}) \ . \tag{4.4}$$

Roughly speaking, we should anticipate that (4.2) holds with m and σ^2 in (4.3) whenever σ^2 in (4.4) is finite. However, additional conditions are actually required in the theorems. From a practical perspective, however, in applications it *is* usually reasonable to act as if the FCLT is valid if σ^2 in (4.4) is finite, and the main challenge is to find effective ways to calculate or estimate the asymptotic variance σ^2. There is a large literature on estimating the asymptotic variance σ^2 in (4.3), because the asymptotic variance is used to determine confidence intervals around the sample mean for estimates of the steady-state mean; for the sample mean $\bar{X}_n \equiv n^{-1} S_n$,

$$Var\, \bar{X}_n = n^{-2} Var\, S_n, \quad n \geq 1 .$$

Even for a mathematical model, statistical estimation is a viable way to compute the asymptotic variance. We can either estimate σ^2 from data collected from a system being modelled or from output of a computer simulation of the model. For more information, see Section 3.3 of Bratley, Fox and Schrage (1987), Damerdji (1994, 1995) and references therein.

In order for the FCLT in (4.2) to hold, the degree of dependence in the sequence $\{X_n\}$ needs to be controlled. One way to do this is via *uniform mixing conditions.* Here we follow Chapter 4 of Billingsley (1999); also see the papers in Section 2 of Eberlein and Taqqu (1986). To define uniform mixing conditions, let $\mathcal{F}_n \equiv \sigma[X_k : k \leq n]$ be the σ-field generated by $\{X_k : k \leq n\}$ and let $\mathcal{G}_n \equiv \sigma[X_k : k \geq n]$ be the σ-field generated by $\{X_k : k \geq n\}$. We write $X \in \mathcal{F}_k$ to indicate that X is $\mathcal{F}_k$-measurable. Let

$$\alpha_n \equiv \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_k,\ B \in \mathcal{G}_{k+n}\} \tag{4.5}$$

$$\begin{aligned} \rho_n \equiv \sup\{|E[XY]| : X \in \mathcal{F}_k,\ EX = 0, \\ EX^2 \leq 1,\ Y \in \mathcal{G}_{k+n},\ EY = 0,\ EY^2 \leq 1\} \end{aligned} \tag{4.6}$$

$$\phi_n \equiv \sup\{|P(B|A) - P(B)| : A \in \mathcal{F}_k, P(A) > 0, B \in \mathcal{G}_{k+n}\} . \tag{4.7}$$

It turns out that these three measures of dependence are ordered by

$$\alpha_n \leq \rho_n \leq 2\sqrt{\phi_n} .$$

Theorem 4.4.1. (FCLT for stationary sequence with uniform mixing) *Assume that $\{X_n : -\infty < n < \infty\}$ is a two-sided stationary sequence with $Var\, X_n < \infty$ and*

$$\sum_{n=1}^{\infty} \rho_n < \infty . \tag{4.8}$$

for ρ_n in (4.6). *Then the series in* (4.4) *converges absolutely and the FCLT* (4.2) *holds with $m = EX_1$ and σ^2 being the asymptotic variance in* (4.4).

In many applications condition (4.8) will be hard to verify, but it does apply directly to finite-state discrete-time Markov chains (DTMC's), as shown on p. 201 of Billingsley (1999).

Theorem 4.4.2. (FCLT for stationary DTMCs) *Suppose that* $\{Y_n : -\infty < n < \infty\}$ *is the stationary version of an irreducible finite-state Markov chain and let* $X_n = f(Y_n)$ *for a real-valued function* f *on the state space. Then the conditions on* $\{X_n\}$ *in Theorem* 4.4.1 *and the conclusions there hold.*

It is also possible to replace the quantitative measure of dependence in (4.6) with a qualitative characterization of dependence. We say that the sequence $\{X_n : -\infty < n < \infty\}$ is *associated* if, for any k and any two (coordinatewise) nondecreasing real-valued functions f_1 and f_2 on $\mathbb{R}^k$ for which $E[f_i(X_1, \ldots, X_k)^2] < \infty$ for $i = 1, 2$,

$$Cov(f_1(X_1, \ldots, X_k),\ f_2(X_1, \ldots, X_k)) \geq 0\ .$$

(For further discussion of associated processes in queues and other discrete-event systems, see Glynn and Whitt (1989) and Chapter 8 of Glasserman and Yao (1994).) The following FCLT is due to Newman and Wright (1981). See Cox and Grimmett (1984) and Dabrowski and Jakubowski (1994) for extensions.

Theorem 4.4.3. (FCLT for associated process) *If* $\{X_n : -\infty < n < \infty\}$ *is an associated stationary sequence with* $EX_n^2 < \infty$ *and* $\sigma^2 < \infty$ *for* σ^2 *in* (4.4)*, then the FCLT* (4.2) *holds.*

Instead of uniform mixing conditions, we can use ergodicity and martingale properties; see p. 196 of Billingsley (1999). For a stationary process $\{X_n\}$, ergodicity essentially means that the SLLN holds: $n^{-1}S_n \to EX_1$ w.p.1 as $n \to \infty$, where $E|X_1| < \infty$; e.g., see Chapter 6 of Breiman (1968). The sequence of centered partial sums $\{S_n - mn : n \geq 1\}$ is a martingale if $E|X_1| < \infty$ and $E[X_n - m|\mathcal{F}_{n-1}] = 0$ for all $n \geq 1$, where as before $\mathcal{F}_n$ is the σ-field generated by $X_1, \ldots, X_n$.

Theorem 4.4.4. (stationary martingale FCLT) *Suppose that* $\{X_n : -\infty < n < \infty\}$ *is a two-sided stationary ergodic sequence with* $Var\, X_n = \sigma^2$, $0 < \sigma^2 < \infty$, *and* $E[X_n - m|\mathcal{F}_{n-1}] = 0$ *for all* n *for some constant* m. *Then the FCLT* (4.2) *holds with* (m, σ^2) *specified in the conditions here.*

There are two difficulties with the FCLT's stated so far. First, they require stationarity and, second, they do not contain tractable expressions for the asymptotic variance. In many applications, the stochastic process of interest does not start in steady state, but it is asymptotically stationary, and that should be enough. For those situations, it is convenient to exploit regenerative structure. Regenerative structure tends to encompass Markovian structure as a special case. The additional Markovian structure enables us to obtain formulas and algorithms for computing the asymptotic variance. We discuss FCLT's in Markov and regenerative settings in Section 2.3 of the Internet Supplement.

4.5. The Noah Effect: Heavy Tails

In the previous section we saw that the conclusion of Donsker's theorem still holds when the IID assumption is relaxed, with the finite-second-moment condition maintained; only the asymptotic-variance parameter σ^2 in (4.3) and (4.4) needs to be revised, with the key condition being that σ^2 be finite. We now see what happens when we keep the IID assumption but drop the finite-second-moment condition.

As we saw in Chapter 1, when the second moment is infinite, there is a dramatic change! When the second moments are infinite, there still may be limits, but the limits are very different. First, unlike in the finite-second-moment case, there may be no limit at all; the existence of a limit depends critically on regular behavior of the tails of the underlying probability distribution (of X_1). But that regular tail behavior is very natural to assume. When that regular tail behavior holds with infinite second moments, we obtain limits, but limits with different scaling and different limit processes.

Of particular importance to us, the new limit processes have discontinuous sample paths, so that the space D becomes truly important. In this setting we do not need the M_1 topology to establish the FCLT for partial sums of IID random variables, but we do often need the M_1 topology to successfully apply the continuous-mapping approach starting from the initial FCLTs to be described in this section. We illustrate the importance of the M_1 topology in Sections 6.3 and 7.3 below when we discuss FCLTs for counting processes.

The framework here will be a single sequence $\{X_n : n \geq 1\}$ of IID random variables, where $EX_n^2 = \infty$. As before, we will focus on the associated partial sums $S_n = X_1 + \cdots + X_n$, $n \geq 1$, with $S_0 = 0$. We form the normalized processes

$$\mathbf{S}_n(t) \equiv c_n^{-1}(S_{\lfloor nt \rfloor} - m_n nt), \quad t \geq 0 \ , \tag{5.1}$$

in D where $\{m_n : n \geq 1\}$ and $\{c_n : n \geq 1\}$ are general deterministic sequences with $c_n \to \infty$ as $n \to \infty$. Usually we will have $m_n = m$ as in (2.1), but we need translation constants depending on n in one case (when the stable index is $\alpha = 1$). In Sections 4.3 and 4.4 we always had $c_n = \sqrt{n}$. Here will have $c_n/\sqrt{n} \to \infty$; a common case is $c_n = n^{1/\alpha}$ for $0 < \alpha < 2$, where α depends on the asymptotic behavior of the tail probability $P(|X_1| > t)$ as $t \to \infty$. Under regularity conditions, the normalized partial-sum process $\mathbf{S}_n$ in (5.1) will converge in (D, J_1) to a process called stable Lévy motion.

We consider the more general double-sequence (or triangular array) framework using $\{X_{n,k} : n \geq 1, k \geq 1\}$ in Section 2.4 of the Internet Supplement. Unlike in Sections 4.3 and 4.4 above, with heavy-tailed distributions, there is a big difference between a single sequence and a double sequence, because the class of possible limits is much larger in the double-sequence framework: With IID conditions, the possible limits in the framework of double sequences are all Lévy processes. Like the stable Lévy motion considered in this section, general Lévy processes have stationary and independent increments, but the marginal distributions need *not* be stable laws; the marginal distributions of Lévy processes are infinitely divisible distributions (a surprisingly large class). The smaller class of limits we obtain in the single-sequence framework has the advantage of producing more robust approximations; the larger class we obtain in the double-sequence framework has the advantage of producing more flexible approximations.

A *stable stochastic process* is a stochastic process all of whose finite-dimensional distributions are stable laws. The Gaussian distribution is a special case of a stable law, and a Gaussian process is a special case of a stable process, but the limits with infinite second moments will be non-Gaussian stable processes, whose finite-dimensional distributions are non-Gaussian stable laws. The non-Gaussian stable distributions have heavy tails, so that exceptionally large increments are much more likely with a non-Gaussian stable process than with a Gaussian process. We refer to Samorodnitsky and Taqqu (1994) for a thorough treatment of non-Gaussian stable laws and non-Gaussian stable processes. For additional background, see Bertoin (1996), Embrechts, Klüppelberg and Mikosch (1997), Feller (1971), Janicki and Weron (1993) and Zolotarev (1986).

4.5.1. Stable Laws

A random variable X is said to have a *stable law* if, for any positive numbers a_1 and a_2, there is a real number $b \equiv b(a_1, a_2)$ and a positive number $c \equiv c(a_1, a_2)$ such that

$$a_1 X_1 + a_2 X_2 \stackrel{\mathrm{d}}{=} b + cX\,, \tag{5.2}$$

where X_1 and X_2 are independent copies of X and $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution. A stable law is *strictly stable* if (5.2) holds with $b = 0$. Except in the pathological case $\alpha = 1$, a stable law always can be made strictly stable by appropriate centering. Note that a random variable concentrated at one point is always stable; that is a *degenerate* special case.

