Chapter 2

Stochastic-Process Limits

2.1. Introduction

Chapters 4 and 7 of the book present a panorama of stochastic-process limits. In this chapter we present even more material. In Section 2.2 we present an introduction to strong approximations and the rates of convergence in the setting of Donsker's theorem that they imply using the Prohorov metric. In Section 2.3 we present additional Brownian limits under weak dependence; here we focus on Markov and regenerative structure.

In Section 2.4 we briefly discuss the convergence to general Lévy processes that holds when we have a sequence of random walks (based on a double sequence of random walk steps). Finally, in Section 2.5 we point out that the linear-process representation assumed with strong dependence in Sections 4.6 and 4.7 of the book arises naturally from modelling when we take a time-series perspective.

2.2. Strong Approximations and Rates of Convergence

In Sections 1.4 and 4.3 of the book we noted that the CLT and FCLT are invariance principles, meaning that the same limits occur in great generality. In the IID case we only need the summands X_n to have finite variance. However, the quality of the approximation for any given n is affected by the distribution of X_n. Indeed, that is obvious for the CLT: If $X_n \stackrel{\mathrm{d}}{=} N(0, \sigma^2)$, then the limit can be replaced by equality in distribution. Moreover, the closer the distribution of X_n is to the normal distribution, the better the normal approximation for the scaled partial sum should be. More generally, the advantage of extra structure in the distribution of X_n can be seen from more refined results giving bounds on the rate of convergence and asymptotic expansions. We review some of these results in this section.

2.2.1. Rates of Convergence in the CLT

A bound on the rate of convergence in the basic CLT, given a finite third absolute moment of a summand, is provided by the Berry-Esseen theorem; see p. 542 of Feller (1971). To state it, we use the uniform metric on cdf's, defined by

$$\|F_1 - F_2\| \equiv \sup_x |F_1(x) - F_2(x)| \; . \tag{2.1}$$

As before, let Φ be the standard normal cdf.

Theorem 2.2.1. (Berry-Esseen theorem) *Let $\{X_n\}$ be a sequence of IID random variables with $EX_1 = 0$, $E[X_1^2] = \sigma^2$ and $E[|X_1|^3] = \delta_3 < \infty$. Then*

$$\|F_n - \Phi\| \le 3\delta_3/\sigma^3\sqrt{n} \quad \text{for all} \quad n \; ,$$

where $F_n(x) \equiv P((n\sigma^2)^{-1/2}(X_1 + \cdots + X_n) \le x)$.

Theorem 2.2.1 implies that for given n and σ^2, the bound on the distances decreases as the third absolute moment δ_3 decreases. We now describe the Edgeworth expansion, which shows how further regularity conditions can improve the quality of the normal approximation; see p. 535 of Feller (1971). We also get convergence of pdf's.

Theorem 2.2.2. (Edgeworth expansion) *If, in addition to the assumptions of Theorem 2.2.1 above, moments $E[X_1^k]$ exist for $3 \le k \le r$ and $|E[exp(itX_1)|^\nu$ is integrable for some $\nu \ge 1$, then $(n\sigma^2)^{-1/2}(X_1 + \cdots + X_n)$ has a pdf f_n for all n and*

$$f_n(x) = n(x)[1 + \sum_{k=3}^{r} n^{-(k-2)/2} P_k(x) + o(n^{-(r-2)/2})]$$

as $n \to \infty$, uniformly in x, where n is the standard normal pdf and $P_k(x)$ is a real polynomial depending on the first k moments of X_1, with the property that $P_k(x) = 0$ if the first k moments of X_1 agree with those of the standard normal distribution.

Note that the rate of convergence in Theorem 2.2.2 is $O(n^{-1/2})$ if $E[X_1^3] \neq 0$, but is $O(n^{-1})$ or better if $E[X_1^3] = 0$. When $E[X_1^3] \neq 0$, the refinement provided by the second term can be useful.

2.2.2. Rates of Convergence in the FCLT

We now turn to Donsker's FCLT. From the Lipschitz mapping theorem, Theorem 3.4.2 in the book, we can deduce a bound on the rate of convergence in the CLT from a bound on a rate of convergence in the FCLT. Hence, we can see in advance that the rate of convergence in the FCLT, given a finite third absolute moment, can be no better than the $O(n^{-1/2})$ bound provided by the Berry-Esseen theorem. In fact, the best possible bound for the FCLT, under an even stronger regularity condition, is somewhat worse, being larger by a factor of $\log n$. From a practical perspective, though, the difference is not great.

We now give the final rate-of-convergence result, expressed in terms of the Prohorov metric π from Section 3.2 of the book; see (2.2) here. For this application, it is convenient to let the underlying function space be the set $D_Q \equiv D_Q([0,1],\mathbb{R})$ of functions in $D \equiv D([0,1],\mathbb{R})$ with discontinuities only at rational points in the domain $[0,1]$, endowed with the uniform metric $\|\cdot\|$; we refer to the space as (D_Q,U). The space (D_Q,U) is a separable metric space and the stochastic processes considered here all have sample paths in this space. Thus, the Prohorov metric π is defined on the space $\mathcal{P}((D_Q,U))$, the space of all probability measures on (D_Q,U). Since

$$d_{M_1}(x_1,x_2) \le d_{J_1}(x_1,x_2) \le \|x_1 - x_2\| \quad \text{for} \quad x_1, x_2 \in D \ ,$$

the result also holds for the spaces (D, d_{J_1}) and (D, d_{M_1}).

The following combines Theorems 1.16 and 1.17 in Csörgő and Horváth (1993).

Theorem 2.2.3. (bounds on the rate of convergence in Donsker's FCLT) *Let $\{X_n\}$ be a sequence of IID random variables with $EX_n = 0$ and $E[X_1^2] = \sigma^2$. If, in addition, $E[exp(tX_1)] < \infty$ for t in a neighborhood of the origin, then there exist positive constants C_1 and C_2 such that*

$$C_1 \log n/\sqrt{n} \le \pi(\mathbf{S}_n, \sigma\mathbf{B}) \le C_2 \log n/\sqrt{n} \tag{2.2}$$

for all n, where π is the Prohorov metric on the space $\mathcal{P}((D_Q,U))$, $\mathbf{B}$ is standard Brownian motion and $\mathbf{S}_n(t) \equiv n^{-1/2}S_{\lfloor nt \rfloor}$, $0 \le t \le 1$. If, instead, only $E[|X_1|^p < \infty$ for some $p > 2$, then there is a constant C such that

$$\pi(\mathbf{S}_n, \sigma\mathbf{B}) \le Cn^{-(p-2)/2(p+1)} \tag{2.3}$$

for all n. Moreover, for any sequence $\{a_n\}$ with $a_n \to \infty$ as $n \to \infty$, there is a random variable X_1 with $E[|X_i|^p] < \infty$ such that

$$\overline{\lim_{n\to\infty}} \ a_n n^{(p-2)/2(p+1)} \pi(\mathbf{S}_n, \sigma\mathbf{B}) = \infty \ . \tag{2.4}$$

The lower bound in (2.2) and the limit in (2.4) show that the upper bounds in Theorem 2.2.3 are indeed best possible. Note that the rate $O(\log n/\sqrt{n})$ in (2.2) exceeds the Berry-Esseen bound $O(1/\sqrt{n})$ by a factor of $\log n$. We regard that difference as negligible.

However, there is a big difference between the bounds in (2.3) and in Theorem 2.2.2. When there is only a finite third absolute moment, we have (2.3) with $p = 3$, which only yields the rate $O(n^{-1/8})$. For finite p^{th} moment with $p > 2$, (2.3) gives a rate that can be substantially worse than $O(n^{-1/2})$, while Theorem 2.2.2 gives rates that can be much better than $O(n^{-1/2})$. It should be recognized that the conditions are quite different though.

By the Lipschitz mapping theorem, Theorem 3.4.2 of the book, the rate of convergence in Theorem 2.2.3 is inherited by Lipschitz functions. For real-valued Lipschitz functions, we then can obtain bounds on the uniform metric for cdf's.

