Chapter 3

The Framework for Stochastic-Process Limits

3.1. Introduction

In Chapters 1 and 2 we saw that plots of stochastic-process sample paths can suggest stochastic-process limits. Now we want to define precisely what we mean by those stochastic-process limits.

The main idea is to think of a stochastic process as a random function. With that mindset, convergence of a sequence of stochastic processes naturally becomes convergence of a sequence of probability measures on a function space (space of functions). There then remain three problems: First, what should we mean by the convergence of a sequence of probability measures on an abstract space? Second, what should be the underlying function space containing the sample paths of the stochastic processes? And, third, what should be the topology (notion of convergence) in the underlying function space?

We start in Section 3.2 by defining the standard notion of convergence for a sequence of probability measures on a metric space. We also define the Prohorov metric on the space of all probability measures on the metric space, which induces that convergence.

In Section 3.3 we discuss the function space D that we will use to represent the space of possible sample paths of the stochastic processes. We define two different metrics on the functions space D: One is the standard J_1 metric, which induces the Skorohod (1956) J_1 topology. The other is the M_1 metric, which induces the Skorohod (1956) M_1 topology. The commonly used J_1 topology is often referred to as "the Skorohod topology." We use the M_1 topology in order to be able to establish stochastic-process limits with unmatched jumps in the limit process.

In Section 3.4 we state three versions of the continuous-mapping theorem that support the continuous-mapping approach for obtaining new stochastic-process limits from established stochastic-process limits. In Section 3.5 we introduce useful functions mapping D or the product space $D \times D$ into D that preserve convergence and thus facilitate the continuous-mapping approach. We conclude in Section 3.6 by describing the organization of the book.

This chapter is intended to be brief, providing background for the introductory chapters. We elaborate in Chapter 11 and refer to Billingsley (1968, 1999) for more details.

3.2. The Space $\mathcal{P}$

Our goal is to precisely define what we mean by a *stochastic-process limit*, i.e., the convergence of a sequence of stochastic processes. We use metrics for that purpose. We define a metric on a space of stochastic processes in two steps: First, we define a metric on the space of probability measures on a general metric space and, second, we define a metric on the underlying function space containing the sample paths of the stochastic processes.

A *metric* is a distance function satisfying certain axioms. In particular, a metric m on a set S is a nonnegative real-valued function on the product space $S \times S \equiv \{(s_1, s_2) : s_1 \in S, s_2 \in S\}$ such that $m(x,y) = 0$ if and only if $x = y$, satisfying the *symmetry property*

$$m(x,y) = m(y,x) \quad \text{for all} \quad x, y \in S$$

and the *triangle inequality*

$$m(x,z) \leq m(x,y) + m(y,z) \quad \text{for all} \quad x, y, z \in S \ .$$

A *sequence* in a set S is a function mapping the positive integers into S. A sequence $\{x_n : n \geq 1\}$ in a metric space (S, m) *converges* to a limit x in S if, for all $\epsilon > 0$, there exists an integer n_0 such that $m(x_n, x) < \epsilon$ for all $n \geq n_0$. If we use the metric only to specify which sequences converge, then we characterize the topology induced by the metric: In a metric space, the topology is a specification of which sequences converge. Topology is the more general concept, because different metrics can induce the same topology. For further discussion about topologies, see Section 11.2.

As a regularity condition, we assume that the metric space (S, m) is *separable*, which means that there is a countable dense subset; i.e., there is a countably infinite (or finite) subset S_0 of S such that, for all $x \in S$ and all $\epsilon > 0$, there exists $y \in S_0$ such that $m(x, y) < \epsilon$.

We first consider probability measures on a general separable metric space (S, m). In our applications, the underlying metric space S will be the function space D, but now S can be any nonempty set. To consider probability measures on (S, m), we make S a *measurable space* by endowing it with a σ-field of measurable sets (discussed further in Section 11.3). For the separable metric space (S, m), we always use the *Borel σ-field* $\mathcal{B}(S)$, which is the smallest σ-field containing the *open balls*

$$B_m(x, r) \equiv \{y \in S : m(x, y) < r\}.$$

The elements of $\mathcal{B}(S)$ are called measurable sets. We mention measurability and σ-fields because, in general, it is not possible to define a probability measure (satisfying the axioms of a probability measure) on all subsets; see p. 233 of Billingsley (1968).

We say that a sequence of probability measures $\{P_n : n \geq 1\}$ on (S, m) *converges weakly* or just *converges* to a probability measure P on (S, m), and we write $P_n \Rightarrow P$, if

$$\lim_{n \to \infty} \int_S f dP_n = \int_S f dP \tag{2.1}$$

for all functions f in $C(S)$, the space of all continuous bounded real-valued functions on S. The metric m enters in by determining which functions f on S are continuous. It remains to show that this is a good definition; we discuss that point further in Section 11.3.

We now define the *Prohorov metric* on the space $\mathcal{P} \equiv \mathcal{P}(S)$ of all probability measures on the metric space (S, m); the metric was orginally defined by Prohorov (1956); see Dudley (1968) and Billingsley (1999). Let A^ϵ be the *open ϵ-neighborhood* of A, i.e.,

$$A^\epsilon \equiv \{y \in S : m(x, y) < \epsilon \quad \text{for some} \quad x \in A\} \ .$$

For P_1, $P_2 \in \mathcal{P}(S)$, the *Prohorov metric* is defined by

$$\pi(P_1, P_2) \equiv \inf\{\epsilon > 0 : P_1(A) \leq P_2(A^\epsilon) + \epsilon \quad \text{for all} \quad A \in \mathcal{B}(S)\} \ . \tag{2.2}$$

At first glance, it may appear that π in (2.2) lacks the symmetry property, but it holds. We prove the following theorem in Section 1.2 of the Internet Supplement.

Theorem 3.2.1. (the Prohorov metric on $\mathcal{P}$) *For any separable metric space (S, m), the function π on $\mathcal{P}(S)$ in (2.2) is a separable metric. There is convergence $\pi(P_n, P) \to 0$ in $\mathcal{P}(S)$ if and only if $P_n \Rightarrow P$, as defined in (2.1).*

We primarily want to specify when weak convergence $P_n \Rightarrow P$ holds, thus we are primarily interested in the topology induced by the Prohorov metric. Indeed, there are other metrics inducing this topology; e.g., see Dudley (1968).