It turns out that the constant c in (5.2) must be related to the constants a_1 and a_2 there by

$$a_1^\alpha + a_2^\alpha = c^\alpha \tag{5.3}$$

for some constant α, $0 < \alpha \leq 2$. Moreover, (5.2) implies that, for any $n \geq 2$, we must have

$$X_1 + \cdots + X_n \stackrel{\mathrm{d}}{=} n^{1/\alpha} X + b_n \tag{5.4}$$

where $X_1, \ldots, X_n$ are independent copies of X and α is the same constant appearing in (5.3), which is called the *index* of the stable law.

The stable laws on $\mathbb{R}$ can be represented as a four-parameter family. Following Samorodnitsky and Taqqu (1994), let $S_\alpha(\sigma, \beta, \mu)$ denote a stable law (also called α-stable law) on the real line. Also let $S_\alpha(\sigma, \beta, \mu)$ denote a real-valued random variable with the associated stable law. The four parameters of the stable law are: the *index* α, $0 < \alpha \leq 2$; the *scale parameter* $\sigma > 0$; the *skewness parameter* β, $-1 \leq \beta \leq 1$; and the location or *shift parameter* μ, $-\infty < \mu < \infty$. When $1 < \alpha < 2$, the shift parameter is the mean. When $\alpha \leq 1$, the mean is infinite. The logarithm of the characteristic function of $S_\alpha(\sigma, \beta, \mu)$ is

$$\log E e^{i\theta S_\alpha(\sigma,\beta,\mu)} = \begin{cases} -\sigma^\alpha |\theta|^\alpha (1 - i\beta (\operatorname{sign}\theta) \tan(\pi\alpha/2)) + i\mu\theta\,, & \alpha \neq 1 \\ -\sigma|\theta|(1 + i\beta(2/\pi)(\operatorname{sign}\theta) ln(|\theta|)) + i\mu\theta\,, & \alpha = 1\,, \end{cases} \tag{5.5}$$

where $\operatorname{sign}(\theta) = 1$, 0 or -1 for $\theta > 0$, $\theta = 0$ and $\theta < 0$.

The cases $\alpha = 1$ and $\alpha = 2$ are singular cases, with special properties and special formulas. They are boundary cases, at which abrupt change of behavior occurs. The normal law is the special case with $\alpha = 2$; then μ is the mean, $2\sigma^2$ is the variance and β plays no role because $\tan(\pi) = 0$; i.e., $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$. When $\beta = 1$ ($\beta = -1$), the stable distribution is said to be *totally skewed* to the right (left). For limits involving nonnegative summands, we will be interested in the centered totally-skewed stable laws $S_\alpha(\sigma, 1, 0)$.

With the notation $S_\alpha(\sigma, \beta, \mu)$ for stable laws, we can refine the stability property (5.4). If $X_1, \ldots, X_n$ are IID random variables distributed as $S_\alpha(\sigma, \beta, \mu)$, then

$$X_1 + \cdots + X_n \stackrel{\mathrm{d}}{=} \begin{cases} n^{1/\alpha} X_1 + \mu(n - n^{1/\alpha})\,, & \alpha \neq 1 \\ n X_1 + \frac{2}{\pi}\sigma\beta n\, ln\,(n)\,, & \alpha = 1\,. \end{cases} \tag{5.6}$$

From (5.6), we see that $S_\alpha(\sigma, \beta, 0)$ is strictly stable for all $\alpha \neq 1$ and that $S_1(\sigma, 0, \mu)$ is strictly stable.

All stable laws have continuous pdf's, but there are only three classes of these pdf's with convenient closed-form expressions: The first is the Gaussian distribution; as indicated above, $S_2(\sigma, 0, \mu) = N(\mu, 2\sigma^2)$. The second is the *Cauchy distribution* $S_1(\sigma, 0, \mu)$, whose pdf is

$$f(x) \equiv \frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)} \,.$$

In the case $\mu = 0$,

$$P(S_1(\sigma, 0, 0) \leq x) \equiv \frac{1}{2} + \frac{1}{\pi} \operatorname{Arctan}\left(\frac{x}{\sigma}\right) .$$

The third is the *Lévy distribution* $S_{1/2}(\sigma, 1, \mu)$, whose pdf is

$$f(x) = \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x-\mu)^{3/2}} \exp\left(\frac{-\sigma}{2(x-\mu)}\right), \quad x > \mu \,.$$

For the case $\mu = 0$, the cdf is

$$P(S_{1/2}(\sigma, 1, 0) \leq x) = 2(1 - \Phi(\sqrt{\sigma/x}), \quad x > 0 \,,$$

where Φ is the standard normal cdf.

There are simple scaling relations among the non-Gaussian stable laws: For any non-zero constant c,

$$S_\alpha(\sigma, \beta, \mu) + c \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, \beta, \mu + c) \,, \tag{5.7}$$

$$cS_\alpha(\sigma, \beta, \mu) \stackrel{\mathrm{d}}{=} \begin{cases} S_\alpha(|c|\sigma, \operatorname{sign}(c)\beta, c\mu) & \text{if} \quad \alpha \neq 1 \\ S_1(|c|\sigma, \operatorname{sign}(c)\beta, c\mu - \frac{2c}{\pi}(ln(|c|)\sigma\beta) & \text{if} \quad \alpha = 1 \,, \end{cases} \tag{5.8}$$

$$-S_\alpha(\sigma, \beta, 0) \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, -\beta, 0) \,. \tag{5.9}$$

If $S_\alpha(\sigma_i, \beta_i, \mu_i)$ are two independent α-stable random variables, then

$$S_\alpha(\sigma_1, \beta_1, \mu_1) + S_\alpha(\sigma_2, \beta_2, \mu_2) \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, \beta, \mu) \tag{5.10}$$

for

$$\sigma^\alpha = \sigma_1^\alpha + \sigma_2^\alpha, \quad \beta = \frac{\beta_1\sigma_1^\alpha + \beta_2\sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2 \,. \tag{5.11}$$

In general, the stable pdf's are continuous, positive and unimodal on their support. (Unimodality means that there is an argument t_0 such that the pdf is nondecreasing for $t < t_0$ and nonincreasing for $t > t_0$.) The stable laws $S_\alpha(\sigma, 1, \mu)$ with $0 < \alpha < 1$ have support (μ, ∞), while the stable laws $S_\alpha(\sigma, -1, \mu)$ with $0 < \alpha < 1$ have support $(-\infty, \mu)$. All other stable laws (if $\alpha \geq 1$ or if $\alpha < 1$ and $\beta \neq 1$) have support on the entire real line. See Samorodnitsky and Taqqu (1994) for plots of the pdf's.

It is significant that the non-Gaussian stable laws have *power tails*. As in (4.6) in Section 1.4, we write $f(x) \sim g(x)$ as $x \to \infty$ if $f(x)/g(x) \to 1$ as $x \to \infty$. For $0 < \alpha < 2$,

$$P(S_\alpha(\sigma, \beta, \mu) > x) \sim x^{-\alpha} C_\alpha \frac{(1+\beta)}{2} \sigma^\alpha \tag{5.12}$$

and

$$P(S_\alpha(\sigma, \beta, \mu) < -x) \sim x^{-\alpha} C_\alpha \frac{(1-\beta)}{2} \sigma^\alpha , \tag{5.13}$$

where

$$C_\alpha \equiv \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \text{if} \quad \alpha \neq 1 \\ 2/\pi & \text{if} \quad \alpha = 1 \end{cases} \tag{5.14}$$

with $\Gamma(x)$ being the gamma function.

Note that there is an abrupt change in tail behavior at the boundary $\alpha = 2$. For all $\alpha < 2$, the stable pdf has a power tail, but for $\alpha = 2$, the pdf is of order $e^{-x^2/2}$. There also is a discontinuity in the constant C_α in (5.14) at $\alpha = 1$; as $\alpha \to 1$, $C_\alpha \to 1$, but $C_1 = 2/\pi$.

When $\beta = 1$ ($\beta = -1$), the left (right) tail is asymptotically negligible. When also $\alpha < 1$, there is no other tail. When $1 < \alpha < 2$ and $\beta = 1$, the left tail decays faster than exponentially. Indeed, when $1 < \alpha < 2$,

$$\begin{aligned} &P(S_\alpha(\sigma, 1, 0) < -x) \\ &\quad \sim A \left(\frac{x}{\alpha \hat{\sigma}_\alpha} \right)^{-\alpha/(2(\alpha-1))} \exp \left(-(\alpha - 1) \left(\frac{x}{\alpha \hat{\sigma}_\alpha} \right)^{\alpha/(\alpha-1)} \right) \end{aligned} \tag{5.15}$$

where

$$A \equiv (2\pi\alpha(\alpha-1))^{-1/2} \quad \text{and} \quad \hat{\sigma}_\alpha \equiv \sigma(\cos((\pi/2)(2-\alpha))^{-1/\alpha} .$$

When $\alpha = 1$ and $\beta = 1$,

$$P(S_1(\sigma, 1, 0) < -x) \sim \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(\pi/2\sigma)x - 1}{2} - e^{(\pi/2\sigma)x-1} \right) . \tag{5.16}$$

Consequently, the Laplace transform of $S_\alpha(\sigma, 1, 0)$ is well defined, even though the pdf has the entire real line for its support. In particular, the logarithm of the Laplace transform of $S_\alpha(\sigma, 1, 0)$ is

$$\psi_\alpha(s) \equiv \log E e^{-sS_\alpha(\sigma,1,0)} = \begin{cases} -\sigma^\alpha s^\alpha / \cos(\pi\alpha/2), & \alpha \neq 1 \\ 2\sigma s \ \ln(s)/\pi, & \alpha = 1\,, \end{cases} \tag{5.17}$$

for $\mathrm{Re}(s) \geq 0$.

From the asymptotic form above, we can deduce properties of the moments. In particular, for $0 < \alpha < 2$,

$$\begin{aligned} E|S_\alpha(\sigma, \beta, \mu)|^p < \infty \quad &\text{for} \quad 0 < p < \alpha \qquad \text{and} \\ E|S_\alpha(\sigma, \beta, \mu)|^p = \infty \quad &\text{for} \quad p \geq \alpha\ . \end{aligned} \tag{5.18}$$

4.5.2. Convergence to Stable Laws

We now discuss convergence to stable laws. A cdf F on $\mathbb{R}$ is said to be in the *domain of attraction* of the stable law $S_\alpha(\sigma, \beta, \mu)$ if there exist constants m_n and c_n such that

$$c_n^{-1}(S_n - m_n) \Rightarrow S_\alpha(\sigma, \beta, \mu)\ , \tag{5.19}$$

where $S_n = X_1 + \cdots + X_n$, $n \geq 1$, $\{X_n : n \geq 1\}$ is a sequence of IID random variables and X_1 has cdf F. By (5.4), $S_\alpha(\sigma, \beta, \mu)$ is contained in the domain of attraction of $S_\alpha(\sigma, \beta, \mu)$ for all $(\alpha, \sigma, \beta, \mu)$. Clearly, by scaling, it suffices to let $\sigma = 1$ and $\mu = 0$. Hence, only the parameters α and β are unaltered by scaling. A cdf F is said to be in the *normal domain of attraction* of the stable law $S_\alpha(\sigma, \beta, \mu)$ if, in addition to being in the domain of attraction, the constants c_n in (5.19) can be chosen so that $\mu = 0$, $\sigma = 1$ and $c_n = cn^{1/\alpha}$ for some constant c.