Corollary 2.2.1. (bounds on the uniform metric for cdf's of the images of real-valued Lipschitz maps) *Suppose that* $g : (D_Q, U) \to \mathbb{R}$ *is a Lipschitz function and that* $g(\mathbf{B})$ *has a bounded pdf. If the conditions of Theorem 2.2.3 hold with* $E exp(tX_1) < \infty$ *for* t *in a neighborhood of the origin, then there is a positive constant* C *such that*

$$\sup_x |P(g(\mathbf{S}_n) \leq x) - P(g(\sigma \mathbf{B}) \leq x)| \leq C \log n/\sqrt{n} \tag{2.5}$$

for all $n \geq 1$.

We can apply Corollary 2.2.1 to obtain a bound on the rate of convergence in the CLT; we use the projection map $\pi_1(x) \equiv x(1)$, which is easily seen to be Lipschitz. However, the bound is not as good as provided by the Berry-Esseen theorem, so the bound may no longer be best possible when we consider the image measure associated with a single Lipschitz map.

We can also apply Theorem 2.2.3 to establish bounds on the rate of convergence in heavy-traffic FCLTs for queues. We illustrate by stating a result for the queueing model in Section 1.6. We use the fact that the two-sided reflection map $\phi_K : D \to D$ is Lipschitz; see Theorem 13.10.1. An early result of this kind is Kennedy (1973). That served as motivation for the Lipschitz mapping theorem in Whitt (1974).

Corollary 2.2.2. (bounds on the rate of convergence in a heavy-traffic stochastic-process limit for queues) *Consider the queueing model in Section* 2.3 *of the book with IID inputs* V_k *with mean* m_v *and variance* σ^2.

If, in addition, $K_n = n^{1/2}K$ *and* $\mu_n = m_v + mn^{-1/2}$ *for all* n *and with* $E[exp(tV_1)] < \infty$ *for some* $t > 0$*, then there exists a constant* C *such that*

$$\pi(\mathbf{W}_n, \phi_K(\sigma\mathbf{B} - m\mathbf{e})) \leq C \log n / n^{1/2} \ ,$$

where $\mathbf{W}_n$ *is the scaled workload process in equation* (2.3.6) *of the book and* ϕ_K *is the two-sided reflection map.*

2.2.3. Strong Approximations

Theorem 2.2.3 can be extablished by applying *strong approximations.* Like the Skorohod and Strassen representation theorems in Chapters 3 and 11 of the book, strong approximations are special constructions of random objects on the same underlying probability space, often called couplings; see Lindvall (1992).

We start by stating the Komlós, Major and Tusnády (1975, 1976) strong approximation theorems for partial sums of IID random variables; see Chapter 2 of Csörgő and Révész (1981) and Chapter 1 of Csörgő and Horváth (1993). See Philipp and Stout (1975) for extensions to the weakly dependent case and Einmahl (1989) for extensions to the multivariate case. See Csörgő and Horvath (1993) for strong approximations of renewal processes and random sums. For applications of strong approximations to queues, see Zhang et al. (1990), Horváth (1990), Glynn and Whitt (1991a,b) and Chen and Mandelbaum

Theorem 2.2.4. (strong approximation with finite moment generating function) *Let* $\{X_n : n \geq 1\}$ *be a sequence of IID random variables with* $EX_1 = 0$*,* $EX_1^2 = 1$ *and* $Ee^{tX_1} < \infty$ *for* t *in a neighborhood of the origin. Let* $S_n \equiv X_1 + \cdots + X_n$*,* $n \geq 1$*, with* $S_0 \equiv 0$*. Then there exists a standard Brownian motion* $\mathbf{B} \equiv \{\mathbf{B}(t) : t \geq 0\}$ *such that, for all real* x *and every* $n \geq 1$*,*

$$P\left(\max_{1\leq k\leq n} |S_k - \mathbf{B}(k)| > C_1 \log n + x\right) < C_2 e^{-\lambda x} \ , \tag{2.6}$$

where C_1*,* C_2 *and* λ *are positive constants depending upon the distribution of* X_1*.*

As a consequence of Theorem 2.2.4, we can deduce that

$$S_n - \mathbf{B}(n) = O(\log n) \quad \text{as} \quad n \to \infty \quad \text{w.p.1} \ ; \tag{2.7}$$

i.e., there is a constant C such that

$$P(|S_n - \mathbf{B}(n)| > C \log n \quad \text{infinitely often}) = 0 \ . \tag{2.8}$$

Note that (2.8) follows from (2.6) by substituting $C' \log n$ for x in (2.6) for suitably large C' and then applying the Borel-Cantelli theorem.

We now relax the extra condition on the tail of the ccdf $P(|X_1| > t)$, at the expenses of obtaining a slower rate.

Theorem 2.2.5. (strong approximation with p^{th} moment) *Let $\{X_n : n \geq 1\}$ be a sequence of IID random variables with $EX_1 = 0$, $EX_1^2 = 1$ and $E|X_1|^p < \infty$ for some $p > 2$. Let $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, with $S_0 \equiv 0$. Then there exists a standard Brownian motion $\mathbf{B}$ such that*

$$n^{-1/p}|S_n - \mathbf{B}(n)| \to 0 \quad \text{w.p.1} \tag{2.9}$$

To apply Theorems 2.2.4 and 2.2.5 to establish Theorem 2.2.3, we need to relate Brownian motion $\mathbf{B}$ to the associated processes

$$\mathbf{B}_n^1(t) \equiv n^{-1/2}\mathbf{B}(\lfloor nt \rfloor), \quad \mathbf{B}_n^2(t) \equiv \mathbf{B}(\lfloor nt \rfloor/n), \quad \mathbf{B}_n^3(t) \equiv n^{-1/2}\mathbf{B}(nt)$$

for $0 \leq t \leq 1$. By the self-similarity property, $\mathbf{B} \stackrel{\text{d}}{=} \mathbf{B}_n^3$ and $\mathbf{B}_n^1 \stackrel{\text{d}}{=} \mathbf{B}_n^2$ for all $n \geq 1$. We can relate $\mathbf{B}_n^2$ to $\mathbf{B}$ by bounding the fluctuations of Brownian motion. The following is Lemma 1.1.1 of Csörgő and Révész (1981).

Theorem 2.2.6. (uniform bound on the fluctuations of Brownian motion) *For any $\epsilon > 0$, there exists a constant $C = C(\epsilon)$ such that*

$$P\Big(\sup_{0 \leq t \leq T-h}\ \sup_{0 \leq s \leq h} |\mathbf{B}(t+s) - \mathbf{B}(t)| \geq \nu\sqrt{h}\Big) \leq (CT/h)exp(-\nu^2/(2+\epsilon)) \tag{2.10}$$

for all positive ν, T, and h, $0 < h < T$.

Theorem 2.2.6 can be applied to determine the precise modulus of continuity of Brownian sample paths (originally determined by Lévy); see Theorem 1.1 of Csörgő and Révész (1981).

Theorem 2.2.7. (modulus of continuity of Brownian paths) *If $\mathbf{B}$ is Brownian motion, then*

$$\lim_{h \to 0}\ \sup_{0 \leq s \leq 1}\ \sup_{0 \leq t \leq h} \frac{|\mathbf{B}(s+t) - \mathbf{B}(s)|}{\sqrt{2h \log h^{-1}}} = 1 \quad \textit{w.p.1} \ .$$

From Theorem 2.2.7, we see that the sample paths of Brownian motion are continuous but not differentiable; the largest increment of length h is almost surely of order $O(\sqrt{2h \log h^{-1}})$. We can also apply Theorem 2.2.6 to determine the following bound on the in-probability distance $p(\mathbf{B}, \mathbf{B}_n^2)$ and the Prohorov distance $\pi(\mathbf{B}, \mathbf{B}_n^1)$, where π is defined on the space $\mathcal{P}((C, U))$.

Corollary 2.2.3. *There exists a constant* C_1 *such that*

$$\pi(\mathbf{B}, \mathbf{B}_n^1) \leq p(\mathbf{B}, \mathbf{B}_n^2) \leq C_1 \sqrt{\log n / n}$$

for all $n \geq 1$.