Instead of directly referring to probability measures, we often use random elements. A *random element* X of $(S, \mathcal{B}(S))$ is a (measurable; see Section 11.3) mapping from some underlying probability space $(\Omega, \mathcal{F}, P)$ to $(S, \mathcal{B}(S))$. (In the underlying probability space, Ω is a set, $\mathcal{F}$ is a σ-field and P is a probability measure.) The *probability law* of X or the *probability distribution* of X is the image probability measure PX^{-1} induced by X on $(S, \mathcal{B}(S))$; i.e.,

$$\begin{aligned} PX^{-1}(A) &\equiv P(X^{-1}(A)) \equiv P(\{\omega \in \Omega : X(\omega) \in A\}) \\ &\equiv P(X \in A) \quad \text{for} \quad A \in \mathcal{B}(S) \ , \end{aligned}$$

where P is the probability measure in the underlying probability space $(\Omega, \mathcal{F}, P)$. We often use random elements, but when we do, we usually are primarily interested in their probability laws. Hence the underlying probability space $(\Omega, \mathcal{F}, P)$ is often left unspecified.

We say that a sequence of random elements $\{X_n : n \geq 1\}$ of a metric space (S, m) *converges in distribution* or *converges weakly* to a random element X of (S, m), and we write $X_n \Rightarrow X$, if the image probability measures converge weakly, i.e., if

$$P_n X_n^{-1} \Rightarrow PX^{-1} \quad \text{on} \quad (S, m) \ ,$$

using the definition in (2.1), where P_n and P are the underlying probability measures associated with X_n and X, respectively. It follows from (2.1) that $X_n \Rightarrow X$ if and only if

$$\lim_{n \to \infty} Ef(X_n) = Ef(X) \quad \text{for all} \quad f \in C(S) \ . \tag{2.3}$$

Thus convergence in distribution of random elements is just another way to talk about weak convergence of probability measures. When S is a function space, such as D, a random element of S becomes a *random function*, which we also call a *stochastic process*.

We can use the Skorohod representation theorem, also from Skorohod (1956), to help understand the topology of weak convergence in $\mathcal{P}(S)$. As before, $\stackrel{\mathrm{d}}{=}$ means equal in distribution.

Theorem 3.2.2. (Skorohod representation theorem) *If $X_n \Rightarrow X$ in a separable metric space (S, m), then there exist other random elements of (S, m), $\tilde{X}_n, n \geq 1$, and $\tilde{X}$, defined on a common underlying probability space, such that*

$$\tilde{X}_n \stackrel{\mathrm{d}}{=} X_n, n \geq 1, \quad \tilde{X} \stackrel{\mathrm{d}}{=} X$$

and

$$P(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}) = 1 \ .$$

The Skorohod representation theorem is useful because it lets us relate the structure of the space of probability measures $(\mathcal{P}, \pi)$ to the structure of the underlying metric space (S, m). It also serves as a basis for the continuous-mapping approach; see Section 3.4 below. We prove the Skorohod representation theorem in Section 1.3 of the Internet Supplement.

3.3. The Space *D*

We now consider the underlying function space of possible sample paths for the stochastic processes. Since we want to consider stochastic processes with discontinuous, but not too irregular, sample paths, we consider the space D of all right-continuous $\mathbb{R}^k$-valued functions with left limits defined on a subinterval I of the real line, usually either $[0,1]$ or $\mathbb{R}_+ \equiv [0,\infty)$; see Section 12.2 for additional details. We refer to the space as $D(I, \mathbb{R}^k)$, $D([0,1], \mathbb{R}^k)$ or $D([0,\infty), \mathbb{R}^k)$, depending upon the function domain, or just D when the function domain and range are clear from the context. The space D is also known as the space of *cadlag* or *càdlàg* functions – an acronym for the French *continu à droite, limites à gauche.*

The space D includes all continuous functions and the discontinuous functions of interest, but has useful regularity properties facilitating the development of a satisfactory theory. Let $C(I, \mathbb{R}^k)$, $C([0,1], \mathbb{R}^k)$ and $C([0,\infty), \mathbb{R}^k)$, or just C, denote the corresponding subsets of continuous functions.

We start by considering $D([0,1], \mathbb{R})$, i.e., by assuming that the domain is the unit interval $[0,1]$ and the range is $\mathbb{R}$. Recall that the space $D([0,1], \mathbb{R})$ was appropriate for the stochastic-process limits suggested by the plots in

Chapter 1. The reference metric is the *uniform metric* $\|x_1 - x_2\|$, defined in terms of the *uniform norm*

$$\|x\| \equiv \sup_{0 \le t \le 1} \{|x(t)|\} \ . \tag{3.1}$$

On the subspace C the uniform metric works well, but it does *not* on D: When functions have discontinuities, we do not want to insist that corresponding jumps occur exactly at the same times in order for the functions to be close. Appropriate topologies were introduced by Skorohod (1956). For a celebration of Skorohod's impressive contributions to probability theory, see Korolyuk, Portenko and Syta (2000).

To define the first metric on D, let Λ be the set of strictly increasing functions λ mapping the domain $[0,1]$ onto itself, such that both λ and its inverse λ^{-1} are continuous. Let e be the *identity map* on $[0,1]$, i.e., $e(t) = t$, $0 \le t \le 1$. Then the standard J_1 metric on $D \equiv D([0,1], \mathbb{R})$ is

$$d_{J_1}(x_1, x_2) \equiv \inf_{\lambda \in \Lambda} \{\|x_1 \circ \lambda - x_2\| \vee \|\lambda - e\|\} \ , \tag{3.2}$$

where $a \vee b \equiv max\{a, b\}$.

The general idea in going from the uniform metric $\|\cdot\|$ *to the* J_1 *metric* d_{J_1} *is to say functions are close if they are uniformly close over* $[0,1]$ *after allowing small perturbations of time* (*the function argument*). For example, $d_{J_1}(x_n, x) \to 0$ as $n \to \infty$, while $\|x_n - x\| \ge 1$ for all n, in $D([0,1], \mathbb{R})$ when $x = I_{[2^{-1},1]}$ and $x_n = (1 + n^{-1}) I_{[2^{-1}+n^{-1},1]}$, $n \ge 3$.