This limit theory is classical; see Gnedenko and Kolmogorov (1968), Feller (1971) and p. 50 of Samorodnitsky and Taqqu (1994). Naturally, a key role is played by the cdf F of X_1. A big role is also played by the cdf of $|X_1|$; let G be its cdf and $G^c \equiv 1 - G$ its complementary cdf (ccdf), i.e.,

$$G^c(x) \equiv P(|X_1| > x) = 1 - F(x) + F(-x)\ . \tag{5.20}$$

The conditions make use of regularly varying functions; see Appendix A. We write $G^c \in \mathcal{R}(-\alpha)$ if the ccdf is regularly varying with index $-\alpha$. That holds if and only if $G^c(x) = x^{-\alpha}L(x)$ for some slowly varying function L.

Theorem 4.5.1. (stable-law CLT) *Let $\{X_n : n \geq 1\}$ be an IID sequence of real-valued random variables with cdf F. The cdf F belongs to the domain of attraction of $S_\alpha(1, \beta, 0)$ for $0 < \alpha < 2$, i.e., (5.19) holds for $\sigma = 1$ and $\mu = 0$, if and only if both $G^c \in \mathcal{R}(-\alpha)$, i.e.,*

$$x^\alpha G^c(x) = L(x) \tag{5.21}$$

for G^c in (5.20), where L is slowly varying, and

$$F^c(x)/G^c(x) \to \frac{1+\beta}{2} \quad as \quad x \to \infty \; . \tag{5.22}$$

The space-scaling constants c_n in (5.19) then must satisfy

$$\lim_{n\to\infty} \frac{nL(c_n)}{c_n^\alpha} = C_\alpha \; , \tag{5.23}$$

for C_α in (5.14) and L in (5.21). The translation constants m_n in (5.19) may be chosen to satisfy

$$m_n = \begin{cases} 0 & if \quad 0 < \alpha < 1 \\ nc_n \int_{-\infty}^{\infty} \sin(x/c_n) dF(x) & if \quad \alpha = 1 \\ n \int_{-\infty}^{\infty} x dF(x) & if \quad 1 < \alpha < 2 \; . \end{cases} \tag{5.24}$$

If c_n satisfies (5.23), then $c_n = n^{1/\alpha} L_0(n)$, where L_0 is slowly varying (in general different from L in (5.21)).

At the expense of changing the scaling constants σ and μ in the limit, the normalization constants c_n in Theorem 4.5.1 can be chosen to be the $(1 - n^{-1})^{\text{th}}$ percentile of the cdf G instead of (5.23); i.e., we can let

$$c_n \equiv (1/G^c)^{\leftarrow}(n) \equiv \inf\{y : G(y) \geq n\} \; ; \tag{5.25}$$

see p. 3 of Resnick (1987) and p. 78 of Embrechts et. al. (1997).

Theorem 4.5.1 contains the result about normal domains of attraction as a special case. Note that the condition has the summand having a power law.

Theorem 4.5.2. (normal domain of attraction of a stable law) *Let $\{X_n : n \geq 1\}$ be an IID sequence with cdf F. The cdf belongs to the normal domain of attraction of $S_\alpha(1, \beta, 0)$ for $0 < \alpha < 2$, i.e., (5.19) holds with $c_n = cn^{1/\alpha}$, $\sigma = 1$ and $\mu = 0$, if and only if both*

$$G^c(x) \sim Ax^{-\alpha} \quad as \quad x \to \infty \tag{5.26}$$

for G^c in (5.20) and positive constants A and α, and

$$\frac{F^c(x)}{G^c(x)} \to \frac{1+\beta}{2} \quad as \quad x \to \infty \; . \tag{5.27}$$

The space-scaling constants can then be

$$c_n = (A/C_\alpha)^{1/\alpha} n^{1/\alpha} \; , \tag{5.28}$$

where the pair (A, α) is from (5.26) and C_α is the stable-law asymptote in (5.14). The translation constant m_n can then be as in (5.24).

Proof. Given Theorem 4.5.1, for $0 < \alpha < 2$, c_n can be chosen to be of the form $cn^{1/\alpha}$ for some constant c, while satisfying (5.23), if and only if the slowly varying function $L(t)$ approaches a constant as $t \to \infty$. Thus, a cdf belongs to the normal domain of attraction of a stable law of index α if and only if (5.21) and (5.22) hold with $L(t) \to A$ as $t \to \infty$ for some constant A. In other words, for the normal domain of attraction, (5.21) should be restated as (5.26). Then the left side of (5.23) becomes nA/c_n^α. If $nA/c_n^\alpha \to C_\alpha$ as $n \to \infty$, then $n^{1/\alpha}A^{1/\alpha}/c_n \to C_\alpha^{1/\alpha}$ as $n \to \infty$, so that it suffices to use (5.28). ■

It is useful to have a sanity check to verify the form of the space-scaling constants in (5.28). That is provided by considering the special case in which

$$X_n \stackrel{\mathrm{d}}{=} (A/C_\alpha)^{1/\alpha} S_\alpha(1, \beta, 0) \; .$$

Note that this X_n satisfies (5.26) and (5.27); e.g., by (5.12) and (5.13),

$$P(A/C_\alpha)^{1/\alpha}|S_\alpha(1,\beta,0)| > x) = P(|S_\alpha(1,\beta,0)| > (C_\alpha/A)^{1/\alpha}x) \sim Ax^{-\alpha} \; .$$

However, by (5.6),

$$(C_\alpha/nA)^{1/\alpha}(X_1 + \cdots + X_n) \stackrel{\mathrm{d}}{=} S_\alpha(1,\beta,0) \quad \text{for all} \quad n \geq 1 \; .$$

Hence we must have (5.28).

From a mathematical perspective, Theorem 4.5.1 is appealing because it fully characterizes when the limit exists and gives its value. However, from a practical perspective, the special case in Theorem 4.5.2 may be more useful because it yields a more parsimonious approximation as a function of n. For the case $0 < \alpha < 2$, Theorem 4.5.1 yields the approximation

$$S_n \stackrel{\mathrm{d}}{\approx} ES_n + c_n S_\alpha(1,\beta,0) \; ,$$

with the approximation as a function of n being a function of α, β and the entire (in general complicated) sequence $\{c_n : n \geq 1\}$. On the other hand, for the same case, Theorem 4.5.2 yields the approximation

$$S_n \stackrel{\mathrm{d}}{\approx} ES_n + cn^{1/\alpha} S_\alpha(1, \beta, 0) \tag{5.29}$$

with the approximation as a function of n being a function only of the three parameters α, β and c.

In applications it is usually very difficult to distinguish between a power tail and a regularly-varying non-power tail of the same index. Even estimating the stable index α itself can be a challenge; see Embrechts et al. (1997), Resnick (1997) and Adler, Feldman and Taqqu (1998).

4.5.3. Convergence to Stable Lévy Motion

We now want to obtain the FCLT generalization of the stable-law CLT in Theorem 4.5.1. The limit process is a stable Lévy motion, which is a special case of Lévy process. A *Lévy process* is a stochastic process $\mathbf{L} \equiv \{\mathbf{L}(t) : t \geq 0\}$ with sample paths in D such that $\mathbf{L}(0) = 0$ and $\mathbf{L}$ has stationary and independent increments; we discuss Lévy processes further in Section 2.4 of the Internet Supplement. A standard *stable (or α-stable) Lévy motion* is a Lévy process $\mathbf{S} \equiv \{\mathbf{S}(t) : t \geq 0\}$ such that the increments have stable laws, in particular,

$$\mathbf{S}(t+s) - \mathbf{S}(s) \stackrel{\mathrm{d}}{=} S_\alpha(t^{1/\alpha}, \beta, 0) \stackrel{\mathrm{d}}{=} t^{1/\alpha} S_\alpha(1, \beta, 0) \tag{5.30}$$

for any $s \geq 0$ and $t > 0$, for some α and β with $0 < \alpha \leq 2$ and $-1 \leq \beta \leq 1$. The adjective "standard" is used because the shift and scale parameters of the stable law in (5.30) are $\mu = 0$ and $\sigma = t^{1/\alpha}$ (without an extra multiplicative constant). When we want to focus on the parameters, we call the process a standard (α, β)-stable Lévy motion. Formula (5.30) implies that a stable Lévy motion has stationary increments. When $\alpha = 2$, stable Lévy motion is Brownian motion. Except in the cases when $\alpha = 1$ and $\beta \neq 1$, a stable (or α-stable) Lévy motion is self-similar with self-similarity index $H = 1/\alpha$, i.e.,

$$\{\mathbf{S}(ct) : t \geq 0\} \stackrel{\mathrm{d}}{=} \{c^{1/\alpha}\mathbf{S}(t) : t \geq 0\} .$$

In many ways, non-Brownian ($\alpha < 2$) stable Lévy motion is like Brownian motion ($\alpha = 2$), but it is also strikingly different. For example, Brownian motion has continuous sample paths, whereas stable Lévy motion, except for its deterministic drift, is a *pure-jump process.* It has infinitely many discontinuities in any finite interval w.p.1. On the positive side, there is a version with sample paths in D, and we shall only consider that version. For $0 < \alpha < 1$ and $\beta = 1$, stable Lévy motion has nondecreasing sample paths, and is called a *stable subordinator.*

For $\alpha \geq 1$ and $\beta = 1$, stable Lévy motion has no negative jumps; it has positive jumps plus a negative drift. For $\alpha > 1$, stable Lévy motion (like Brownian motion) has sample paths of unbounded variation in each bounded interval. Like Brownian motion, stable Lévy motion has complicated structure from some points of view, but also admits many simple formulas.

In the case of IID summands (for both double and single sequences), Skorohod (1957) showed that all ordinary CLT's have FCLT counterparts in (D, J_1); see Jacod and Shiryaev (1987) for further discussion, in particular, see Theorems 2.52 and 3.4 on pages 368 and 373. Hence the FCLT generalization of Theorem 4.5.1 requires no new conditions.