Proof. The first inequality holds because $\mathbf{B}_n^1 \stackrel{\mathrm{d}}{=} \mathbf{B}_n^2$ and $\pi \leq p$. For the second inequality, let $\nu = \sqrt{c \log n}$ for $c > 4$ in (2.10). Then the right hand side of (2.10) for $T = 1$ becomes $C' n^{-(1+\delta)}$ for $\delta > 0$ and constant C'. ■

Partial proof of Theorem 2.2.3. For the upper bound in (2.1), let $x = C_3 \log n$ in (2.6) to obtain

$$\pi(\mathbf{S}_n, \mathbf{B}_n^1) \leq p(\mathbf{S}_n, \mathbf{B}_n^1) \leq C \log n / \sqrt{n} \ .$$

Then use the triangle inequality with Corollary 2.2.3. ■

Theorem 2.2.4 can be applied to obtain a strong approximation for a Lévy process, i.e., a random element of D with stationary and independent increments; see Corollary 5.5 on p. 359 of Ethier and Kurtz (1986).

Theorem 2.2.8. (strong approximation for a Lévy process) *Let* $\{\mathbf{L}(t) : t \geq 0\}$ *be a real-valued Lévy process. Assume that*

$$Ee^{\alpha \mathbf{L}(1)} < \infty \tag{2.11}$$

for all α *with* $|\alpha| \leq \alpha_0$ *for some* $\alpha_0 > 0$. *Then there exist versions of the Lévy process* $\mathbf{L}$ *and a standard Brownian motion* $\mathbf{B}$ *on a common probability space such that*

$$|\mathbf{L}(t) - mt - \sigma \mathbf{B}(t)| = O(\log t) \quad \text{as} \quad t \to \infty \quad \text{w.p.1} \ , \tag{2.12}$$

where $m = E\mathbf{L}(1)$ *and* $\sigma^2 = Var\, \mathbf{L}(1)$.

A precursor to the strong approximation theorems, of interest in its own right, is the Skorohod (1961) embedding theorem; see p. 88 of Csörgő and Révész (1981).

Theorem 2.2.9. (Skorohod embedding theorem) *Let* $\{X_n : n \geq 1\}$ *be a sequence of IID real-valued random variables with* $EX_1 = 0$ *and* $EX_1^2 = 1$. *Let* $S_n \equiv X_1 + \cdots + X_n$, $n \geq 1$, *with* $S_0 \equiv 0$. *There exists a probability space supporting a standard Brownian motion* $\mathbf{B}$ *and a sequence* $\{T_n : n \geq 1\}$ *of nonnegative IID random variables such that* *(i)* $\{\mathbf{B}(T_1 + \cdots T_n) : n \geq 1\} \stackrel{d}{=} \{S_n : n \geq 1\}$ *in* $\mathbb{R}^\infty$*;*

(ii) $\{T_1 + \cdots + T_n : n \geq 1\}$ *is a sequence of stopping times, i.e., the event* $\{T_1 + \cdots + T_n \leq t\}$ *is contained in the* σ*-field generated by* $\{\mathbf{B}(s) : 0 \leq s \leq t\}$ *for all* $t \geq 0$*;*

(iii) $ET_1 = 1$*;*

(iv) $ET_1^k < \infty$ *if, in addition,* $EX^{2k} < \infty$ *for positive integer* k*.*

As a consequence of Theorem 2.2.9,

$$\begin{aligned}\{n^{-1/2}S_{\lfloor nt \rfloor} : t \geq 0\} &\stackrel{d}{=} \{n^{-1/2}B(T_1 + \cdots + T_{\lfloor nt \rfloor}) : t \geq 0\} \\ &\stackrel{d}{=} \{B(n^{-1}(T_1 + \cdots + T_{\lfloor nt \rfloor}) : t \geq 0\} \ .\end{aligned}$$

By the FSLLN,

$$\sup_{0 \leq t \leq u} |n^{-1}(T_1 + \cdots + T_{\lfloor nt \rfloor}) - t| \to 0 \quad \text{w.p.1} \ ,$$

so that Donsker's theorem again is a consequence. Rate of convergence results follow too.

2.3. Weak Dependence from Regenerative Structure

This section is a sequel to Section 4.4 in the book, in which we showed that many Brownian limits still hold for random walks $\{S_n : n \geq 0\}$ when the IID condition on the sequence of steps $\{X_n : n \geq 1\}$ is relaxed, with the finite-second-moment condition $EX_n^2 < \infty$ remaining in place. We now obtain results for stochastic-processes with regenerative structure.

This new setting allows us to abandon the assumption of stationarity and obtain explicit expressions for the asymptotic variance σ^2, defined by

$$\sigma^2 \equiv \lim_{n \to \infty} \frac{Var(S_n)}{n} \ . \tag{3.1}$$

For a stationary sequence $\{X_n\}$, the asymptotic variance has the representation

$$\sigma^2 = Var\, X_n + 2\sum_{k=1}^{\infty} Cov(X_1, X_{1+k}) \ . \tag{3.2}$$

We now obtain more explicit representations for the asymptotic variance in terms of basic model elements.

2.3.1. Discrete-Time Markov Chains

We start by stating results for finite-state Markov chains. We first consider discrete-time chains and then we consider continuous-time chains. Afterwards, we state results for general regenerative processes, which cover more general Markov processes and non-Markov processes. The first result for DTMC's extends Theorem 4.4.2 in the book. An important point is that an explicit expression can be given for the asymptotic variance σ^2. It is expressible as a function of the fundamental matrix of the DTMC. The most effective way to calculate the asymptotic variance is usually to solve a system of equations, collectively known as the *Poisson equation*.

Let P be the transition matrix of an irreducible k-state DTMC and let Π be a matrix with each row being the steady-state vector π. (We will work with row vectors; let A^t be the transpose of a matrix A, so that the column vector associated with a row vector x is x^t.) Then the *fundamental matrix* of the DTMC is

$$Z \equiv (I - P + \Pi)^{-1} ; \tag{3.3}$$

see pp. 75, 100 of Kemeny and Snell (1960). (The matrix $I - P + \Pi$ is nonsingular.)

Theorem 2.3.1. (FCLT for a DTMC with explicit asymptotic variance) *Let $\{Y_n : n \geq 1\}$ be an irreducible k-state DTMC and let $X_n = f(Y_n)$ for a real-valued function f. Then the FCLT*

$$\mathbf{S}_n \Rightarrow \sigma \mathbf{B} \quad in \quad (D, J_1) , \tag{3.4}$$

where $\mathbf{B}$ is standard Brownian motion and

$$\mathbf{S}_n(t) = n^{-1/2}(S_{\lfloor nt \rfloor} - mnt), \quad t \geq 0 , \tag{3.5}$$

holds with

$$m \equiv \sum_{i=1}^{k} \pi_i f(i) ,$$

$$\sigma^2 \equiv 2 \sum_{i=1}^{k} \sum_{j=1}^{k} (f(i) - m) \pi_i Z_{i,j} (f(j) - m) - \sum_{i=1}^{k} \pi_i (f(i) - m)^2 , \tag{3.6}$$

π the steady-state vector and $Z \equiv (Z_{i,j})$ the fundamental matrix in (3.3).

As a quick sanity check on (3.6), note that in the IID case we have $P = A$, $Z = I$ and, from (3.6),

$$\sigma^2 = \sum_{i=1}^{k} \pi_i (f(i) - m)^2 \ ,$$

as we should.

It is significant that we can calculate π, m, Z and σ^2 in Theorem 2.3.1 by solving the Poisson equation(s). We state both row-vector and column-vector versions. Let $\mathbf{1} \equiv (1, \ldots, 1)$ be a vector of 1's and $\mathbf{0} \equiv (0, \ldots, 0)$ be a vector of 0's.