In the example above, the limit function has a single jump of magnitude 1 at time 2^{-1}. The converging functions have jumps of size $1 + n^{-1}$ at time $2^{-1} + n^{-1}$; both the magnitudes and locations of the single jump in x_n converge to those of the limit function x. That is a characteristic property of the J_1 topology. Indeed, from definition (3.2) it follows that, if $d_{J_1}(x_n, x) \to 0$ in $D([0,1], \mathbb{R})$, then for any t with $0 < t \le 1$ there necessarily exists a sequence $\{t_n : n \ge 1\}$ such that $t_n \to t$, $x_n(t_n) \to x(t)$, $x_n(t_n-) \to x(t-)$ and

$$x_n(t_n) - x_n(t_n-) \to x(t) - x(t-) \quad \text{as} \quad n \to \infty \ ;$$

i.e., the jumps converge. (It suffices to let $t_n = \lambda_n(t)$, where $\| \lambda_n - e \| \to 0$ and $\| x_n \circ \lambda_n - x \| \to 0$.) Thus, if x has a jump at t, i.e., if $x(t) \ne x(t-)$, and if $x_n \to x$, then for all n sufficiently large x_n must have a "matching jump" at some time t_n. That is, for any $\epsilon > 0$, we can find n_0 such that, for all $n \ge n_0$, there is t_n with $|t_n - t| < \epsilon$ and

$$|(x_n(t_n) - x_n(t_n-)) - (x(t) - x(t-))| < \epsilon \, .$$

We need a different topology on D if we want the jump in a limit function to be unmatched in the converging functions. For example, we want to allow continuous functions to be arbitrarily close to a discontinuous function; e.g., we want to have $d(x_n, x) \to 0$ when $x = I_{[2^{-1},1]}$ and

$$x_n = n(t - 2^{-1} + n^{-1}) I_{[2^{-1}-n^{-1}, 2^{-1})} + I_{[2^{-1},1]} \ ,$$

as shown in Figure 3.1. (We include dips in the axes because the points $2^{-1} - n^{-1}$ and 2^{-1} are not in scale. And similarly in later figures.) Notice

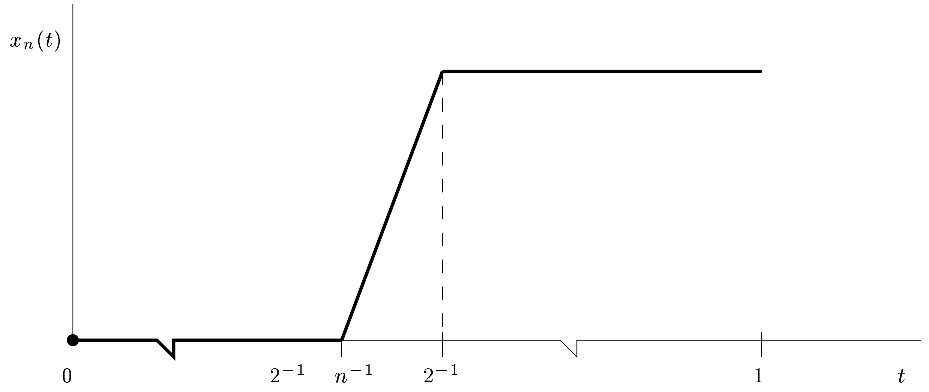

Figure 3.1: The continuous functions x_n that we want converging to the indicator function $x = I_{[2^{-1},1]}$ in D.

that both $\|x_n - x\| = 1$ and $d_{J_1}(x_n, x) = 1$ for all n, so that both the uniform metric and the J_1 metric on D are too strong.

Another example has discontinuous converging functions, but converging functions in which a limiting jump is approached in more than one jump. With the same limit x above, let

$$x_n = 2^{-1} I_{[2^{-1}-n^{-1}, 2^{-1})} + I_{[2^{-1},1]} \ ,$$

as depicted in Figure 3.2. Again, $\|x_n - x\| \not\to 0$ and $d_{J_1}(x_n, x) \not\to 0$ in D as $n \to \infty$.

In order to establish limits with unmatched jumps in the limit function, we use the M_1 metric. We define the M_1 metric using the completed graphs of the functions. For $x \in D([0,1], \mathbb{R})$, the *completed graph* of x is the set

$$\begin{aligned} \Gamma_x \equiv & \{(z,t) \in \mathbb{R} \times [0,1] : \\ & z = \alpha x(t-) + (1-\alpha) x(t) \quad \text{for some} \quad \alpha, \quad 0 \le \alpha \le 1\} \ , \end{aligned} \tag{3.3}$$

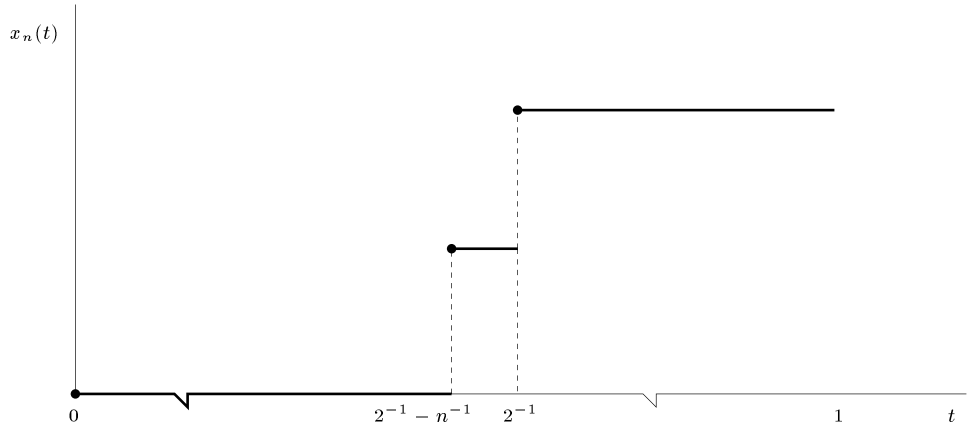

Figure 3.2: The two-jump discontinuous functions x_n that we want converging to the indicator function $x = I_{[2^{-1},1]}$.

where $x(t-)$ is the left limit of x at t. The completed graph is a connected subset of the plane $\mathbb{R}^2$ containing the line segment joining $(x(t), t)$ and $(x(t-), t)$ for all discontinuity points t. To illustrate, a function and its completed graph are displayed in Figure 3.3.