Theorem 4.5.3. (stable FCLT) *Under the conditions of Theorem 4.5.1, in addition to the CLT*

$$c_n^{-1}(S_n - m_n) \Rightarrow S_\alpha(1, \beta, 0) \quad in \quad \mathbb{R} ,$$

there is convergence in distribution

$$\mathbf{S}_n \Rightarrow \mathbf{S} \quad in \quad (D, J_1) \tag{5.31}$$

for the associated normalized process

$$\mathbf{S}_n(t) \equiv c_n^{-1}(S_{\lfloor nt \rfloor} - m_n t), \quad t \geq 0 , \tag{5.32}$$

where the limit $\mathbf{S}$ *is a standard* (α, β)*-stable Lévy motion, with*

$$\mathbf{S}(t) \stackrel{\mathrm{d}}{=} t^{1/\alpha} S_\alpha(1, \beta, 0) \stackrel{\mathrm{d}}{=} S_\alpha(t^{1/\alpha}, \beta, 0) .$$

We have seen the stable FCLT in action in Chapter 1. Plots of random walks with IID steps having Pareto(p) distributions converging to stable Lévy motion with $\alpha = p$ are shown in Figures 1.20, 1.21 and 1.22 for $p = 3/2$ and in Figures 1.19, 1.25 and 1.26 for $p = 1/2$. We have also seen how the stable FCLT can be applied with the continuous-mapping approach to establish stochastic-process limits for queueing models. Plots of workload processes converging to reflected stable Lévy motion appear in Figures 2.3 and 2.4.

Of course, there is a corresponding FCLT generalization of Theorem 4.5.2. There also is a k-dimensional generalization of Theorem 4.5.3 paralleling Theorem 4.3.5 in Section 4.3. The proof is just like that for Theorem 4.3.5, again exploiting the Cramér-Wold device in Theorem 4.3.3. To apply the Cramér-Wold device, we use the fact that a stochastic process is strictly stable (stable with index $\alpha \geq 1$) if and only if all linear combinations (over time points and coordinates) of the process are again strictly stable (stable with index $\alpha \geq 1$); combine Theorems 2.1.5 and 3.1.2 of Samorodnitsky and Taqqu (1994). (For $\alpha \neq 1$, we always work with the centered stable laws having $\mu = 0$, so that they are strictly stable.)

4.5.4. Extreme-Value Limits

We have observed that the sample paths of stable Lévy motion are discontinuous. For that to hold, the maximum jump X_n must be asymptotically of the same order as the centered partial sum $S_n - mn$ for $\alpha > 1$ and the uncentered sum S_n for $\alpha < 1$. That was illustrated by the random-walk sample paths in Section 1.4. Further insight into the sample-path structure, and to the limit more generally, can be obtained from extreme-value theory, for which we draw upon Resnick (1987) and Embrechts et al. (1997). We will focus on the successive maxima of the random variables $|X_n|$. Let

$$M_n \equiv \{|X_1|, |X_2|, \ldots, |X_n|\}, \quad n \geq 1 . \tag{5.33}$$

As in (5.20), $|X_1|$ has ccdf G^c. Extreme-value theory characterizes the possible limit behavior of the successive maxima M_n, with scaling. Of special concern to us is the case in which the limiting cdf is the *Fréchet cdf*

$$\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0, \end{cases} \tag{5.34}$$

which is defined for all $\alpha > 0$. Let Φ_α also denote a random variable with cdf Φ_α. Here is the relevant extreme-value theorem (which uses the concept of regular variation; see Appendix A and Section 1.2 of Resnick (1987):

Theorem 4.5.4. (extreme-value limit) *Suppose that $\{|X_n| : n \geq 0\}$ is a sequence of IID random variables having cdf G with $EX_n^2 = \infty$. There exist constants c_n and b_n such that $c_n(M_n - b_n)$ converges in distribution to a nondegenerate limit for M_n in (5.33) if and only if $G^c \in \mathcal{R}(-\alpha)$, in which case*

$$c_n^{-1} M_n \Rightarrow \Phi_\alpha \quad in \quad \mathbb{R} , \tag{5.35}$$

where Φ_α *has the Fréchet cdf in* (5.34) *and the scaling constants may be*

$$c_n \equiv (1/G^c)^{\leftarrow}(n)$$

as in (5.25).

As noted after Theorem 4.5.1, we can also use the scaling constant c_n in (5.25) in the CLT and FCLT for partial sums; i.e., under the conditions of Theorem 4.5.1, we have

$$c_n^{-1} M_n \Rightarrow \Phi_\alpha \ ,$$

$$c_n^{-1}(S_n - nm_n) \Rightarrow S_\alpha(\sigma, \beta, 0)$$

and

$$\mathbf{S}_n \Rightarrow \mathbf{S} \ ,$$

where

$$\mathbf{S}_n(t) = c_n^{-1}(S_{\lfloor nt \rfloor} - m_n nt), \quad t \geq 0 \ ,$$

$$m_n = \begin{cases} 0 & \text{if} \quad 0 < \alpha < 1 \\ EX_1 & \text{if} \quad 1 < \alpha < 2 \ , \end{cases} \tag{5.36}$$

and $\mathbf{S}$ is a nondegenerate stable Lévy motion with

$$\mathbf{S}(1) \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, \beta, 0)$$

for some σ, β and the scaling constants c_n throughout being as in (5.25).

It turns out that we can also obtain a limit for M_n by applying the continuous mapping theorem with the FCLT in (5.31). For that purpose, we exploit the *maximum-jump functional* $J : D \to \mathbb{R}$ defined by

$$J(x) = \sup_{0 \leq t \leq 1} \{|x(t) - x(t-)|\} \ . \tag{5.37}$$

In general, the maximum-jump function is not continuous on D, but it is almost surely with respect to stable Lévy motion; see p. 303 of Jacod and Shiryaev (1987). As before, let $Disc(x)$ be the set of discontinuities of x.

Theorem 4.5.5. (maximum jump function) *The maximum-jump function* J *in* (5.37) *is measurable and continuous on* (D, J_1) *at all* $x \in D$ *for which* $1 \in Disc(x)^c$. *Hence,* J *is continuous almost surely with respect to stable Lévy motion.*

Hence we can apply the continuous mapping theorem in Section 2.7 with Theorems 4.5.3–4.5.5 to obtain the following result. See Resnick (1986) for related results.

Theorem 4.5.6. (joint limit for normalized maximum and sum) *Under the conditions of Theorem 4.5.1, we have the FCLT* (5.31) *with* (5.32) *for* c_n *in* (5.25) *and*

$$c_n^{-1}(M_n, S_n - nm_n) \Rightarrow (J(\mathbf{S}), \mathbf{S}(1)) \quad in \quad \mathbb{R}^2 ,$$

where

$$J(\mathbf{S}) \stackrel{\mathrm{d}}{=} \Phi_\alpha$$

for J *in* (5.37), Φ_α *in* (5.34) *and* m_n *in* (5.36). *Consequently, on any positive interval the stable process* $\mathbf{S}$ *has a jump w.p.1. and* $(S_n - nm_n)/M_n$ *has a nondegenerate limit as* $n \to \infty$.

More generally, it is interesting to identify cases in which the largest single term M_n among $\{X_1, \ldots, X_n\}$, when $X_i \geq 0$, is (i) asymptotically negligible, (ii) asymptotically of the same order, or (iii) asymptotically dominant compared to the partial sum S_n or its centered version. Work on this problem is reviewed in Section 8.15 of Bingham et al. (1989); we summarize the main results below.

Theorem 4.5.7. (asymptotics for the ratio of the maximum to the sum) *Let* $\{X_n : n \geq 1\}$ *be a sequence of IID random variables with cdf* F *having support on* $(0, \infty)$. *Let* S_n *be the* n^{th} *partial sum and* M_n *the* n^{th} *maximum. Then*

(a) $M_n/S_n \Rightarrow 0$ *if and only if* $\int_0^x y dF(y)$ *is slowly varying;*

(b) $M_n/S_n \Rightarrow 1$ *if and only if* F^c *is slowly varying;*

(c) M_n/S_n *converges in distribution to a nondegenerate limit if and only if* F^c *is regularly varying of index* $-\alpha$ *for some* α, $0 < \alpha < 1$.

(d) If, in addition F *has finite mean* μ, *then* $(S_n - n\mu)/M_n$ *converges in distribution to a nondegenerate limit if and only if* F^c *is regularly varying of index* $-\alpha$ *for some* α, $1 < \alpha < 2$.

4.6. The Joseph Effect: Strong Dependence

In Section 4.4 we saw that the conclusion of Donsker's theorem still holds when the independence condition is replaced with weak dependence, provided that the finite-second-moment condition is maintained. The situation is very different when there is *strong dependence*, also called *long-range dependence*.

In fact, all hell breaks loose. The statistical regularity we have seen, both with light and heavy tails, depends critically on the independence. As we saw in Section 4.4, we can relax the independence considerably, but the results depend on the dependence being suitably controlled. By definition, strong dependence occurs when that control is lost.

When we allow too much dependence, many bizarre things can happen. A simple way to see the possible difficulties is to consider the extreme case in which the random variables X_n are all copies of a single random variable X, where X can have any distribution. Then the scaled partial sum $n^{-1}S_n$ has the law of X, so $n^{-1}S_n \Rightarrow X$. Obviously there is no unifying stochastic-process limit in this degenerate case.

Nevertheless, it is important to study strong dependence, because it can be present. With strong dependence, we need to find some appropriate way to introduce strong structure to replace the independence we are giving up. Fortunately, ways to do this have been discovered, but no doubt many more remain to be discovered. We refer to Beran (1994), Eberlein and Taqqu (1986) and Samorodnitsky and Taqqu (1994) for more discussion and references.

We will discuss two approaches to strong dependence in this section. One is to exploit Gaussian processes. Gaussian processes are highly structured because they are fully characterized by their first and second moments, i.e., the mean function and the covariance function. The other approach is to again exploit independence, but in a modified form.

When we introduce this additional structure, it often becomes possible to establish stochastic-process limits with strong dependence. Just as with the heavy tails considered in Section 4.5, the strong dependence has a dramatic impact on the form of the stochastic-process limits, changing both the scaling and the limit process.

4.6.1. Strong Positive Dependence

Consider a stationary sequence $\{X_n : n \geq 1\}$ with $EX_n = 0$ and $Var\, X_n < \infty$. Since the variance $Var\, X_n$ is assumed to be finite, we call this the light-tailed case; in the next section we consider the heavy-tailed case in which $Var\, X_n = \infty$. Strong dependence can be defined by saying that the natural mixing conditions characterizing weak dependence, as in Theorem 4.4.1, no longer hold. However, motivated by applications, we are interested in a particular form of strong dependence called *strong positive dependence.* Roughly speaking, with positive dependence, we have

$$Var(S_n) > nVar(X_1) \quad \text{for} \quad n > 1 \ ,$$

i.e., the variance of the n^{th} partial sum is greater than it would be in the IID case. We are interested in the case in which this is true for all sufficiently large n (ignoring departures from the assumption in a short time scale).

Even though $Var\, X_n$ is finite, there may be so much dependence among the successive variables X_n that the variance of the partial sum $S_n \equiv X_1 + \cdots + X_n$ is not of order n. Unlike (4.3), we are now primarily interested in the case in which

$$\lim_{n\to\infty} \frac{Var(S_n)}{n} = \infty\,. \tag{6.1}$$

In particular, we assume that $Var(S_n)$ is a regularly varying function with index $2H$ for some H with

$$1/2 < H < 1\,, \tag{6.2}$$

i.e.,

$$Var(S_n) = n^{2H} L(n) \quad \text{as} \quad n \to \infty\,, \tag{6.3}$$

where $L(t)$ is a slowly varying function; see Appendix A. The principal case of interest for applications is $L(t) \to c$ as $t \to \infty$ for some constant c. When (6.2) and (6.3) hold, we say that $\{X_n\}$ and $\{S_n\}$ exhibit *strong positive dependence.* Since $Var(S_n) \le n^2\, Var\, X_1$, (6.2) covers the natural range of possibilities when (6.1) holds. In fact, we allow $0 < H < 1$, which also includes negative dependence.