Theorem 2.3.2. (Poisson equations for a DTMC) *Consider an irreducible finite-state DTMC with transition matrix P. The row-vector version of the Poisson equation*

$$x(I - P) = y \tag{3.7}$$

has a solution x for given y if and only if $y\mathbf{1}^t = 0$. All solutions to (3.7) *are of the form*

$$x = yZ + (x\mathbf{1}^t)\pi \ .$$

The column-vector version of the Poisson equation

$$(I - P)x^t = y^t \tag{3.8}$$

has a solution x^t for given y^t if and only if $\pi y^t = 0$. All solutions to (3.8) *are of the form*

$$x^t = Zy^t + (\pi x^t)\mathbf{1} \ .$$

Proof. We consider only the row-vector form. Clearly $y\mathbf{1}^t = 0$ is necessary, because $(I - P)\mathbf{1}^t = 0^t$. Given (3.7),

$$x(I - P + \Pi) = y + (x\mathbf{1}^t)\pi \ ,$$

but $I - P + \Pi$ is nonsingular with inverse Z, so that

$$x = yZ + (x\mathbf{1}^t)\pi Z = yZ + (x1^t)\pi$$

since $\pi Z = Z$. ∎

Theorem 2.3.3. (Poisson equations for the steady-state vector and the asymptotic variance of a DTMC) *For an irreducible finite-state DTMC, the steady-state vector π is the unique solution x to the Poisson equation* (3.7) *with* $y = (0, \ldots, 0)$ *and* $x\mathbf{1}^t = 1$. *The asymptotic variance can be expressed as*

$$\sigma^2 = 2\sum_{i=1}^{k} x_i(f(i) - m)$$

where m *is the mean and* x *solves the Poisson equation* (3.7) *with*

$$y_i = (f(i) - m)\pi_i, \quad 1 \le i \le k \ .$$

2.3.2. Continuous-Time Markov Chains

We now turn to the continuous-time processes. There are analogs of the DTMC results in Theorems 2.3.1–2.3.3 for CTMC's. Let $\{(Y(t) : t \ge 0\}$ be an irreducible k-state CTMC. Then the limit is for the integral

$$S(t) \equiv \int_0^t f(Y(s))ds, \quad t \ge 0 \ .$$

The associated normalized processes in D are

$$\mathbf{S}_n(t) \equiv n^{-1/2}(S(nt) - mnt), \quad t \ge 0 \ . \tag{3.9}$$

Given transition matrices $P(t) \equiv (P_{i,j}(t))$, where

$$P_{i,j}(t) \equiv P(Y(t) = j | Y(0) = i) \ ,$$

the *infinitesimal generator matrix* of the CTMC is $Q \equiv (Q_{i,j})$ where

$$Q \equiv \lim_{t \downarrow 0}(P(t) - I)$$

and the *fundamental matrix* is $Z \equiv (Z_{i,j})$ where

$$Z_{i,j} \equiv \int_0^\infty (P_{i,j}(t) - \pi_j)dt$$

and

$$Z = (\Pi - Q)^{-1} - \Pi \tag{3.10}$$

see Kemeny and Snell (1961) and Whitt (1992). A CTMC model is usually specified by giving the infinitesimal generator matrix Q. For an irreducible finite-state CTMC, the steady-state vector π is the unique vector with sum 1 that satisfies

$$\pi Q = 0 \ .$$

Paralleling (3.1) and (3.2) above, the asymptotic variance in this continuous-time framework is

$$\sigma^2 \equiv \lim_{t\to\infty} \frac{Var(S(t))}{t} = 2\int_0^\infty r(t)dt \ ,$$

where $r(t)$ is the *(auto) covariance function*, i.e.,

$$r(t) \equiv E[X(0)X(t)] - (E[X(0)])^2$$

for $X(t) \equiv f(Y(t))$, $t \geq 0$.

The following is the continuous-time analog of Theorem 2.3.1.

Theorem 2.3.4. (FCLT for a CTMC with explicit asymptotic variance) *Let $\{Y(t) : t \geq 0\}$ be an irreducible k-state CTMC, and let $X(t) = f(Y(t))$ for a real-valued function f. Then the FCLT* (3.4) *holds for $\mathbf{S}_n$ in* (3.9) *with m the steady-state mean and σ^2 the asymptotic variance, which can be expressed as*

$$\sigma^2 \equiv 2\sum_{i=1}^{k}\sum_{j=1}^{k} f(i)\pi_i Z_{i,j} f(j) \ ,$$

where Z is the fundamental matrix in (3.10).

We can calculate π, m, Z and σ^2 by solving Poisson equations for CTMC's; see Whitt (1992). The following is the continuous-time analog of Theorem 2.3.2.

Theorem 2.3.5. (Poisson equations for a CTMC) *Consider an irreducible finite-state CTMC with infinitesimal generator matrix Q. The row-vector version of the Poisson equation*

$$xQ = y \tag{3.11}$$

has a solution x for given y if and only if $y\mathbf{1}^t = 0$. All solutions to (3.11) *are of the form*

$$x = -yZ + (x\mathbf{1}^t)\pi \ .$$

The column-vector version of the Poisson equation

$$Qx^t = y^t$$

has a solution x^t for given y^t if and only if $\pi y^t = 0$. All solutions are of the form

$$x^t = -Zy^t + (\pi x^t)\mathbf{1}^t \ .$$

The following is the continuous-time analog of Theorem 2.3.3.

Theorem 2.3.6. (Poisson equations for the steady-state vector and the asymptotic variance of a CTMC) *For an irreducible finite-state CTMC, the steady-state vector π is the unique solution x to the Poisson equation* (3.11) *with $y = (0, \ldots, 0)$ and $x\mathbf{1}^t = 1$. The asymptotic variance can be expressed as*

$$\sigma^2 = 2\sum_{i=1}^{k} x_i f_i \ ,$$

where x is the unique solution to the Poisson equation (3.11) *with*

$$y_i = (f_i - m)\pi_i \quad \text{and} \quad \sum_{i=1}^{k} x_i = 0 \ .$$

and m is the mean.

We can also obtain even more explicit expressions for the asymptotic variance in Markov chains with additional structure. For example, suppose that the CTMC $\{Y(t) : t \geq 0\}$ is a birth-and-death processes on the integers $\{0, 1, \ldots, n\}$ with positive birth rates λ_i, death rates μ_i and stationary probabilities

$$\pi_i = \frac{\pi_0 \lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \ . \tag{3.12}$$

If the process is irreducible, then the process must be reflecting at 0 and n; i.e., $\lambda_n = \mu_0 = 0$.) The following is Proposition 1 of Whitt (1992). Corresponding results for diffusion processes are also stated there.

Theorem 2.3.7. (asymptotic variance of a birth-and-death process) *Suppose that $X(t) = f(Y(t))$, where f is a real-valued function and $\{Y(t) : t \geq 0\}$ is an irreducible birth-and-death process on the integers $\{0, 1, \ldots, n\}$ with birth rates λ_i and death rates μ_i. Then the asymptotic variance can be expressed as*

$$\sigma^2 = 2\sum_{j=0}^{n-1} (\lambda_j \pi_j)^{-1} \left[\sum_{i=0}^{j} (f(i) - m)\pi_i \right]^2$$

for m the mean and π in (3.12) *above.*

We now state a corollary of Theorem 2.3.7 for an elementary queueing model – the $M/M/1$ queue. The queue-length process in an $M/M/1$ queue is a birth-and-death process with $\lambda_i = \lambda$ and $\mu_i = \mu$ when positive. The following would properly be a corollary to Theorem 2.3.7 except for the fact that the state space is infinite. Extensions to countably infinite and more general state spaces are covered by the results for regenerative processes below.

Corollary 2.3.1. (asymptotic variance for the queue-length process in the M/M/1 queue) *For the queue-length (number in system) process in the M/M/1 queue with traffic intensity $\rho \equiv \lambda/\mu < 1$, the asymptotic variance is*

$$\sigma^2 = \frac{2\rho(1+\rho)}{(1-\rho)^4} \ . \tag{3.13}$$

The $(1-\rho)^4$ term in the denominator of (3.13) shows that very long simulation runs are required to directly estimate the steady-state mean of the queue-length process by the sample mean when ρ is close to its upper limit 1. That insight is important for related models for which we do not already know the steady-state distribution, so that simulation is actually needed. We discuss applications of stochastic-process limits to obtain insights about simulation in Section 5.9 of the book.

For a birth-and-death process it is also possible, and usually preferable, to recursively solve the Poisson equation, see Remarks 1, 2 and 5 of Whitt (1992). For more on Poisson equations, see Glynn (1994) and Glynn and Meyn (1996).

2.3.3. Regenerative FCLT

Donsker's theorem itself applies quite directly when we have regenerative structure, as in the case of DTMC's and CTMC's in Theorem 2.3.1 and 2.3.4 above. For this discussion, we use the classical definition of regenerative process, meaning that the process splits into IID cycles; see p. 125 of Asmussen (1987). We will present the result in continuous time, following Glynn and Whitt (1993), but corresponding results hold in discrete time, as in Glynn and Whitt (1987). An earlier related Markov chain FCLT is due to Maigret (1978).