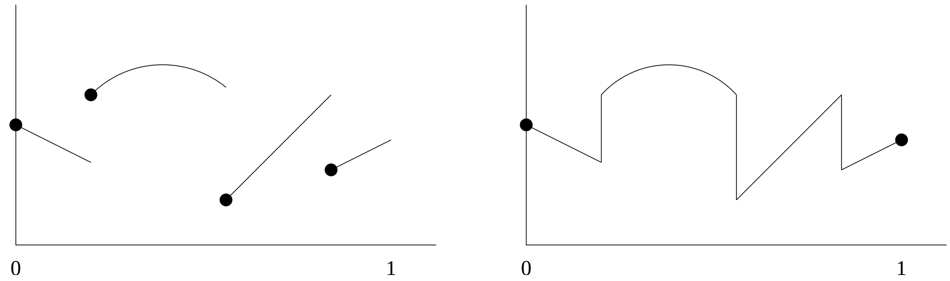

Figure 3.3: A function in $D([0, 1], \mathbb{R})$ and its completed graph.

We define the M_1 metric using the uniform metric defined on parametric representations of the completed graphs of the functions. To define the parametric representations, we need an order on the completed graphs. We define an *order* on the graph Γ_x by saying that $(z_1, t_1) \le (z_2, t_2)$ if either (i) $t_1 < t_2$ or (ii) $t_1 = t_2$ and $|x(t_1-) - z_1| \le |x(t_2-) - z_2|$. Thus the order is a total order, starting from the "left end" of the completed graph and concluding on the "right end".

A *parametric representation* of the completed graph Γ_x (or of the function x) is a continuous nondecreasing function (u,r) mapping $[0,1]$ onto Γ_x, with u being the spatial component and r being the time component. The parametric representation (u,r) is nondecreasing using the order just defined on the completed graph Γ_x.

Let $\Pi(x)$ be the set in $D \equiv D([0,1],\mathbb{R})$. For any $x_1,\ x_2 \in D$, the M_1 metric is

$$d_{M_1}(x_1,x_2) \equiv \inf_{\substack{(u_j,r_j)\in\Pi(x_j)\\ j=1,2}} \{\|u_1-u_2\| \vee \|r_1-r_2\|\} \ , \tag{3.4}$$

where again $a \vee b \equiv max\{a,b\}$. It turns out that d_{M_1} in (3.4) is a bonafide metric on D. (The triangle inequality is not entirely obvious; see Theorem 12.3.1.)

It is easy to see that, if x is continuous, then $d_{M_1}(x_n,x) \to 0$ if and only if $\|x_n - x\| \to 0$. It is also easy to see that $d_{M_1}(x_n,x) \to 0$ as $n \to \infty$ for the examples in Figures 3.1 and 3.2. To illustrate, we display in Figure 3.4 specific parametric representations (u,r) and (u_n,r_n) of the completed graphs of x and x_n for the functions in Figure 3.1 that yield the distance $d_{M_1}(x_n,x) = n^{-1}$. The spatial components u and u_n are identical. The time components satisfy $\| r_n - r \| = n^{-1}$.

For applications, it is significant that previous limits for stochastic processes with the familiar J_1 topology on D will also hold when we use the M_1 topology instead, because the J_1 topology is stronger (or finer) than the M_1 topology; see Theorem 12.3.2.

We now want to modify the space $D([0,1],\mathbb{R})$ in two ways: We want to extend the range of the functions from $\mathbb{R}$ to $\mathbb{R}^k$ and we want to allow the domain of the functions be the semi-infinite interval $[0,\infty)$ instead of the unit interval $[0,1]$. First, the J_1 and M_1 metrics extend directly to $D^k \equiv D([0,1],\mathbb{R}^k)$ when the norm $|\cdot|$ on $\mathbb{R}$ in (3.1) is replaced by a corresponding norm on $\mathbb{R}^k$ such as the maximum norm

$$\|a\| \equiv \max_{1\le i\le k} |a^i| \ ,$$

for $a \equiv (a^1,\ldots,a^k) \in \mathbb{R}^k$. With the maximum norm on $\mathbb{R}^k$, we obtain the standard or strong J_1 and M_1 metrics on D^k. We call the topology induced by these metrics the standard or *strong topology*, and denote it by SJ_1 and SM_1, respectively.

We also use the *product topology* on D^k, regarding D^k as the product space $D \times \cdots \times D$, which has $x_n \to x$ as $n \to \infty$ for $x_n \equiv (x_n^1,\ldots,x_n^k)$ and

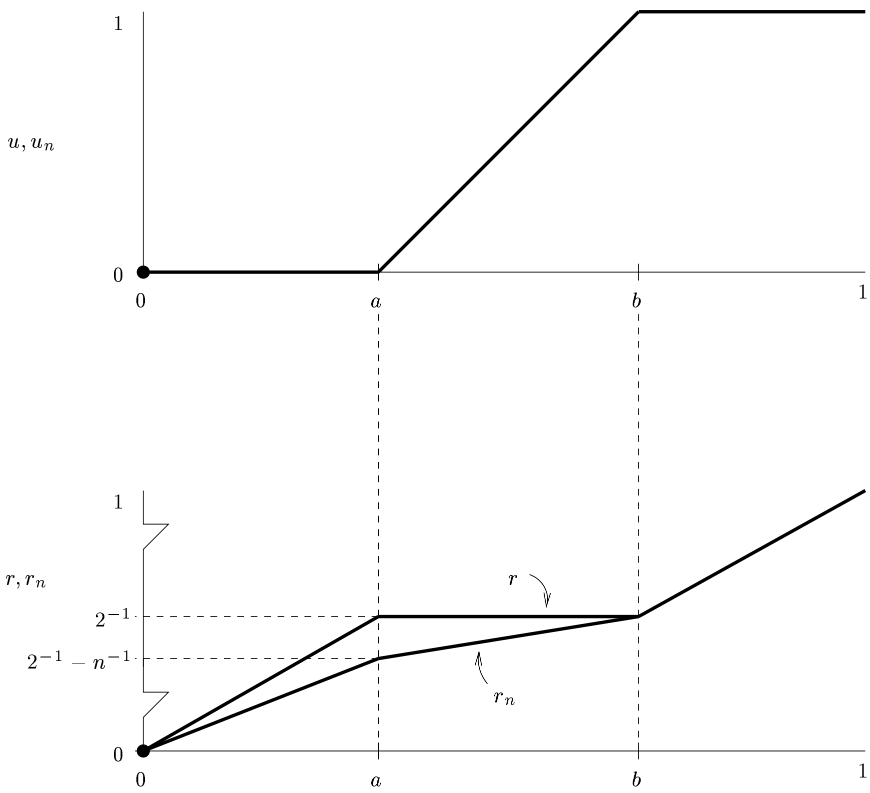

Figure 3.4: Plots of parametric representations (u, r) of Γ_x and (u_n, r_n) of Γ_{x_n} yielding $d_{M_1}(x_n, x) = n^{-1}$ for the functions in Figure 3.1. The points a and b are arbitrary, satisfying $0 < a < b < 1$.