We primarily characterize and quantify the strong dependence through the asymptotic form of the variance of the partial sums, as in (6.3). However, it is important to realize that we still need to impose additional structure in order to allow us to focus only on these variances. We will impose appropriate structure below.

It is natural to deduce the asymptotic form of the variance $\mathrm{Var}(S_n)$ in (6.3) directly, but we could instead start with a detailed characterization of the covariances between variables in the sequence $\{X_n\}$. We want to complement the weak-dependent case in (4.3) and (4.4), so we focus on the cases with $H \neq 1/2$. We state the result as a lemma; see p. 338 of Samorodnitsky and Taqqu (1994). We state the result for pure power asymptotics, but there is an extension to regularly varying functions.

Lemma 4.6.1. (from covariance asymptotics to variance asymptotics) *Suppose that the covariances have the asymptotic form*

$$r_n \equiv Cov(X_1, X_{1+n}) \equiv E[(X_1 - EX_1)(X_{1+n} - EX_{1+n})] \sim cn^{2H-2}$$

as $n \to \infty$. *If* $c > 0$ *and* $1/2 < H < 1$, *then*

$$Var(S_n) \sim c\frac{n^{2H}}{H(2H-1)} \quad as \quad n \to \infty\,.$$

If $c < 0$ *and* $0 < H < 1/2$, *then*

$$Var(S_n) \sim |c|\frac{n^{2H}}{H(1-2H)} \quad as \quad n \to \infty\,.$$

In this setting with $\mathrm{Var}(X_n) < \infty$ and centering to zero mean, the natural scaled process is

$$\mathbf{S}_n(t) \equiv c_n^{-1} S_{\lfloor nt \rfloor}, \quad t \geq 0\,, \tag{6.4}$$

where

$$c_n \equiv (Var(S_n))^{1/2}\,. \tag{6.5}$$

With the scaling in (6.4), we have

$$E\mathbf{S}_n(t) = 0 \quad \text{and} \quad Var(\mathbf{S}_n(t)) = t, \quad t \geq 0\,.$$

Space scaling asymptotically equivalent to (6.5) is required to get convergence of the second moments to a proper limit. We will find conditions under which $\mathbf{S}_n \Rightarrow \mathbf{S}$ in D and identify the limit process $\mathbf{S}$. Note that the strong positive dependence causes $c_n/\sqrt{n} \to \infty$ as $n \to \infty$.

4.6.2. Additional Structure

We now impose the additional structure needed in order to obtain a FCLT. As indicated above, there are two cases that have been quite well studied. In the first case, $\{X_n\}$ is a zero-mean Gaussian sequence. Then the finite-dimensional distributions are determined by the covariance function. A generalization of this first case in which $X_n = g(Y_n)$, where $\{Y_n\}$ is Gaussian and $g : \mathbb{R} \to \mathbb{R}$ is a smooth nonlinear function, has also been studied, e.g., see Taqqu (1975, 1979) and Dobrushin and Major (1979), but we will not consider that case. It gives some idea of the complex forms possible for limit processes with strong dependence.

In the second case, the basic stationary sequence $\{X_n\}$ has the *linear-process representation*

$$X_n \equiv \sum_{j=0}^{\infty} a_j Y_{n-j}, \quad n \geq 1\,, \tag{6.6}$$

where $\{Y_n : -\infty < n < \infty\}$ is a two-sided sequence of IID random variables with $EY_n = 0$ and $EY_n^2 = 1$, and $\{a_j : j \geq 0\}$ is a sequence of (deterministic, finite) constants with

$$\sum_{j=0}^{\infty} a_j^2 < \infty\,. \tag{6.7}$$

With (6.6), the stochastic process $\{X_n\}$ is said to obtained from the underlying process $\{Y_n\}$ by applying a *linear filter*; e.g., see p. 8 of Box, Jenkins and Reinsel (1994).

In fact, the linear-process representation tends to include the Gaussian sequence as a special case, because if $\{X_n\}$ is a stationary Gaussian process, then under minor regularity conditions, $\{X_n\}$ can be represented as in (6.6), where $\{Y_n\}$ is a sequence of IID random variables distributed as $N(0,1)$; e.g., see Hida and Hitsuda (1976). Of course, in general the random variables in the linear-process representation need not be normally distributed. Thus, the linear-process representation includes the Gaussian sequence as a special case.

The second case can also be generalized by considering variables $g(X_n)$ for X_n in (6.6) and $g : \mathbb{R} \to \mathbb{R}$ a smooth nonlinear function, see Avram and Taqqu (1987) and references there, but we will not consider that generalization either. It provides a large class of stochastic-process limits in a setting where the strong dependence is still quite tightly controlled by the underlying linear-process representation.

It is elementary that $\{X_n\}$ in (6.6) is a stationary process and

$$Var(X_n) = \sum_{j=0}^{\infty} a_j^2\,,$$

so that condition (6.7) ensures that $Var(X_n) < \infty$, as assumed before. It is also easy to determine the covariance function for X_n:

$$r_n = \sum_{j=0}^{\infty} a_j a_{j+n}.$$

The n^{th} partial sum can itself be represented as a weighted sum of the variables from the underlying sequence $\{Y_n\}$, namely,

$$S_n = \sum_{k=-\infty}^{n} Y_k a_{n,k}\ ,$$

where

$$a_{n,k} = \begin{cases} \sum_{j=0}^{n-k} a_j & , \quad 1 \le k \le n, \\ \sum_{j=k}^{n+k} a_j, & \quad k \le 0 . \end{cases}$$

Example 4.6.1. *Power weights.*

Suppose that the weights a_j in (6.6) have the relatively simple form

$$a_j = cj^{-\gamma} . \tag{6.8}$$

To get strong positive dependence with (6.7), we need to require that $1/2 < \gamma < 1$. The associated covariances are

$$r_n = c^2 \sum_{j=0}^{\infty} j^{-\gamma}(j+n)^{-\gamma} .$$

By applying the Euler-Maclaurin formula, Chapter 8 of Olver (1974), and the change of variables $x = nu$, we obtain the asymptotic form of r_n:

$$r_n \sim c^2 \int_0^\infty x^{-\gamma}(x+n)^{-\gamma}\, dx \sim n^{1-2\gamma} c^2 \int_0^\infty u^{-\gamma}(1+u)^{-\gamma}\, du$$

as $n \to \infty$, where

$$\int_0^\infty u^{-\gamma}(1+u)^{-\gamma}\, du = B(1-\gamma, 2\gamma-1) = \Gamma(1-\gamma)\Gamma(2\gamma-1)/\Gamma(\gamma)$$

with $B(z,w)$ and $\Gamma(z)$ the beta and gamma functions; see 6.1.1, 6.2.1 and 6.2.2 of Abramowitz and Stegun (1972). Hence

$$r_n \sim C_1 n^{1-2\gamma} \quad \text{as} \quad n \to \infty , \tag{6.9}$$

where

$$C_1 = c^2\Gamma(1-\gamma)\Gamma(2\gamma-1)/\Gamma(\gamma) . \tag{6.10}$$

By Lemma 4.6.1, $H = (3-2\gamma)/2$ and

$$Var(S_n) \sim C_2 n^{3-2\gamma} \quad \text{as} \quad n \to \infty , \tag{6.11}$$

where

$$C_2 = 2c^2\Gamma(1-\gamma)\Gamma(2\gamma-1)/\Gamma(\gamma)(3-2\gamma)^2 . \tag{6.12}$$

For instance, if $\gamma = 3/4$, then $H = 3/4$ and $C_2 = 41.95c^2$ for c in (6.8). ∎

4.6.3. Convergence to Fractional Brownian Motion

We can deduce that the limit process $\mathbf{S}$ for $\mathbf{S}_n$ in (6.4), with (6.6) holding, must be a Gaussian process. First, if the basic sequence $\{X_n : n \geq 1\}$ is a Gaussian process, then the scaled partial-sum process $\{\mathbf{S}_n(t) : t \geq 0\}$ must also be a Gaussian process for each n, which implies that $\mathbf{S}$ must be Gaussian if $\mathbf{S}_n \Rightarrow \mathbf{S}$. Hence, if a limit holds more generally without the Gaussian condition, then the limit process must be as determined for the special case.

Alternatively, starting from the linear-process representation (6.6) with a general sequence $\{Y_n\}$ of IID random variables with $EY_n = 0$ and $EY_n^2 = 1$, we can apply the central limit theorem for non-identically distributed summands, e.g., as on p. 262 of Feller (1971), and the Cramer-Wold device in Theorem 4.3.3 to deduce that

$$(\mathbf{S}_n(t_1), \ldots, \mathbf{S}_n(t_k)) \Rightarrow (\mathbf{S}(t_1), \ldots, \mathbf{S}(t_k)) \quad \text{in} \quad \mathbb{R}^k$$

for all positive integers k and all k-tuples $(t_1, \ldots, t_k)$ with $0 \leq t_1 < \cdots < t_k$, where $(\mathbf{S}(t_1), \ldots, \mathbf{S}(t_k))$ must have a Gaussian distribution. Thus, weak convergence in D only requires in addition showing tightness.

The limit process in the FCLT is *fractional Brownian motion* (FBM). Standard FBM is the zero-mean Gaussian process $\mathbf{Z}_H \equiv \{\mathbf{Z}_H(t) : t \geq 0\}$ with covariance function

$$r_H(s,t) \equiv Cov(\mathbf{Z}_H(s), \mathbf{Z}_H(t)) \equiv \frac{1}{2}(t^{2H} + s^{2H} - (t-s)^{2H}) \ , \qquad (6.13)$$

where any H with $0 < H < 1$ is allowed. For $H = 1/2$, standard FBM reduces to standard Brownian motion.

Standard FBM can also be expressed as a stochastic integral with respect to standard Brownian motion; in particular,

$$\mathbf{Z}_H(t) = \int_{-\infty}^{t} w_H(t,u) d\mathbf{B}(u) \, , \qquad (6.14)$$

where

$$w_H(t,u) = \begin{cases} 0, & u \geq t, \\ (t-u)^{H-1/2}, & 0 \leq u < t \, , \\ (t-u)^{H-1/2} - (-u)^{H-1/2} & u < 0 \, . \end{cases} \qquad (6.15)$$

Of course, some care is needed in defining the stochastic integral with respect to Brownian motion, because the paths are of unbounded variation, but this problem has been addressed; e.g., see Karatzas and Shreve (1988), Protter (1992), Section 2.4 of Beran (1994) and Chapter 7 of Samorodnitsky and Taqqu (1994).