Consider a stochastic process $\{Y(t) : t \geq 0\}$ with general state space and a measurable real-valued function f on that state space. We assume that the stochastic process $\{Y(t) : t \geq 0\}$ is regenerative with respect to regeneration times T_i satisfying

$$0 \leq T_0 < T_1 < \cdots$$

with $T_{-1} \equiv 0$. We focus on the associated *cumulative process*

$$C(t) \equiv \int_0^t f(Y(s))ds, \quad t \geq 0 , \tag{3.14}$$

and consider the associated normalized processes

$$\mathbf{C}_n(t) \equiv n^{-1/2}(C(nt) - mnt), \quad t \geq 0 \tag{3.15}$$

where m is a real number yet to be specified. The key random variables associated with the regenerative cycles are

$$\begin{aligned} \tau_i &\equiv T_i - T_{i-1} , \\ X_i &\equiv X_i(m) \equiv \int_{T_{i-1}}^{T_i} [f(Y(u)) - m]du , \\ Z_i &\equiv Z_i(m) \equiv \sup_{0 \leq s \leq \tau_i} \left| \int_0^s [f(Y(T_{i-1} + u)) - m]du \right| . \end{aligned} \tag{3.16}$$

By *regenerative structure* we mean that the three-tuples (τ_i, X_i, Z_i) are IID for $i \geq 1$. We also assume that $E\tau_i < \infty$ and

$$\int_0^t |f(Y(s))|ds < \infty \text{ w.p.1} \quad \text{for each} \quad t ,$$

which implies that the cumulative process has continuous sample paths w.p.1.

The general idea is that the cumulative process C in (3.14) is approximately equal to a random sum. In particular,

$$C(t) = S_{N(t)} + R_1(t) + R_2(t), \quad t \geq 0 ,$$

where

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

for X_i in (3.16) with $S_0 \equiv 0$, $N \equiv \{N(t) : t \geq 0\}$ is the (possibly delayed) renewal counting process associated with the regeneration times, i.e.,

$$N(t) \equiv \max\{i : T_i \leq t\}, \quad t \geq 0 ,$$

and $R_i \equiv \{R_i(t) : t \geq 0\}$ are remainder processes, defined by

$$R_1(t) \equiv \int_0^{\min\{t,T_0\}} f(Y(s))ds \tag{3.17}$$

and

$$R_2(t) = \int_{T_{N(t)}}^{t} f(Y(s))ds, \quad t \geq 0 . \tag{3.18}$$

Since $E\tau_1 < \infty$, we have

$$t^{-1}N(t) \to \lambda \equiv 1/E\tau_1, \quad \text{as} \quad t \to \infty \quad \text{w.p.1} . \tag{3.19}$$

Under (3.19), FCLTs for partial sums tend to extend to random sums, as we see in Chapter 13 of the book. The major difficulty here is treating the two remainder terms in (3.17). Since $|R_1(t)| \leq Z_0$, the first remainder term in (3.17) is easily dispensed with in limit theorems. The second remainder term is more complicated; the key bound is

$$|R_2(t)| \leq Z_{N(t)+1}, \quad t \geq 0 .$$

Then we observe that $\{R_2(t) : t \geq 0\}$ is tight without space scaling. Thus, after space scaling, it is asymptotically neglible.

Theorem 2.3.8. (FCLT for regenerative processes) *With the regenerative structure above, there is convergence in distribution*

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

for $\mathbf{C}_n$ *in* (3.15) *and* $\mathbf{B}$ *standard BM if and only if there is a constant* m *such that*

$$EX_1(m) = 0, \quad EX_1(m)^2 < \infty$$

and

$$t^2 P(Z_1(m) > t) \to 0 \quad \textit{as} \quad t \to \infty . \tag{3.20}$$

for $X_1(m)$ *and* $Z_1(m)$ *in* (3.16). *Then the asymptotic variance is*

$$\sigma^2 = EX_1(m)^2 .$$

A sufficient condition for the regularity condition (3.20) is $EZ_1(m)^{2+\epsilon} < \infty$ for some $\epsilon > 0$. (A finite second moment is not enough. We remark that condition (3.20) does not appear in the ordinary CLT; see Glynn and Whitt (1993, 2000).) The role of the regularity condition (3.20) can be understood from the following lemma.

Lemma 2.3.1. (condition for the scaled maximum to be asymptotically negligible) *Let* $\{Z_i : i \geq 1\}$ *be a sequence of IID real-valued random variables and let* $\psi : R_+ \to R_+$ *be a function such that* $\psi(t) \to \infty$ *as* $t \to \infty$. *Then*

$$\psi(n)^{-1} \max_{1 \leq i \leq n} \{|Z_i|\} \Rightarrow 0 \quad \textit{as} \quad n \to \infty$$

if and only if

$$tP(|Z_1| > \epsilon\psi(t)) \to 0 \quad as \quad t \to \infty \quad for\ all \quad \epsilon > 0\,. \tag{3.21}$$

Proof. Let $M_n \equiv \max\{|Z_i| : 1 \le i \le n\}$ and $F(t) \equiv P(|Z_1| \le t)$, $t \ge 0$. Note that $\psi(n)^{-1}M_n \Rightarrow 0$ if and only if, for all $\epsilon > 0$, $P(\psi(n)^{-1}M_n > \epsilon) \to 0$ as $n \to \infty$. However,

$$P(M_n > \epsilon\psi(n)) < \delta$$

if and only if

$$P(M_n) \le \epsilon\psi(n)) \ge 1 - \delta\ ,$$

where

$$\begin{aligned} P(M_n \le \epsilon\psi(n)) &= F(\epsilon\psi(n))^n \\ &= (1 - n^{-1}n(1 - F(\epsilon\psi(n)))^n \\ &= (1 - n^{-1}nF^c(\epsilon\psi(n)))^n \\ &\to 1 \quad \text{as} \quad n \to \infty \end{aligned} \tag{3.22}$$

if and only if

$$nF^c(\epsilon\psi(n)) \to 0 \quad \text{as} \quad n \to \infty$$

or, equivalently, (3.21). ∎

Corollary 2.3.2. *If the conditions of Lemma* 2.3.1 *hold with* $\psi(t) = t^\alpha$ *for* $\alpha > 0$*, then condition* (3.21) *is equivalent to*

$$t^{1/\alpha}P(|Z_1| > t) \to 0 \quad as \quad t \to \infty\,.$$

Proof. Under the assumption, condition (3.21) becomes

$$tP(|Z_1| > \epsilon t^\alpha) \to 0 \quad \text{as} \quad t \to \infty \quad \text{for all} \quad \epsilon > 0\,,$$

which first is equivalent to

$$\epsilon^\alpha(\epsilon^{-\alpha}t)P(|Z_1| > (\epsilon^{-\alpha}t)^\alpha)$$

and then is equivalent to

$$\epsilon^\alpha tP(|Z_1|) > t^\alpha) \to 0 \quad \text{as} \quad t \to \infty \quad \text{for all} \quad \epsilon > 0\,,$$

which in turn is equivalent to the stated result. ∎

A general application of Theorem 2.3.8 is to obtain a FCLT for the counting processes associated with a batch Markovian arrival process (BMAP) as in Lucantoni (1993) or, equivalently, the virtual Markovian point process in Neuts (1989). An explicit formula for the variance of the number of arrivals in $[0, t]$ in a BMAP, from which the asymptotic variance easily can be obtained, is given on p. 284 of Neuts (1989).

2.3.4. Martingale FCLT

Martingale FCLTs are versatile tools for many applications. We have stated one martingale FCLT in Theorem 4.4.4 of the book, but there are others. We conclude this section by stating another. It is Theorem 18.1 of Billingsley (1999).

We start with the double sequence $\{X_{n,i} : n \geq 1, i \geq 1\}$ and an associated double sequence of σ-fields $\{\mathcal{F}_{n,k} : n \geq 1, k \geq 1\}$. We assume that $X_{n,k}$ is a *martingale difference* with respect to these σ-fields, i.e., $X_{n,k}$ is $\mathcal{F}_{n,k}$-measurable and

$$E[X_{n,k}|\mathcal{F}_{n,k-1}] = 0 \quad \text{for all} \quad n \quad \text{and} \quad k\,.$$

Suppose that $EX_{n,k}^2 < \infty$ and put

$$V_{n,k} \equiv E[X_{n,k}^2|\mathcal{F}_{n,k-1}]\,. \tag{3.23}$$

Note that $V_{n,k}$, being a conditional expectation, is a random variable. If the martingale is originally defined only for $1 \leq k \leq k_n$, let $X_{n,k} = 0$ and $\mathcal{F}_{n,k} = \mathcal{F}_{n,k_n}$ for $k > n$. Assume that $\sum_{k=1}^{\infty} X_{n,k}$ and $\sum_{k=1}^{\infty} V_{n,k}$ converge w.p.1 for each n.