$x \equiv (x^1, \ldots, x^k)$ in D^k if $x_n^i \to x^i$ as $n \to \infty$ in D for each i. The product topology on D^k is induced by the metric

$$d_p(x, y) \equiv \sum_{i=1}^{k} d(x^i, y^i) \ , \tag{3.5}$$

where d is the metric on D^1. Since convergence in the strong topology implies convergence in the product topology, we also call the product topology the *weak topology*, and we denote it by WJ_1 and WM_1.

The definitions for $D([0,1], \mathbb{R}^k)$ extend directly to $D([0,t], \mathbb{R}^k)$ for any $t > 0$. It is natural to characterize convergence of a sequence $\{x_n : n \geq 1\}$ in $D([0, \infty), \mathbb{R}^k)$ in terms of associated convergence of the restrictions of x_n

to the subintervals $[0,t]$ in the space $D([0,t],\mathbb{R}^k)$ for all $t>0$. However, note that we encounter difficulties in $D([0,t],\mathbb{R}^k)$ if the right endpoint t is a discontinuity point of a prospective limit function x. For example, if $t_n \to t$ as $n\to\infty$, but $t_n > t$ for all n, then the restrictions of $I_{[t_n,\infty)}$ to the subinterval $[0,t]$ are the zero function, while $I_{[t,\infty)}$ is not, so we cannot get the desired convergence $I_{[t_n,\infty)} \to I_{[t,\infty)}$ we want. Thus we say that the sequence $\{x_n : n\ge 1\}$ converges to x as $n\to\infty$ in $D([0,\infty),\mathbb{R}^k)$ if the restrictions of x_n to $[0,t]$ converge to the restriction of x to $[0,t]$ in $D([0,t],\mathbb{R}^k)$ for all $t>0$ that are continuity points of x.

The mode of convergence just defined can be achieved with metrics. Given a metric d_t on $D([0,t],\mathbb{R})$ applied to the restrictions of the functions to $[0,t]$, we define a metric d_∞ on $D([0,\infty),\mathbb{R})$ by letting

$$d_\infty(x_1,x_2) \equiv \int_0^\infty e^{-t}[d_t(x_1,x_2)\wedge 1]dt \ ,$$

where $a\wedge b \equiv \min\{a,b\}$ and $d_t(x_1,x_2)$ is understood to mean the distance d_t (either J_1 or M_1) applied to the restrictions of x_1 and x_2 to $[0,t]$.

The function space D with the J_1 or M_1 topology is somewhat outside the mainstream of traditional functional analysis, because *addition is not a continuous map* from the product space $D\times D$ with the product topology to D.

Example 3.3.1. *Addition is not continuous.*

A simple example has $x=-y=I_{[2^{-1},1]}$ with

$$x_n = I_{[2^{-1}-n^{-1},1]} \quad \text{and} \quad y_n = -I_{[2^{-1}+n^{-1},1]} \ .$$

Then $(x+y)(t)=0 \quad$ for all $\quad t$, while

$$x_n + y_n = I_{[2^{-1}-n^{-1},2^{-1}+n^{-1}]} \ .$$

With the non-uniform Skorohod topologies, $x_n\to x$ and $y_n\to y$ as $n\to\infty$, but $x_n+y_n \not\to x+y$ as $n\to\infty$. ∎

Thus, even though D is a vector space (we can talk about the linear combinations $ax+by$ for functions x and y in D and numbers a and b in $\mathbb{R}$), *D is not a topological vector space* (and thus not a Banach space) with the J_1 and M_1 topologies (because those structures require addition to be continuous).

Nevertheless, in applications of the continuous-mapping approach to establish stochastic-process limits, we will often want to add or subtract two functions. Thus it is very important that addition can be made to preserve convergence. It turns out that addition on $D \times D$ is measurable and it is continuous at limits in a large subset of $D \times D$. For any of the non-uniform Skorohod topologies, it suffices to assume that the two limit functions x and y have no common discontinuity points. With the M_1 topology (but not the J_1 topology), it suffices to assume that the two limit functions x and y have no common discontinuity points with jumps of opposite sign. (For instance, in Example 3.3.1, $x_n - y_n \to x - y$ in (D, M_1).) In many applications, we are able to show that the two-dimensional limiting stochastic process has sample paths in one of those subsets of pairs (x, y) w.p.1. Then we can apply the continuous-mapping theorem with addition.

3.4. The Continuous-Mapping Approach

The continuous-mapping approach to stochastic-process limits exploits previously established stochastic-process limits and the continuous-mapping theorem to obtain new stochastic-process limits of interest. Alternative approaches are the compactness approach described in Section 11.6 and various stochastic approaches (which usually exploit the compactness approach), which exploit special stochastic structure, such as Markov and martingale structure; e.g., see Billingsley (1968, 1999), Ethier and Kurtz (1986), Jacod and Shiryaev (1987) and Kushner (2001).

Here is a simple form of the continuous-mapping theorem:

Theorem 3.4.1. (simple continuous-mapping theorem). *If $X_n \Rightarrow X$ in (S, m) and $g : (S, m) \to (S', m')$ is continuous, then*

$$g(X_n) \Rightarrow g(X) \quad in \quad (S', m') \ .$$

Proof. Since g is continuous, $f \circ g$ is a continuous bounded real-valued function on (S, m) for each continuous bounded real-valued function f on (S', m'). Hence, under the conditions,

$$E[f \circ g(X_n)] \to E[f \circ g(X)]$$

for each continuous bounded real-valued function f on (S', m'), which implies the desired conclusion by (2.3). ∎

Paralleling the simple continuous-mapping theorem above, we can use a *Lipschitz-mapping theorem* to show that distances, and thus rates of convergence with the Prohorov metric, are preserved under Lipschitz mappings: A function g mapping a metric space (S, m) into another metric space (S', m') is said to be Lipschitz continuous, or just *Lipschitz*, if there exists a constant K such that

$$m'(g(x), g(y)) \leq Km(x, y) \quad \text{for all} \quad x, y \in S \ . \tag{4.1}$$

The infimum of all constants K for which (4.1) holds is called the *Lipschitz constant*. As before, let $a \vee b \equiv max\{a, b\}$. The following Lipschitz mapping theorem, taken from Whitt (1974a), is proved in Section 1.5 of the Internet Supplement. Applications to establish rates of convergence in stochastic-process limits are discussed in Section 2.2 of the Internet Supplement. We write $\pi(X, Y)$ for the distance between the probability laws of the random elements X and Y.