Note that (6.14) should be consistent with our expectations, given the initial weighted sum in (6.6). From (6.14) we can see how the dependence appears in FBM. We also see that FBM is a smoothed version of BM. For example, from (6.14) it is evident that FBM has continuous sample paths. The process FBM is also H-self-similar, which can be regarded as a consequence of being a weak-convergence limit, as discussed in Section 4.2.

We are now ready to state the FCLT, which is due to Davydov (1970); also see p. 288–289 of Taqqu (1975). Note that the theorem always holds for $1/2 \leq H < 1$, but also holds for $0 < H < 1/2$ under extra moment conditions (in (6.17) below). These extra moment conditions are always satisfied in the Gaussian case. For refinements, see Avram and Taqqu (1987) and references therein.

Theorem 4.6.1. (FCLT for strong dependence and light tails) *Suppose that the basic stationary sequence* $\{X_n : n \geq 1\}$ *is either a zero-mean Gaussian process or a zero-mean linear process as in* (6.6) *and* (6.7) *with* $E[X_n^2] < \infty$. *If*

$$Var(S_n) = c_n^2 \equiv n^{2H} L(n), \quad n \geq 1 \ , \tag{6.16}$$

for $0 < H < 1$, *where* L *is slowly varying, and, in the non-Gaussian case,*

$$E|S_n|^{2a} \leq K(E[S_n^2]^a) \quad \textit{for some} \quad a > 1/H \tag{6.17}$$

for some constant K, *then*

$$\mathbf{S}_n \Rightarrow \mathbf{Z}_H \quad \textit{in} \quad (D, J_1) \tag{6.18}$$

for $\mathbf{S}_n$ *in* (6.4) *with* c_n *in* (6.16) *and* $\mathbf{Z}_H$ *standard FBM with self-similarity index* H.

Remark 4.6.1. *Applying the continuous-mapping approach.* Considering the linear-process representations in (6.6) and (6.14), it is natural to view the limit in (6.18) as convergence of stochastic integrals

$$\int w_n dB_n \to \int w dB \,, \tag{6.19}$$

where the integrands are deterministic, the limiting stochastic integral corresponds to (6.14) and

$$\mathbf{B}_n(t) = n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} Y_i, \quad t \geq 0 \,.$$

Donsker's theorem states that $\mathbf{B}_n \Rightarrow \mathbf{B}$ in D. It remains to show that $w_n \to w$ in a manner so that (6.19) holds. An approach to weak convergence of linear processes along this line is given by Kasahara and Maejima (1986). An earlier paper in this spirit for the special case of discounted processes is Whitt (1972). For more on convergence of stochastic integrals, see Kurtz and Protter (1991) and Jakubowski (1996). The point of this remark is that Theorem 4.6.1 should properly be viewed as a consequence of Donsker's FCLT and the continuous-mapping approach. ∎

The linear-process representation in (6.6) is convenient mathematically to impose structure, because we have constructed the stationary sequence $\{X_n\}$ from an underlying sequence of IID random variables with finite second moments, which we know how to analyze. What may not be evident, however, is that the linear-process representation can arise naturally from modelling. We show that it can arise naturally from time-series modeling in Section 2.5 of the Internet Supplement.

In Chapter 1, the random-walk simulations suggested stochastic-process limits. Having already proved convergence to FBM, we now can use the stochastic-process limits to provide a way to simulate FBM.

Example 4.6.2. *Simulating FBM.* We can simulate FBM, or more properly an approximation of FBM, by simulating a random walk $\{S_n\}$ with steps X_n satisfying the linear-process representation in (6.6), where $\{Y_n\}$ is IID with mean 0 and variance 1. We will let $Y_i \stackrel{\mathrm{d}}{=} N(0,1)$. As part of the approximation, we truncate the series in (6.6). That can be done by assuming that $a_j = 0$ for $j \geq N$, where N is suitably large.

As in Chapter 1, the plotter does the appropriate space scaling automatically. In order to verify that what we see is consistent with the theory, we calculate the appropriate space-scaling constants. To be able to do so conveniently, we use the power weights in Example 4.6.1 with $c = 1$ and $\gamma = 3/4$. As indicated there, then the self-similarity index is $H = 3/4$, $\mathrm{Var}(S_n) \sim 41.95n^{2H}$ and the space-scaling constants are

$$c_n = \sqrt{Var(S_n)} = 6.477n^{3/4} \ .$$

We plot S_k for $0 \leq k \leq n$ for $n = 10^2$ and $n = 10^3$ in Figures 4.1 and 4.2. We plot four independent replications in each case. In these examples, we let $N = 10^4$. We use smaller n than in the IID case, because the computation is more complex, tending to require work of order nN. Comparing Figure 4.1 to Figures 1.3 and 1.4, we see that the sample paths of FBM are smoother than the paths of BM, as we should anticipate from (6.14).

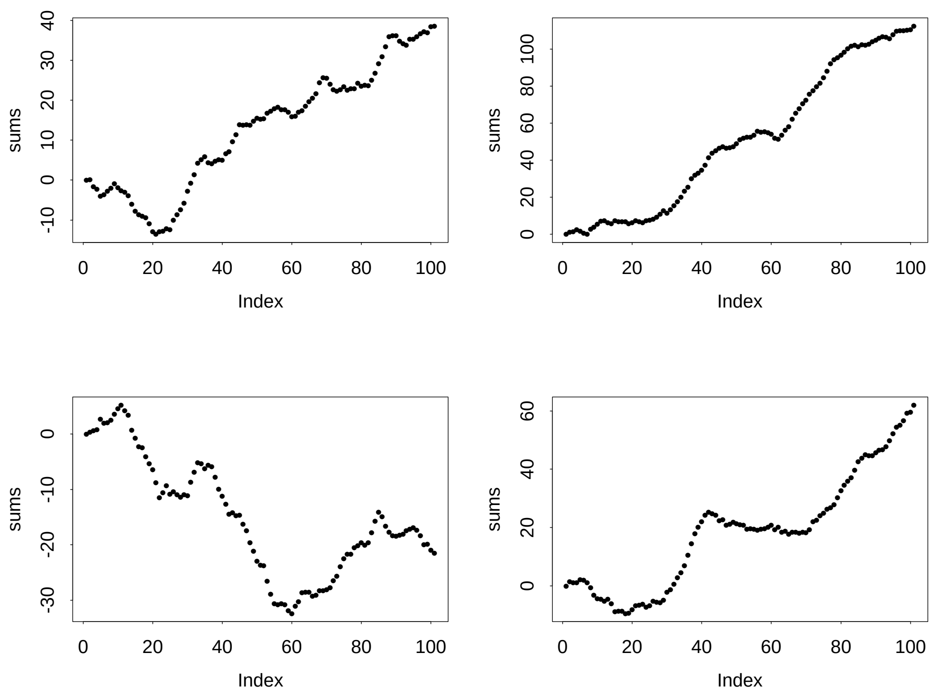

Figure 4.1: Four independent realizations of the first 10^2 steps of the unscaled random walk $\{S_k : 0 \leq k \leq n\}$ associated with the strongly dependent steps in Example 4.6.1.

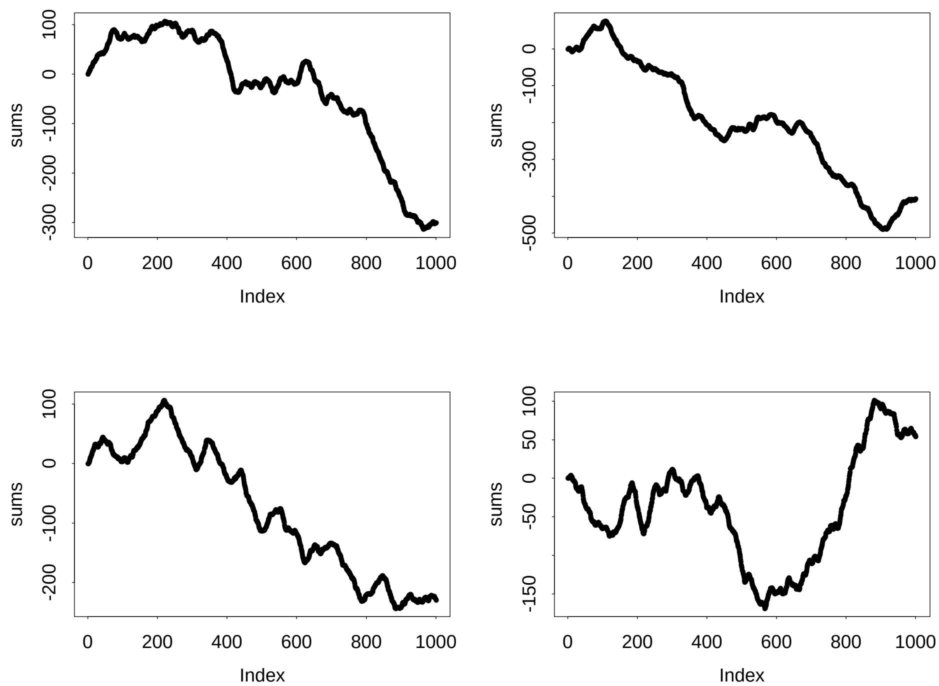

Figure 4.2: Four independent realizations of the first 10^3 steps of the unscaled random walk $\{S_k : 0 \le k \le n\}$ associated with the strongly dependent steps in Example 4.6.1.

As in Chapter 1, we can see emerging statistical regularity by considering successively larger values of n. The plots tend to look the same as n increases. However, as with the heavy-tailed distributions (the Noah effect), there is more variability from sample path to sample path than in the IID light-tailed case, as depicted in Figures 1.3 and 1.4. Even though the steps have mean 0, the strong dependence often make the plots look like the steps have non-zero mean. These sample paths show that it would be impossible to distinguish between strong dependence and nonstationarity from only a modest amount of data, e.g., from only a few sample paths like those in Figures 4.1 and 4.2.

The standard deviations of S_n for $n = 100$ and $n = 1,000$ are 205 and 1152, respectively. That is consistent with the final positions seen in Figures 4.1 and 4.2.

Since it is difficult to simulate the random walk S_n with dependent steps X_n, it is natural to seek more efficient methods to simulate FBM. For discussion of alternative methods, see pp. 370, 588 of Samorodnitsky and Taqqu (1994).

■

The strong dependence poses a difficulty because of increased variability. The increased variability is indicated by the growth rate of $Var(S_n)$ as $n \to \infty$. However, the strong dependence also has a positive aspect, providing an opportunity for better *prediction*.