Theorem 2.3.9. (martingale FCLT) *If, in addition to the assumptions above,*

$$\sum_{k=1}^{\lfloor nt \rfloor} V_{n,k} \Rightarrow \sigma^2 t \quad \textit{as} \quad n \to \infty \quad \textit{for every} \quad t > 0 \tag{3.24}$$

with $V_{n,k}$ in (3.23) and the Lindeberg condition

$$\sum_{k=1}^{\lfloor nt \rfloor} E[X_{n,k}^2 I_{\{|X_{n,k}| \geq \epsilon\}}] \to 0 \quad \textit{as} \quad n \to \infty$$

holds for every $t > 0$ and $\epsilon > 0$, then

$$\mathbf{S}_n \Rightarrow \sigma \mathbf{B} \quad \textit{in} \quad D\,,$$

where σ is determined by (3.24),

$$\mathbf{S}_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} X_{n,k}, \quad t \geq 0\,,$$

and $\mathbf{B}$ is standard Brownian motion.

Generalizations and other variations of Theorem 2.3.9 are contained on p. 339 of Ethier and Kurtz (1986) and Jacod and Shiryaev (1987).

2.4. Double Sequences and Lévy Limits

We have seen that there are only a few possible limits for normalized partial-sum processes with weak dependence when we work in the framework of a single sequence $\{X_n : n \geq 1\}$. In addition to the Brownian motion limits discussed in Sections 4.3 and 4.4 of the book, there are the stable Lévy motion limits discussed in Sections 4.5 and 4.7 of the book. However, there are many more possible limits for normalized partial-sum processes with weak dependence when we work in the framework of a double sequence $\{X_{n,k} : n \geq 1, k \geq 1\}$. We give a brief account in this section.

Throughout this section we assume that the sequence $\{X_{n,k} : k \geq 1\}$ is IID for each n, so that we are in a classic well-studied setting; e.g., see Gnedenko and Kolmogorov (1968) and Feller (1971). Since there is a different sequence for each n, we can incorporate multiplicative and additive normalization constants directly in the variables $X_{n,k}$. Hence we focus on the partial sums

$$S_{n,n} \equiv X_{n,1} + \cdots + X_{n,n} \tag{4.1}$$

without further normalization and the associated random functions in D defined by

$$\mathbf{S}_n(t) \equiv S_{n,\lfloor nt \rfloor}, \quad t \geq 0 . \tag{4.2}$$

The class of limits processes in FCLTs for $\mathbf{S}_n$ now are all Lévy processes. As indicated in Section 4.5 of the book, a *Lévy process* $\mathbf{L} \equiv \{\mathbf{L}(t) : t \geq 0\}$ is a stochastic process with sample paths in $D \equiv D([0,\infty), \mathbb{R})$, $\mathbf{L}(0) = 0$ and stationary and independent increments. Brownian motion and stable Lévy motion are important examples of Lévy processes, but there are many more; see Bertoin (1996) and Jacod and Shiryaev (1987).

The distribution of $\mathbf{L}(t)$ for any t is an infinitely divisible distribution. A probability distribution is *infinitely divisible* if for each n it is the n-fold convolution of another probability distribution; i.e., a random variable X has an infinitely divisible distribution if, for all n, there are IID random variables $X_1, \ldots, X_n$ (depending upon X and n) such that

$$X \stackrel{\mathrm{d}}{=} X_1 + \cdots + X_n .$$

Lévy processes and infinitely divisible distributions are characterized by their characteristic functions. In particular, the one-dimensional marginal distribution of every Lévy process has characteristic function

$$Ee^{i\theta L(t)} = e^{t\psi(\theta)} ,$$

where the *Lévy exponent* $\psi(\theta)$ can be expressed as

$$\psi(\theta) = ib\theta - \frac{\sigma^2\theta^2}{2} + \int_{-\infty}^{\infty} (\exp(i\theta x) - 1 - i\theta h(x))\mu(dx) \ , \qquad (4.3)$$

with b being the *centering coefficient*, $\sigma^2 \geq 0$ is the *Gaussian coefficient*, μ the *Lévy measure* and h a *truncation function*. There is quite a lot of freedom in the choice of the truncation function h. Following Jacod and Shiryaev (1987, pp. 75) we assume that the truncation function has compact support, is bounded and coincides with x in a neighborhood of the origin. To characterize convergence, we also want h to be continuous. A truncation function with all these properties is

$$h(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2 - x, & 1 \leq x \leq 2 \\ -x, & -1 \leq x \leq 0 \\ 2 + x, & -2 \leq x \leq 0 \\ 0, & |x| \geq 2 \,. \end{cases} \qquad (4.4)$$

Other truncation functions are considered in the literature. Changing the truncation function h typically changes the centering coefficient b, but does not change the Gaussian coefficient σ^2 or the Lévy measure μ. The Lévy measure has support on $\mathbb{R} - \{0\}$; it is a bonafide measure with

$$\int_{-\infty}^{\infty} \min\{1, x^2\}\mu(dx) < \infty \ . \qquad (4.5)$$

Given a specific truncation function, such as h in (4.4), there is a one-to-one correspondence between Lévy processes, infinitely distributions and the triple of characteristics (b, σ^2, μ) appearing in (4.3), with $\sigma^2 \geq 0$ and μ being a measure on $\mathbb{R} - \{0\}$ satisfying (4.5).

Brownian motion is the special Lévy process with null Lévy measure, i.e., $\mu(A) = 0$ for all measurable subsets A. Non-Gaussian stable Lévy motions with index α are the special cases with $\sigma^2 = 0$ and

$$\mu(dx) = \begin{cases} c^+ x^{-(1+\alpha)}, & x > 0 \ , \\ c^- |x|^{-(1+\alpha)}, & x < 0 \ , \end{cases} \qquad (4.6)$$

for nonnegative constants c^+ and c^-, where $c^+ + c^- > 0$. From (4.6), we see that the power-tail structure of a stable law is manifested very strongly in the Lévy measure. While the stable law $S_\alpha(\sigma, \beta, \mu)$ has the power-tail asymptotics in equations 4.5.12 and 4.5.13 in the book, the corresponding Lévy measure has simple power densities on $(0, \infty)$ and $(-\infty, 0)$. A stable Lévy motion is totally skewed to the right, so that $\beta = 1$, (left, so that $\beta = -1$) if and only if $c^- = 0$ ($c^+ = 0$).

The Lévy measure μ characterizes the possible jumps of the Lévy process. Indeed, the jump process of the Lévy process is a Poisson random measure on $\mathbb{R} \times \mathbb{R}^+$ with intensity $\mu(dx)dt$; i.e., the number of jumps in the Lévy process falling in any spatial subinterval $[a, b]$ during time subinterval $[c, d]$ for $a < b$ and $0 < c < d$ has a Poisson distribution with mean $\mu([a, b])|d - c|$. As a simple consequence, if the Lévy measure μ has support in $\mathbb{R}^+$, then the Lévy process has no negative jumps. Thus we know that the totally skewed stable Lévy motion with $\beta = 1$ (and thus $c^- = 0$ in (4.6)) has sample paths without negative jumps.

A complication with Lévy processes is the large (in general, infinite) number of very small jumps. For any $c > 0$, a Lévy process has only finitely many jumps of at least size c in any finite interval w.p.1. However, for any $c > 0$, it can have infinitely many jumps of absolute size less than or equal to c in any finite interval. This large number of small jumps is compensated for by deterministic drift built into the final integral in (4.3), in particular, this drift occurs in the region that the truncation function h is positive. Thus the true process drift is the sum of the drift b and the drift associated with h. In general, the total drift may be infinite, which explains why the representation (4.3) does not separate out all the drift.