Theorem 3.4.2. (Lipschitz mapping theorem) *Suppose that* $g : (S, m) \rightarrow (S', m')$ *is Lipschitz as in (4.1) on a subset* B *of* S. *Then*

$$\pi(g(X), g(Y)) \leq (K \vee 1)\pi(X, Y)$$

for any random elements X *and* Y *of* (S, m) *for which* $P(Y \in B) = 1$.

We often need to go beyond the simple continuous-mapping theorem in Theorem 3.4.1. We often need to consider measurable functions that are only continuous almost everywhere or a sequence of such functions. Fortunately, the continuous-mapping theorem extends to such settings. We can work with a sequence of Borel measurable functions $\{g_n : n \geq 1\}$ all mapping one separable metric space (S, m) into another separable metric space $(S^{'}, m^{'})$. It suffices to have $g_n(x_n) \rightarrow g(x)$ as $n \rightarrow \infty$ whenever $x_n \rightarrow x$ as $n \rightarrow \infty$ for a subset E of limits x in S such that $P(X \in E) = 1$. This generalization follows easily from the Skorohod representation theorem, Theorem 3.2.2: Starting with the convergence in distribution $X_n \Rightarrow X$, we apply the Skorohod representation theorem to obtain the special random elements $\tilde{X}_n$ and $\tilde{X}$ with the same distributions as X_n and X such that $\tilde{X}_n \rightarrow \tilde{X}$ w.p.1. Since $\tilde{X} \stackrel{\mathrm{d}}{=} X$ and $P(X \in E) = 0$, we also have $P(\tilde{X} \in E) = 0$. We then apply the deterministic convergence preservation assumed for the functions g_n to get the limit

$$g(\tilde{X}_n) \rightarrow g(\tilde{X}) \quad \text{as} \quad n \rightarrow \infty \quad \text{in} \quad (S', m') \quad \text{w.p.1.}$$

Since convergence w.p.1 implies convergence in distribution, as a consequence we obtain

$$g(\tilde{X}_n) \Rightarrow g(\tilde{X}) \quad \text{in} \quad (S', m') \ .$$

Finally, since X_n and X are respectively equal in distribution to $\tilde{X}_n$ and $\tilde{X}$, also $g_n(X_n)$ and $g(X)$ are respectively equal in distribution to $g_n(\tilde{X}_n)$ and $g(\tilde{X})$. Thus, we obtain the desired generalization of the continuous-mapping theorem:

$$g_n(X_n) \Rightarrow g(X) \quad \text{in} \quad (S', m') \ .$$

It is also possible to establish such extensions of the simple continuous-mapping theorem in Theorem 3.4.1 directly, without resorting to the Skorohod representation theorem. We can use the continuous-mapping theorem or the generalized continuous-mapping theorem, proved in Section 1.5 of the Internet Supplement.

For $g : (S, m) \to (S', m')$, let $Disc(g)$ be the *set of discontinuity points* of g; i.e., $Disc(g)$ is the subset of x in S such that there exists a sequence $\{x_n : n \geq 1\}$ in S with $m(x_n, x) \to 0$ and $m'(g(x_n), g(x)) \not\to 0$.

Theorem 3.4.3. (continuous-mapping theorem) *If $X_n \Rightarrow X$ in (S, m) and $g : (S, m) \to (S', m')$ is measurable with $P(X \in Disc(g)) = 0$, then $g(X_n) \Rightarrow g(X)$.*

Theorem 3.4.4. (generalized continuous-mapping theorem) *Let g and g_n, $n \geq 1$, be measurable functions mapping (S, m) into (S', m'). Let the range (S', m') be separable. Let E be the set of x in S such that $g_n(x_n) \to g(x)$ fails for some sequence $\{x_n : n \geq 1\}$ with $x_n \to x$ in S. If $X_n \Rightarrow X$ in (S, m) and $P(X \in E) = 0$, then $g_n(X_n) \Rightarrow g(X)$ in (S', m').*

Note that $E = Disc(g)$ if $g_n = g$ for all n, so that Theorem 3.4.4 contains both Theorems 3.4.1 and 3.4.3 as special cases.

3.5. Useful Functions

In order to apply the continuous-mapping approach to establish stochastic-process limits, we need initial stochastic-process limits in D, the product space $D^k \equiv D \times \cdots \times D$ or some other space, and we need functions mapping D, D^k or the other space into D that preserve convergence. The initial limit is often Donsker's theorem or a generalization of it; see Chapters 4 and 7.

Since we are interested in obtaining stochastic-process limits, the functions preserving convergence must be D-valued rather than $\mathbb{R}$-valued or $\mathbb{R}^k$-valued. In this section we identify five basic functions from D or $D \times D$ to D that can be used to establish new stochastic-process limits from given ones:

addition, composition, supremum, reflection and inverse. These functions will be carefully examined in Chapters 12 and 13.

The *addition map* takes $(x, y) \in D \times D$ into $x + y$, where

$$(x + y)(t) \equiv x(t) + y(t), \quad t \geq 0 . \tag{5.1}$$

The *composition map* takes $(x, y) \in D \times D$ into $x \circ y$, where

$$(x \circ y)(t) \equiv x(y(t)), \quad t \geq 0 . \tag{5.2}$$

The *supremum map* takes $x \in D$ into $x^{\uparrow}$, where

$$x^{\uparrow}(t) \equiv \sup_{0 \leq s \leq t} x(s), \quad t \geq 0 . \tag{5.3}$$

The (one-sided, one-dimensional) *reflection map* index
takes $x \in D$ into $\phi(x)$, where

$$\phi(x) \equiv x + (-x \vee 0)^{\uparrow}, \quad t \geq 0 , \tag{5.4}$$

with $(x \vee 0)(t) \equiv x(t) \vee 0$. The *inverse* map takes x into x^{-1}, where

$$x^{-1}(t) \equiv \inf\{s \geq 0 : x(s) > t\}, \quad t \geq 0 . \tag{5.5}$$

Regularity conditions are required in order for the composition $x \circ y$ in (5.2) and the inverse x^{-1} in (5.5) to belong to D; those conditions will be specified in Chapter 13. We will also specify the domain for the functions in D; the common case is $\mathbb{R}_+ \equiv [0, \infty)$.