Remark 4.6.2. *Exploiting dependence for prediction.* The strong dependence helps to exploit observations of the past to predict process values in the not-too-distant future. To illustrate, suppose that we have a linear process as in (6.6), and that as time evolves we learn the values of the underlying sequence Y_n, so that after observing X_n and S_n we know the variables Y_j for $j \leq n$. From (6.6), the conditional means and variances are

$$E[X_{n+k}|Y_j, j \leq n] = \sum_{j=0}^{\infty} a_{k+j} Y_{n-j} \ , \tag{6.20}$$

$$Var(X_{n+k}|Y_j, j \leq n) \equiv E[(X_{n+k} - E[X_{n+k}|Y_j, j \leq n])^2] = \sum_{j=0}^{k-1} a_j^2 \ , \tag{6.21}$$

$$E[S_{n+k}|Y_j, j \leq n] = \sum_{j=0}^{\infty} (\sum_{i=j+1}^{j+k} a_i) Y_{n-j} \tag{6.22}$$

and

$$Var(S_{n+k}|Y_j, j \le n) \equiv E[(S_{n+k} - E[S_{n+k}|Y_j, j \le n])^2] = \sum_{j=0}^{k-1} (\sum_{i=0}^{j} a_i)^2 \ .$$

If we use a criterion of mean-squared error, then the conditional mean is the best possible predictor of the true mean and the conditional variance is the resulting mean-squared error. A similar analysis applies to FBM, assuming that we learn the history of the underlying Brownian motion in the linear-process representation in (6.14). However, in many applications we can only directly observe the past of the sequence $\{X_n\}$, or the FBM $Z_H(t)$ in case of the limit process. Fortunately, prediction can still be done by expoliting time-series methods. We discuss prediction in queues in Remark 8.7.2. ■

In some applications (e.g., at the end of Section 7.2 below) we will want continuous-time analogs of Theorem 4.6.1. With continuous-time processes, we need to work harder to establish tightness. We show how this can be done for Gaussian processes with continuous sample paths.

Theorem 4.6.2. (FCLT for Gaussian processes in C) *If $\{Y(t) : t \ge 0\}$ is a zero-mean Gaussian process with stationary increments, sample paths in C, $Y(0) = 0$,*

$$VarY(t) \sim ct^{2H} \quad as \quad t \to \infty \tag{6.23}$$

and

$$VarY(t) \le Kt^{2H} \quad for\ all \quad t \ge 0 \tag{6.24}$$

for some constants c, K and H with $1/2 < H < 1$, then

$$\mathbf{Z}_n \Rightarrow c\mathbf{Z}_H \quad in \quad (C, U) \ ,$$

where $\mathbf{Z}_H$ is standard FBM and

$$\mathbf{Z}_n(t) \equiv n^{-H} Y(nt), \quad t \ge 0 \ .$$

Proof. For each n, $\mathbf{Z}_n$ is a Gaussian process. Given (6.23), it is elementary that $cov(\mathbf{Z}_n(s), \mathbf{Z}_n(t)) \to cov(\mathbf{Z}(s), \mathbf{Z}(t))$ as $n \to \infty$ for all s and t. That establishes convergence of the finite-dimensional distributions. By (6.24),

$$E[(\mathbf{Z}_n(t) - \mathbf{Z}_n(s))^2] = n^{-2H} VarY(n(t-s)) \le K(t-s)^{2H} \ ,$$

which implies tightness by Theorem 11.6.5. ■

4.7. Heavy Tails Plus Dependence

The previous three sections described FCLTs with only heavy tails (Section 4.5) and with only dependence (Sections 4.4 and 4.6). The most complicated case involves both heavy tails and dependence. Unfortunately, there is not yet a well developed theory for stochastic-process limits in this case. Evidently, a significant part of the difficulty stems from the need to use nonstandard topologies on the function space D; e.g., see Avram and Taqqu (1992) and Jakubowski (1996). Hence, this interesting case provides additional motivation for the present book, but it remains to establish important new results.

We start by considering the natural analog of Section 4.4 to the case of heavy tails: stable limits with weak dependence. Since the random variables do not have finite variances, even describing dependence is complicated, because the covariance function is not well defined. However, alternatives to the covariance have been developed; see Samorodnitsky and Taqqu (1994). We understand weak dependence to hold when there is dependence but the stochastic-process limit is essentially the same as in the IID case.

We state one result for stable limits with weak dependence. It is a FCLT for linear processes with heavy tails. However, there is a significant complication caused by having dependence together with jumps in the limit process. To obtain a stochastic-process limit in D, it is necessary to use the M_1 topology on D. Moreover, even with the M_1 topology, it is necessary to impose additional conditions in order to establish the FCLT.

4.7.1. Additional Structure

Just as in the last section, in this section we assume that the basic sequence $\{X_n : n \geq 1\}$ is a stationary sequence with a linear-process representation

$$X_n \equiv \sum_{j=0}^{\infty} a_j Y_{n-j} \,, \tag{7.1}$$

where the *innovation process* $\{Y_n : -\infty < n < \infty\}$ is a sequence of IID random variables, but now we assume that Y_n has a heavy-tailed distribution. In particular, we assume that the distribution of Y_n is in the domain of attraction of a stable law $S_\alpha(1, \beta, 0)$ with $0 < \alpha < 2$; i.e., we assume that (5.21) and (5.22) hold. That in turn implies that $Var(Y_n) = \infty$.

Given the stable index α, we assume that

$$\sum_{j=0}^{\infty} |a_j|^{\alpha-\epsilon} < \infty \quad \text{for some} \quad \epsilon > 0\,. \tag{7.2}$$

Condition (7.2) ensures that the sum (7.1) converges in the L^p space for $p = \alpha - \epsilon$ and w.p.1; see Avram and Taqqu (1992). However, the variance $Var(X_n)$ is necessarily infinite.

We first remark that condition (7.2) permits quite strong dependence, because we can have

$$a_j \sim cj^{-\gamma} \quad \text{as} \quad j \to \infty \quad \text{for} \quad \text{any} \quad \gamma > \alpha^{-1}\,, \tag{7.3}$$

where c is a positive constant, so we might have $\sum_{j=1}^{\infty} |a_j| = \infty$.

For simplicity, we assume that $EY_n = 0$ if $1 < \alpha < 2$ and that the distribution of Y_n is symmetric if $\alpha = 1$. Then, under the assumptions above, Theorems 4.5.1 and 4.5.3 imply that

$$\mathbf{S}_n \Rightarrow \mathbf{S} \quad \text{in} \quad (D, J_1)\,, \tag{7.4}$$

where

$$\mathbf{S}_n(t) \equiv c_n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} Y_i, \quad t \geq 0\,, \tag{7.5}$$

$\mathbf{S}$ is a stable process with $\mathbf{S}(1) \stackrel{\mathrm{d}}{=} S_\alpha(1, \beta, 0)$ and $c_n = n^{1/\alpha} L(n)$ for some slowly varying function L. We are interested in associated FCLTs for

$$\mathbf{Z}_n(t) \equiv c_n^{-1} \sum_{i=1}^{\lfloor nt \rfloor} X_i, \quad t \geq 0\,, \tag{7.6}$$

for $\{X_n\}$ in (7.1) and $\{c_n : n \geq 1\}$ in (7.5).

4.7.2. Convergence to Stable Lévy Motion

In considerable generality, $\mathbf{Z}_n$ in (7.6) satisfies essentially the same FCLT as $\mathbf{S}_n$ in (7.5), with the limit being a constant multiple of the previous limit $\mathbf{S}$. The following result is from Astrauskas (1983), Davis and Resnick (1985) and Avram and Taqqu (1992). Note that the M_1 topology is used. Let $Z_n \Rightarrow Z \quad \text{in} \quad (D, f.d.d.)$ mean that there is convergence of all finite-dimensional distributions.

Theorem 4.7.1. (FCLT for a linear process with heavy tails) *Suppose that the sequence* $\{X_n\}$ *is the linear process in* (7.1) *satisfying the assumptions above, which imply* (7.4). *If, in addition,*

$$\sum_{j=0}^{\infty} |a_j| < \infty\,, \tag{7.7}$$

then

$$\mathbf{Z}_n \Rightarrow (\sum_{j=0}^{\infty} a_j)\mathbf{S} \quad in \quad (D, f.d.d.)$$

for $\mathbf{S}$ *in* (7.4) *and* $\mathbf{Z}_n$ *in* (7.6). *Suppose, in addition, that* $a_i \geq 0$ *for all* i. *If any one of the following conditions hold:*

(i) $0 < \alpha \leq 1$,

(ii) $a_i \neq 0$ *for only finitely many* i,

(iii) $\alpha > 1$, $\sum_{i=1}^{\infty} |a_i|^{\nu} < \infty$ *for some* $\nu < 1$ *and* $\{a_i\}$ *is a monotone sequence,*

then

$$\mathbf{Z}_n \Rightarrow (\sum_{j=0}^{\infty} a_j)\mathbf{S} \quad in \quad (D, M_1)\ .$$

Avram and Taqqu (1992) actually established the M_1-convergence part of Theorem 4.7.1 under a somewhat weaker condition than stated above. Avram and Taqqu (1992) show that the M_1 topology is critical in Theorem 4.7.1; the result does not hold in (D, J_1) if there are at least two nonzero coefficients in (7.1). Indeed, that is evident because an exceptionally large value of Y_n will correspond to more than one exceptionally large value in the X_n; i.e., the jump in the limit process for $\mathbf{Z}_n$ will correspond to more than one jump in the converging processes. The linear-process structure is yet another setting leading to unmatched jumps in the limit process, requiring the M_1 topology instead of the familiar J_1 topology.

Note that the limit process in Theorem 4.7.1 has independent increments. Thus, just as in Section 4.4, the dependence in the original process is asymptotically negligible in the time scaling of the stochastic-process limit. Thus, the predicted value of $S_{\lfloor cn \rfloor}$ for $c > 1$ given $S_j, j \leq n$, is about S_n. At that time scale, there is not much opportunity to exploit past observations, beyond the present value, in order to predict future values.

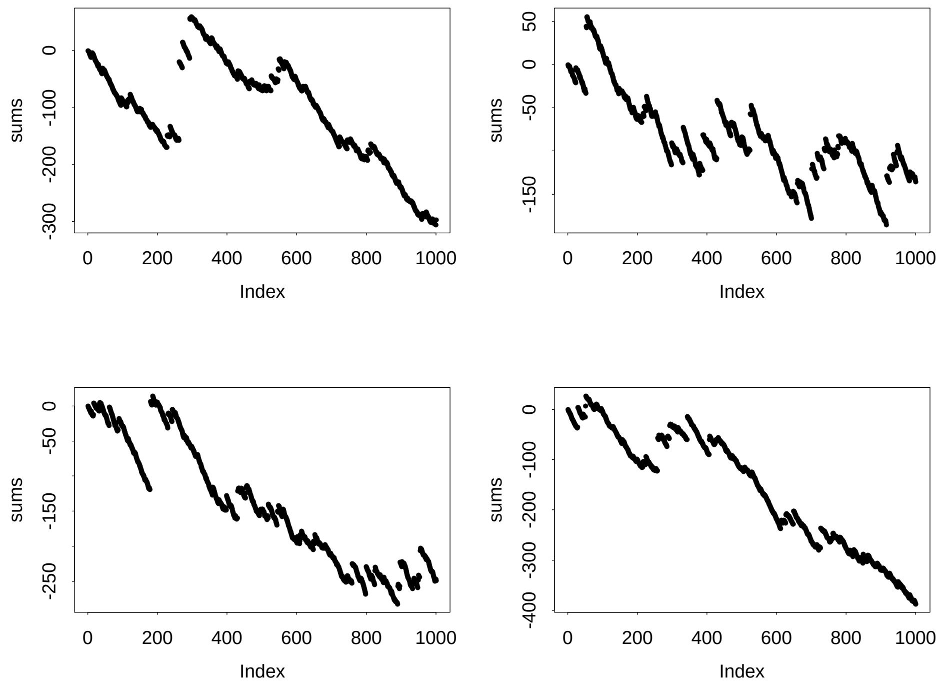

Figure 4.3: Four independent realizations of the first 10^3 steps of the unscaled random walk $\{S_k : 0 \leq k \leq n\}$ associated with the dependent heavy-tailed steps in Example 4.7.1.