It is possible to decompose a Lévy process into the independent sum of component Lévy processes by decomposing the exponent $\psi(\theta)$ in (4.3) into separate pieces; see Theorem 1 of p. 13 of Bertoin (1996) and its proof. The first component Lévy process L_1 has Lévy exponent

$$\psi_1(\theta) \equiv ib\theta - \frac{\sigma^2\theta^2}{2}$$

and is Brownian motion with drift coefficient b and diffusion coefficient σ^2. The second component Lévy process L_2 has exponent

$$\psi_2(\theta) = \int_{|x| \ge 2} (\exp(i\theta x) - 1)\mu(dx)$$

and is a compound Poisson process, with jumps of absolute size at least 2, having Poisson intensity $\lambda_2 \equiv \mu((-\infty, -2]) + \mu((2, \infty)) < \infty$ and jump size probability distribution $\mu(dx)/\lambda_2$ on $(-\infty, -2) \cup (2, \infty)$. The complicated component is the third one. The third component Lévy process L_3 has exponent

$$\psi_3(\theta) = \int_{-2}^{2} (\exp(i\theta x) - 1 - i\theta h(x))\mu(dx) \ .$$

It can be shown to be a pure jump martingale with jumps of absolute size at most 2. It includes some deterministic drift to compensate for the jumps. In summary, we can write

$$\psi(\theta) = \psi_1(\theta) + \psi_2(\theta) + \psi_3(\theta)$$

and

$$L \stackrel{\mathrm{d}}{=} L_1 + L_2 + L_3 \ ,$$

where L_1, L_2 and L_3 are the independent Lévy processes with exponents ψ_1, ψ_2 and ψ_3 defined above.

If an infinitely divisible distribution has finite moments, these moments can be derived by differentiating the characteristic function. For example, if $E|L(1)| < \infty$, then

$$EL(1) = \frac{\psi'(\theta)}{i} = b + \int_{-\infty}^{\infty} [x - h(x)]\mu(dx) \ , \tag{4.7}$$

where, because of the definition of the truncation function h, the integrand is non-zero only in $(-\infty, -1] \cup [1, \infty)$.

An important point is that the class of infinitely divisible distributions is remarkably large. An indication is the fact that infinitely divisible distributions are characterized by the triples (b, σ^2, μ), where μ is a measure on $\mathbb{R} - \{0\}$ satisfying (4.5). Two Lévy processes with triples (b_1, σ_1^2, μ_1) and (b_2, σ_2^2, μ_2) reduce to the same process if and only if $b_1 = b_2$, $\sigma_1^2 = \sigma_2^2$ and $\mu_1(A) = \mu_2(A)$ for all measurable sets $A \subseteq \mathbb{R}$. Nevertheless, infinitely divisible distributions may seem very special. However, over the years, many common distributions have been shown to be infinitely divisible. For example, lognormal distributions, Weibull distributions with ccdf's $e^{-(t/a)^c}$ for $c \leq 1$, Pareto distributions, and all mixtures of exponential distributions are infinitely divisible; see Thorin (1977a,b), p. 452 of Feller (1971), Bondesson (1992) and Abate and Whitt (1996). (The Weibull and Pareto distributions actually are mixtures of exponential distributions so infinite divisibility follows from that structure.) Moreover, the class of infinitely divisible distributions is easily seen to be closed under convolutions.

We now consider convergence in distribution of partial sums to infinitely divisible distributions and Lévy processes. First note that each infinitely divisible distribution can serve as a limit, because if X is infinitely divisible then there is a sequence of sequences $\{X_{n,k} : k \geq 1\}$ of IID random variables such that $X \stackrel{\mathrm{d}}{=} S_n$ for all n by the definition of infinite divisibility.

The following characterization of all possible limits is a consequence of Theorem 2, p. 303, of Feller (1971) and Theorem 2.7 of Skorohod (1957).

Theorem 2.4.1. (Lévy process FCLT for double sequences) *Let $\{X_{n,k} : k \geq 1\}$ be a sequence of IID random variables for each n and let $S_{n,n}$ and $\mathbf{S}_n$ be defined as in* (4.1) *and* (4.2). *If*

$$S_{n,n} \Rightarrow Z \quad in \quad \mathbb{R} ,$$

then Z has an infinitely divisible distribution and

$$\mathbf{S}_n \Rightarrow \mathbf{L} \quad in \quad D([0,\infty), J_1) ,$$

where $\mathbf{L}$ is the Lévy process with $\mathbf{L}(1) \stackrel{\mathrm{d}}{=} Z$.

Necessary and sufficient conditions for the FCLT with convergence to a specific Lévy process are consequences of Theorems 2.35, 2.52 and 3.4 of pp. 362, 368 and 373 of Jacod and Shiryaev (1987). (The partial sum process is both a semimartingale and a process with independent increments (PII) but not a process with stationary independent increments (PIIS).)

Theorem 2.4.2. (criteria for the Lévy-process FCLT) *Let $\{X_{n,k} : k \geq 1\}$ be a sequence of IID random variables for each n, with $\{X_{n,1} : n \geq 1\}$ being infinitesimal, i.e.,*

$$\lim_{n\to\infty} P(|X_{n,1}| > \epsilon) = 0 \quad for\ all \quad \epsilon > 0 . \tag{4.8}$$

Then

$$\mathbf{S}_n \Rightarrow \mathbf{L} \quad in \quad D([0,\infty), \mathbb{R}, J_1) \tag{4.9}$$

for $\mathbf{S}_n$ in (4.2), where $\mathbf{L}$ is a Lévy process with characteristics (b, σ^2, μ), if and only if

$$\text{(i)} \quad \lim_{n\to\infty} nE[h(X_{n,1}] = b , \tag{4.10}$$

$$\text{(ii)} \quad \lim_{n\to\infty} nVar[h(X_{n,1})] = \sigma^2 , \tag{4.11}$$

$$\text{(iii)} \quad \lim_{n\to\infty} nE[g(X_{n,1})] = \int_{-\infty}^{\infty} g(x)\mu(dx) , \tag{4.12}$$

for the truncation function h and all continuous bounded real-valued functions g on $\mathbb{R}$ with $g(x) = 0$ in a neighborhood of 0 and $g(x) \to y$, $-\infty < y < \infty$, as $x \to \pm\infty$.

Note that $h(x) = x$ for $|x| \leq 1$, so that conditions (i) and (ii) above correspond closely to convergence of the scaled means and variances.

Theorem 2.4.2 provides a large class of initial FCLT's to use with the continuous-mapping approach. We have only stated the classical results. Jacod and Shiryaev (1987) go much further, generalizing the characteristics of a Lévy process to define characteristics for semimartingales, allowing for nonstationarity. They also establish conditions for FCLTs in which processes with independent increments converge to other processes with independent increments (Chapter VII), semimartingales converge to processes with independent increments (Chapter VIII) and semimartingales converge to other semimartingales (Chapter IX), all expressed via the process characteristics. Actually verifying these conditions may not be straightforward, however.

2.5. Linear Models

In this section we discuss the linear-process representation in equation 4.6.6 of the book that was critical for obtaining the FCLT with strong dependence. The linear-process representation expresses the basic summands X_n as

$$X_n \equiv \sum_{j=0}^{\infty} a_j Y_{n-j}, \quad n \geq 1\,, \tag{5.1}$$

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

$$\sum_{j=0}^{\infty} a_j^2 < \infty\,. \tag{5.2}$$

We now show that the linear-process representation can arise naturally from modeling. First, however, it is important to repeat our earlier disclaimer. It is important to realize that the stochastic-process limits with strong dependence characterized by (5.1) are less universal. Many other forms of strong dependence are possible. And, if the dependence does not approximately correspond to a linear process, then there may appear a very different limit process or there may even be no stochastic-process limit at all.

Nevertheless, the linear-process representation is very natural. It provides a useful concrete model of strong dependence with an associated FCLT. To explain how linear models can arise, we describe some time-series models.

In particular, we show how the Gaussian linear process arises from a fundamental time-series model. We especially want to show how the Gaussian linear process with strong dependence arises from the *fractional autoregressive integrated moving average* (FARIMA) model; e.g., see Section 2.5 of Beran (1994) and Sections 7.12 and 7.13 of Samorodnitsky and Taqqu (1994).