The general idea is that, by some means, we have already established convergence in distribution

$$\mathbf{X}_n \Rightarrow \mathbf{X} \quad \text{in} \quad D ,$$

and we wish to deduce that

$$\psi(\mathbf{X}_n) \Rightarrow \psi(\mathbf{X}) \quad \text{in} \quad D$$

for one of the functions ψ above. By virtue of the continuous-mapping theorem or the Skorohod representation theorem, it suffices to show that $\psi : D \to D$ is measurable and continuous at all $x \in A$, where $P(\mathbf{X} \in A) = 1$. Equivalently, in addition to the measurability, it suffices to show that ψ preserves convergence in D; i.e., that $\psi(x_n) \to \psi(x)$ whenever $x_n \to x$ for $x \in A$, where $P(\mathbf{X} \in A) = 1$.

There tends to be relatively little difficulty if A is a subset of continuous functions, but we are primarily interested in the case in which the limit has discontinuities. As illustrated by Example 3.3.1 for addition, when $x \notin C$, the basic functions often are not continuous in general. We must then identify an appropriate subset A in D, and work harder to demonstrate that convergence is indeed preserved.

Many applications of interest actually do not involve convergence preservation in such a simple direct form as above. Instead, the limits involve *centering*. In the deterministic framework (obtained after invoking the Skorohod representation theorem), we often start with

$$c_n(x_n - x) \rightarrow y \quad \text{in} \quad D\ , \tag{5.6}$$

where $c_n \rightarrow \infty$, from which we can deduce that

$$x_n \rightarrow x \quad \text{in} \quad D\ . \tag{5.7}$$

From (5.7) we can directly deduce that

$$\psi(x_n) \rightarrow \psi(x) \quad \text{in} \quad D$$

provided that ψ preserves convergence. However, we want more. We want to deduce that

$$c_n(\psi(x_n) - \psi(x)) \rightarrow z \quad \text{in} \quad D \tag{5.8}$$

and identify the limit z. We will want to show that (5.6) implies (5.8).

The common case is for x in (5.6)–(5.8) to be linear, i.e., for

$$x \equiv be, \quad \text{where } b \in \mathbb{R} \text{ and } e \in D \text{ with } e(t) \equiv t \quad \text{for all} \quad t\ .$$

We call that the case of *linear centering.* We will consider both linear and nonlinear centering.

The stochastic applications with centering are less straightforward. We might start with

$$\mathbf{X}_n \Rightarrow \mathbf{U} \quad \text{in} \quad D\,, \tag{5.9}$$

where

$$\mathbf{X}_n \equiv b_n^{-1}(X_n(nt) - \lambda nt), \quad t \geq 0\,,$$

for some stochastic processes $\{X_n(t) : t \geq 0\}$. Given (5.9), we wish to deduce that

$$\mathbf{Y}_n \Rightarrow \mathbf{V} \quad \text{in} \quad D \tag{5.10}$$

and identify the limit process $\mathbf{V}$ for

$$\mathbf{Y}_n \equiv b_n^{-1}(\psi(X_n)(nt) - \mu nt), \quad t \geq 0\,. \tag{5.11}$$

To apply the convergence-preservation results with centering, we can let

$$x_n(t) \equiv (n\lambda)^{-1} X_n(nt), \quad x(t) \equiv e(t) \equiv t, \quad c_n \equiv n\lambda / b_n$$

and assume that $|c_n| \to \infty$. The w.p.1 representation of the weak convergence in (5.9) yields

$$c_n(x_n - x) \to u \quad \text{w.p.1} \quad \text{in} \quad D\,,$$

where u is distributed as $\mathbf{U}$. The convergence-preservation result ((5.6) implies (5.8)) then yields

$$c_n[\psi(x_n) - \psi(x)] \to v \quad \text{w.p.1} \quad \text{in} \quad D\,. \tag{5.12}$$

We thus need to relate the established w.p.1 convergence in (5.12) to the desired convergence in distribution in (5.10). This last step depends upon the function ψ. To illustrate, suppose, as is the case for the supremum and reflection maps in (5.3) and (5.4), that $\psi(e) = e$ and ψ is homogeneous, i.e., that

$$\psi(ax) = a\psi(x) \quad \text{for} \quad x \in D \quad \text{and} \quad a > 0\,.$$

Then

$$c_n[\psi(x_n) - \psi(e)] = b_n^{-1}[\psi(X_n)(nt) - \lambda nt]\,.$$

Thus, under those conditions on ψ, we can deduce that (5.10) holds for $\mathbf{Y}_n$ in (5.11) with $\mu = \lambda$ and $\mathbf{V}$ distributed as v in (5.12).

In applications, our primary goal often is to obtain convergence in distribution for a sequence of real-valued random variables, for which we only need to consider the continuous mapping theorem with real-valued functions. However, it is often convenient to carry out the program in two steps: We start with a FCLT in D for a sequence of basic stochastic processes such as random walks. We then apply the continuous-mapping theorem with the kind of functions considered here to obtain new FCLT's for the basic stochastic processes in applied probability models, such as queue-length stochastic processes in a queueing model. Afterwards, we obtain desired limits for associated random variables of interest by applying the continuous-mapping theorem again with real-valued functions of interest. The final map may be the simple one-dimensional projection map π_t mapping $x \in D$ into $x(t) \in \mathbb{R}^k$ when $\mathbb{R}^k$ is the range of the functions in D, the average $t^{-1}\int_0^t x(s)ds$ or something more complicated.

3.6. Organization of the Book

We now expand upon the description of the organization of the book given at the end of the preface. As indicated there, the book has fifteen chapters, which can be roughly grouped into four parts, ordered according to increasing difficulty. The *first part*, containing the first five chapters, provides an informal introduction to stochastic-process limits and their application to queues.

Chapter 1 exposes the statistical regularity associated with a macroscopic view of uncertainty, with appropriate scaling, via plots of random walks, obtained from elementary stochastic simulations. Remarkably, the plotter automatically does the proper scaling when we plot the first n steps of the random walk for various values of n. The plots tend to look the same for all n sufficiently large, showing that there must be a stochastic-process limit. For random walks with IID steps having infinite variance, the plots show that the limit process must have jumps, i.e., discontinuous sample paths.