Example 4.7.1. *Simulation to experience Theorem 4.7.1.* To illustrate Theorem 4.7.1, suppose that Y_1 has the Pareto(p) distribution with $p = 3/2$, just as in Section 1.4. Let the weights be $a_j = j^{-8}$ for $j \geq 0$. We simulate the random walk just as in Example 4.6.2. Since the weights decay faster here, it suffices to use a smaller truncation point N; we use $N = 100$. We plot four independent replications of the random walk $\{S_k : 0 \leq k \leq n\}$ for $n = 1,000$ in Figure 4.3. The plots look just like the plots of the random walk in the IID case in Figures 1.20, 1.21 and 1.22. Thus the simulation is consistent with Theorem 4.7.1. ■

4.7.3. Linear Fractional Stable Motion

Note that the conditions of Theorem 4.7.1 do not cover the case in which

$$a_j \sim c j^{-\gamma} \quad \text{as} \quad j \to \infty \tag{7.8}$$

for $c > 0$ and $\alpha^{-1} < \gamma \le 1$, where $1 < \alpha < 2$. We include $\gamma > \alpha^{-1}$ in (7.8) so that condition (7.2) is still satisfied, but condition (7.7) is violated. We refer to this case as strong positive dependence with heavy tails.

The limit process when (7.8) holds is *linear fractional stable motion* (LFSM), which is an H-sssi α-stable process with self-similarity index

$$H = \alpha^{-1} + 1 - \gamma > \alpha^{-1} \ , \tag{7.9}$$

where $1 < \alpha < 2$, so that $2^{-1} < \alpha^{-1} < 1$, and $H < 1$; see Sections 7.3 and 7.4 of Samorodnitsky and Taqqu (1994).

Paralleling the representation of FBM as a stochastic integral with respect to standard Brownian motion in (6.14), we can represent LFSM as a stochastic integral with respect to stable Lévy motion; in particular, for $1 < \alpha < 2$ and $\alpha^{-1} < H < 1$,

$$\mathbf{Z}_{H,\alpha}(t) = \int_{-\infty}^{t} w_H(t,u) d\mathbf{S}_\alpha(u) \, , \tag{7.10}$$

where $\mathbf{S}_\alpha$ is an α-stable Lévy motion with $\mathbf{S}_\alpha(1) \stackrel{\mathrm{d}}{=} S_\alpha(1, \beta, 0)$ and

$$w_H(t,u) = \begin{cases} 0, & u \ge t, \\ (t-u)^{H-1/\alpha}, & 0 \le u < t \, , \\ (t-u)^{H-1/\alpha} - (-u)^{H-1/\alpha} & u < 0 \, ; \end{cases} \tag{7.11}$$

The LFSM in (7.10) is natural because $\mathbf{Z}_{H,\alpha}(t)$ depends upon $\mathbf{S}_\alpha$ only over the interval $(-\infty, t]$ for any t, so that we can regard $\mathbf{S}_\alpha$ as an innovation process. For more general LFSMs, see Samorodnitsky and Taqqu (1994). It is significant that the LFSM above has continuous sample paths; see Theorem 12.4.1 of Samorodnitsky and Taqqu (1994).

Theorem 4.7.2. (FCLT with both the Noah and Joseph effects) *Suppose that the basic sequence $\{X_n\}$ has the linear-process representation* (7.1), *where $\{Y_n\}$ is a sequence of IID random variables with Y_1 in the normal domain of attraction of the stable law $S_\alpha(1, \beta, 0)$ i.e., such that* (5.26) *and* (5.27) *hold. If, in addition,* (7.8) *holds, then*

$$\mathbf{Z}_n \Rightarrow \mathbf{Z}_{H,\alpha} \quad in \quad (D, J_1) \ ,$$

where $\mathbf{Z}_{H,\alpha}$ *is LFSM in* (7.10) *and* $\mathbf{Z}_n$ *is the scaled partial-sum process in* (7.6) *with space-scaling constants*

$$c_n = n^H (A/C_\alpha)^{1/\alpha}(c/(1-\gamma)), \quad n \geq 1 \tag{7.12}$$

for A *in* (5.26), C_α *in* (5.14), (c, γ) *in* (7.8) *and the self-similarity index* H *in* (7.9).

By Theorem 4.5.2, under the assumptions in Theorem 4.7.2, the space-scaling constants for the partial sums of Y_n are $c_n = (nA/C_\alpha)^{1/\alpha}$. From (7.12), we see that the linear-process representation produces the extra multiplicative factor $n^{H-\alpha^{-1}}c(1-\gamma)^{-1}$.

We remark that Astrauskas (1983) actually proved a more general result, allowing both the tail probability $P(|Y_1| > x)$ and the weights a_j to be regularly varying at infinity instead of pure power tails. For extensions of Theorems 4.7.1 and 4.7.2, see Hsing (1999) and references therein.

Example 4.7.2. *Simulating LFSM.*

To illustrate Theorem 4.7.2, suppose that Y_1 has the Pareto(p) distribution with $p = 3/2$, just as in Example 4.7.1, but now let the weights be $a_j = j^{-\gamma}$ for $\gamma = 3/4$, just as in Example 4.6.2. Hence we have combined the heavy-tailed feature of Example 4.7.1 with the strong-dependence feature in Example 4.6.2. Since $\alpha = p$,

$$\gamma > \alpha^{-1} = 2/3$$

and (7.8) is satisfied. From (7.9), the self-similarity index in this example is

$$H = \alpha^{-1} + 1 - \gamma = 11/12 ,$$

so that H is much greater than $1/2$.

We simulate the random walk just as in Example 4.7.1, except that we let the truncation point N be higher because of the more slowly decaying weights; in particular, now we let $N = 1,000$. We plot four independent replications of the random walk $\{S_k : 0 \leq k \leq n\}$ for $n = 1,000$ in Figure 4.4.

Unlike the plots in Figure 4.3, it is evident from Figure 4.4 that the sample paths are now continuous. However, the heavy tails plus strong dependence can induce strong surges up and down. The steady downward trend in the first plot occurs because there are relatively few larger values. The sudden steep upward surge at about $j = 420$ in the fourth plot in

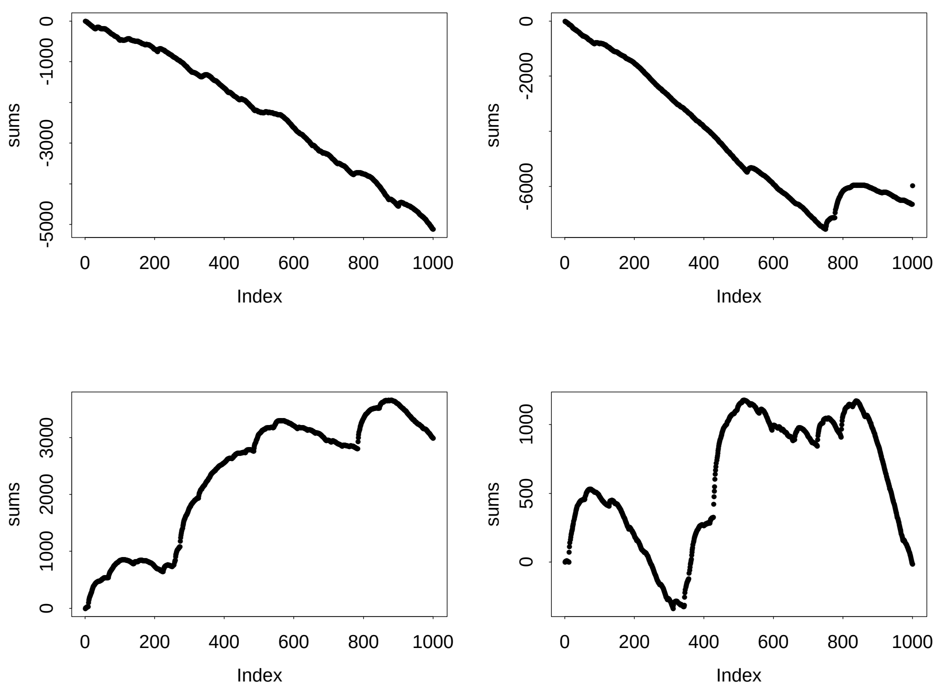

Figure 4.4: Four independent realizations of the first 10^3 steps of the unscaled random walk $\{S_k : 0 \leq k \leq n\}$ associated with the strongly dependent steps in Example 4.7.2.

Figure 4.4 occurs because of a few exceptionally large values at that point. In particular, $Y_{416} = 17.0$, $Y_{426} = 88.1$ and $Y_{430} = 24.3$. In contrast, in the corresponding plot of FBM in Figure 4.2 with the same weights $a_j = j^{-3/4}$, all $2,000$ of the normally distributed Y_j satisfy $|Y_j| \leq 3.2$. Finally, note that the large value of H is consistent with the large observed values of the range in the plots. ■

From the dependent increments in the LFSM limit process, it is evident that there is again (as in Section 4.6) an opportunity to exploit the history of past observations in order to predict future values of the process. With strong dependence plus heavy-tailed distributions, the statistical techniques are more complicated, but there is a growing literature; see Samorodnitsky and Taqqu (1994), Kokoszka and Taqqu (1995, 1996a,b), Montanari, Rosso and Taqqu (1997), Embrechts, Klüppelberg and Mikosch (1997) and Adler, Feldman and Taqqu (1998).

4.8. Summary

We have now presented FCLTs for partial sums in each of the four cases – light or heavy tails with weak or strong dependence. We summarize the results in the table below.

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Dependence</td></tr>
<tr><td>Weak</td><td>Strong
Joseph effect</td></tr>
<tr><td rowspan="2">Tails</td><td>light</td><td>Sections 4.3 and 4.4</td><td>Section 4.6</td></tr>
<tr><td>heavy

Noah
effect</td><td>Section 4.5
Theorem 4.7.1</td><td>Theorem 4.7.2</td></tr>
</table>

Table 4.1: The four kinds of FCLTs established in Sections 4.3–4.7

In conlusion, we observe that the theory seems far from final form for the strong dependence discussed in Sections 4.6 and 4.7 and for heavy tails with any form of dependence. The results should be regarded as illustrative of what is possible. Careful study of specific applications is likely to unearth important new limit processes.

We next show how the continuous-mapping approach can be applied with established stochastic-process limits to establish heavy-traffic stochastic-process limits for queues. In Chapter 7 we present additional established stochastic-process limits.