The starting point is the *autoregressive moving average* (ARMA (p,q)) process, where p and q are nonnegative integers. To define the ARMA (p,q) process, let B be the *backshift operator*, defined by $BX_n \equiv X_{n-1}$, so that *differences* can be expressed as $X_n - X_{n-1} \equiv (1-B)X_n$ and $(X_n - X_{n-1}) - (X_{n-1} - X_{n-2}) \equiv (1-B)^2 X_n$. Let ϕ and ψ be polynomials of degree p and q, respectively, of the form

$$\phi(z) \equiv 1 - \sum_{j=1}^{p} \phi_j z^j$$

and

$$\psi(z) \equiv 1 + \sum_{j=1}^{q} \psi_j z^j ,$$

where z is a complex variable and $\phi_1, \ldots, \phi_p$, $\psi_1, \ldots, \psi_q$ are real coefficients. As regularity conditions, assume that the equations $\phi(z) = 0$ and $\psi(z) = 0$ have no common roots and that all solutions of the equation $\phi(z) = 0$ fall outside the unit disk $\{z : |z| \le 1\}$. An ARMA (p,q) process is defined to be the stationary solution to the equation

$$\phi(B)X_n = \psi(B)Y_n \qquad (5.3)$$

where $\{Y_n : n \ge 1\}$ is a sequence of IID $N(0,1)$ random variables; e.g., see Chapter 3 of Box, Jenkins and Reinsel (1994). In this setting, the sequence $\{Y_n\}$ is called the *innovation process*. Note that the exponential smoothing in Example 1.4.2 in the book is an ARMA$(1,0)$ process.

Theorem 2.5.1. (the ARMA process) *Under the regularity conditions above, the system of ARMA* (p,q) *equations* (5.3) *has a unique solution of the form*

$$X_n = \sum_{j=0}^{\infty} w_j Y_{n-j}, \quad n \ge 1, \qquad (5.4)$$

with real constant coefficients w_j *satisfying* $|w_j| < \delta^j$ *for all sufficiently large* j, *for some* δ, $0 < \delta < 1$. *The coefficients* w_j *in* (5.4) *are the coefficients of the power series* $\psi(z)/\phi(z)$.

Note that the coefficients w_j in the linear-process representation are available via their generating function $\psi(z)/\phi(z)$. Given the polynomials ψ and ϕ, we can thus calculate the coefficients w_j by numerically inverting the generating function; see Abate and Whitt (1992b).

Also note that the coefficients w_j in (5.4) decay exponentially fast, so that an ARMA process only exhibits weak dependence. To obtain strong dependence, we need the coefficients w_j in (5.4) to decay more slowly. We achieve that by considering fractional differencing. We do so by introducing a generalization of the ARIMA model. If instead $\{X_n\}$ is the solution of the equation

$$\phi(B)(1-B)^d X_n = \psi(B)Y_N \;, \tag{5.5}$$

where d is a nonnegative integer and $\{Y_n\}$ is again a sequence of IID $N(0,1)$ random variables, then $\{X_n\}$ is said to be an ARIMA (p,d,q) process, which was introduced by Box and Jenkins (1970); see Chapter 4 of Box, Jenkins and Reinsel (1994).

The FARIMA process is a generalization of the ARIMA process to fractional differencing. The FARIMA generalization of ARIMA was introduced by Granger and Joyeux (1980) and Hosking (1981). The FARIMA model with strong dependence depends on a parameter triple (p,q,d), where p and q are nonnegative integers and $0 < d < 1/2$. (There also are FARIMA models with $-1/2 < d \le 0$, but we will not consider them.) Given (p,q), there are $p+q$ further parameters.

For any real number d, we define the *fractional difference operator*

$$(1-B)^d \equiv \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \;,$$

where

$$\binom{d}{k} \equiv \frac{d!}{k!(d-k)!} \equiv \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

with $\Gamma(x)$ the gamma function. A stationary process $\{X_n\}$ that satisfies (5.5) for positive integers p and q and for $0 < d < 1/2$ is a FARIMA (p,d,q) process. (Values of d with $-1/2 < d \le 0$ are also possible, but we are primarily interested in the range $0 < d < 1/2$.)

Theorem 2.5.2. (the FARIMA process) *Under the regularity conditions above, including* $0 < d < 1/2$, *the system of FARIMA* (p,d,q) *equations* (5.5) *has a unique solution of the form*

$$X_n = \sum_{j=0}^{\infty} a_j Y_{n-j}, \quad n \ge 1 \,,$$

which converges almost surely, where

$$a_j \equiv \sum_{i=0}^{j} w_i b_{j-i}(-d)$$

with $\{w_j\}$ *being the sequence of constant coefficients in* (5.4) *and*

$$b_j(-d) \equiv \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \sim \frac{1}{\Gamma(d)} j^{d-1} \quad as \quad j \to \infty \, .$$

As a consequence,

$$a_j \sim a j^{d-1} \quad as \quad j \to \infty \, ,$$

where

$$a \equiv \sum_{i=0}^{j} w_i / \Gamma(d)$$

for w_i *in* (5.4), *and*

$$r_j \equiv Cov(X_1, X_{1+j}) \sim r j^{2d-1} \quad as \quad j \to \infty \, ,$$

where

$$r \equiv \left(\frac{\psi(1)}{\Gamma(d)\phi(1)} \right)^2 \int_0^\infty g(x) dx$$

for

$$g(x) = x^{2(d-1)} + (1+x)^{2(d-1)} - (x^{d-1} - (1+x)^{d-1})^2 \, .$$

The point of this discussion has been to show that a linear process of the form (5.1) and (5.2), with

$$Var(S_n) = n^{2H} L(n) \quad \text{as} \quad n \to \infty \, , \tag{5.6}$$

where $L(t)$ is a slowly varying function and $H > 1/2$, arises naturally from the FARIMA (p, d, q) model with $0 < d < 1/2$. In the FARIMA case the linear process is also a Gaussian process, but the key relations in Theorems 2.5.1 and 2.5.2 here hold for stationary sequences with finite second moments. We also remark that the parameters H and d are related by

$$d = H - \frac{1}{2} \, .$$

It is also significant that the FARIMA model provides a natural framework to exploit the strong dependence in order to make predictions; see

Beran (1994) for a full account of statistics for strongly dependent, light-tailed processes. We only make a few remarks.

In applications, we may have a stochastic sequence $\{X_n\}$ that we are willing to regard as a zero-mean stationary sequence with $\mathrm{Var}(X_n) < \infty$. We can examine the variance $\mathrm{Var}(S_n)$. If we find that

$$Var(S_n) \sim cn^{2H} \quad \text{as} \quad n \to \infty$$

for $1/2 < H < 1$, then we have the Joseph effect. That can be checked by looking for a linear relationship after taking logarithms; i. e.,

$$\log(Var(S_n)) \sim \log(c) + 2H \log(n) \ .$$

We then can invoke Theorem 4.6.1 in the book, without directly verifying the linear-process representation in (5.1) and without identifying the weights a_j in (5.1), to support the approximation (in distribution)

$$\{(cn^{2H})^{-1/2} S_{\lfloor nt \rfloor} : t \geq 0\} \approx \{Z_H(t) : t \geq 0\} \ , \tag{5.7}$$

where Z_H is standard FBM. Note that we obtain a parsimonious approximation, depending only on the two parameters c and H. Attention naturally focuses on ways to estimate the parameters c and H. That can be done simply from a plot of $\log Var(S_n)$ as a function of $\log n$; see Beran (1994).

It is important to remember that the justification of approximation (5.7) from Theorem 4.6.1 in the book actually depends on the linear-process representation. However, we can directly justify the approximation equation 4.6.13 in the book. by checking that the finite-dimensional distributions are approximately Gaussian and that the covariance function is approximately the covariance function of FBM in equation 4.6.13 in the book. The limit theorem explains why the FBM approximation may be appropriate.

We conclude by remarking that there is again a time-series motivation for considering the linear-process representation in the case of heavy tails plus dependence, discussed in Section 4.7 of the book. Specifically, there is a time-series motivation for the linear-process representation in equation 4.7.1 of the book, where the innovation variables Y_n have heavy tails, just as there was for the light-tailed case in Section 4.6 of the book, because there are analogs of the ARMA, ARIMA and FARIMA processes with stable innovations; i.e., there are analogs of Theorems 2.5.1 and 2.5.2 here for the case in which the innovation process $\{Y_n\}$ is a sequence of IID random variables with a stable law $S_\alpha(\sigma, \beta, \mu)$ for $0 < \alpha < 2$; see Sections 7.12 and 7.13 of Samorodnitsky and Taqqu (1994).