Chapter 2 shows that the abstract random walks considered in Chapter 1 have useful applications. Chapter 2 discusses applications to stock prices, the Kolmogorov-Smirnov statistic and queueing models. Chapter 2 also discusses the engineering significance of the queueing models and the heavy-traffic limits. The engineering significance is illustrated by applications to buffer sizing in network switches and service scheduling for multiple sources.

The present chapter, Chapter 3, introduces the mathematical framework for stochastic-process limits, involving the concept of weak convergence of a sequence of probability measures on a separable metric space and the function space D containing stochastic-process sample paths. Metrics inducing the Skorohod J_1 and M_1 topologies on D are defined. An overview of the continuous-mapping approach to establish stochastic-process limits is also given.

Chapter 4 provides an overview of established stochastic-process limits. These stochastic-process limits are of interest in their own right, but they also serve as starting points in the continuous-mapping approach to establish new stochastic-process limits. The fundamental stochastic-process limit is provided by Donsker's theorem, which was discussed in Chapter 1. The other stochastic-process limits are generalizations of Donsker's theorem. Of particular interest for the limits with jumps, is the generalization of Donsker's theorem in which the random-walk steps are IID with infinite variance. When the random-walk steps have such heavy-tailed distributions, the limit process is a stable Lévy motion in the case of a single sequence or a general Lévy process in the case of a triangular array or double sequence. When these limit processes are not Brownian motion, they have discontinuous sample paths. The stochastic-process limits with jumps in the limit process explain some of the jumps observed in the simulation plots in Chapter 1.

Lévy processes are very special because they have independent increments. Chapter 4 also discusses stochastic-process limits in which the limit process has dependent increments. The principal stochastic-process limits of this kind involve convergence to fractional Brownian motion and linear fractional stable motion. These limit processes with dependent increments arise when there is strong dependence in the converging stochastic processes. These particular limit processes have continuous sample paths, so the topology on D is not critical. Nevertheless, like heavy tails, strong dependence has a dramatic impact on the stochastic-process limit, changing both the scaling and the limit process.

Chapter 5 provides an introduction to heavy-traffic limits for queues. This first queueing chapter focuses on a general fluid queue model that captures the essence of many more-detailed queueing models. This fluid queue model is especially easy to analyze because the continuous-mapping approach with the reflection map can be applied directly. Section 5.5 derives scaling functions, expressed as functions of the traffic intensity in the queue, which provide insight into queueing performance. Proofs are provided in Chapter 5, but the emphasis is on the statement and applied value of the heavy-traffic limits rather than the technical details. This first queueing chapter emphasizes the classical Brownian approximation (involving a reflected Brownian motion limit process). The value of the Brownian approximation is illustrated in the Section 5.8, which discusses its application to plan queueing simulations: The heavy-traffic scaling produces a simple approximation for the simulation run length required to achieve desired statistical precision, as a function of model parameters.

The *second part*, containing Chapters 6 – 10, show how unmatched jumps can arise and expands the treatment of queueing models. Chapter 6 gives several examples of stochastic-process limits with unmatched jumps in the limit process. In all the examples it is obvious that either there are no jumps in the sample paths of the converging processes or the jumps in the converging processes are asymptotically negligible. What is not so obvious is that the limit process actually can have discontinuous sample paths. As in Chapter 1, simulations are used to provide convincing evidence.

Chapter 7 continues the overview of stochastic-process limits begun in

Chapter 4. It first discusses process CLT's, which are central limit theorems for appropriately scaled sums of random elements of D. Process CLT's play an important role in heavy-traffic stochastic-process limits for queues with superposition arrival processes, when the number of component arrival processes increases in the heavy-traffic limit.

Then Chapter 7 discusses CLT's and FCLT's for counting processes. They are shown to be equivalent to corresponding limits for partial sums. Chapter 7 concludes by applying the continuous-mapping approach with the composition and inverse maps, together with established stochastic-process limits in Chapter 4, to establish stochastic-process limits for renewal-reward stochastic processes. The M_1 topology plays an important role in Chapter 7.

The remaining chapters in the second part apply the stochastic-process limits, with the continuous-mapping approach, to obtain more heavy-traffic limits for queues. As in Chapter 5, Chapters 8 – 10 emphasize the applied value of the stochastic-process limits, but now more attention is given to technical details. Chapter 8 considers a more-detailed multi-source on-off fluid-queue model that has been proposed to evaluate the performance of communication networks. That model illustrates how heavy-traffic limits can expose the essential features of complex models. This second queueing chapter also discusses non-classical approximations involving reflected stable Lévy motion and reflected fractional Brownian motion, stemming from heavy-tailed probability distributions and strong dependence.

Chapter 9 focuses on standard single-server queues, while Chapter 10 focuses on standard multi-server queues. In addition to the standard heavy-traffic limits, we consider heavy-traffic limits in which the number of component arrival processes in a superposition arrival process or the number of servers in a multi-server queue increases in the heavy-traffic limit. Those limits tend to capture the behavior of systems with large numbers of sources or servers.

The *third part*, containing Chapters 11 – 14, is devoted to the technical foundations needed to establish stochastic-process limits with unmatched jumps in the limit process. The third part begins with Chapter 11, which provides more details on the mathematical framework for stochastic-process limits, expanding upon the brief introduction in Chapter 3.

Chapter 12 presents the basic theory for the function space D. Four topologies are considered on D: strong and weak versions of the M_1 topology and strong and weak versions of the M_2 topology. The strong and weak topologies differ when the functions have range $\mathbb{R}^k$ for $k > 1$. The strong topologies agree with the standard topologies defined by Skorohod

(1956), while the weak topologies agree with the product topology, regarding $D([0,T],\mathbb{R}^k)$ as the k-fold product of the space $D([0,T],\mathbb{R})$ with itself. The M topologies are defined and characterized in Chapter 12. The main ideas go back to Skorohod (1956), but more details are provided here. For example, several useful alternative characterizations of these topologies are given; e.g., see Theorem 12.5.1.

Chapter 13 focuses on the useful functions from D or $D \times D$ to D introduced in Section 3.5, which preserve convergence with the Skorohod topologies, and thus facilitate the continuous-mapping approach to establish new stochastic-process limits. As illustrated in the queueing chapters, the functions in Chapter 13 can be combined with the FCLT's in Chapter 4 to obtain many new stochastic-process limits.

The third part concludes with a final chapter on queues: Chapter 14 establishes heavy-traffic limits for networks of queues. The extension to networks of queues in Chapter 14 is more complicated because, unlike the one-dimensional reflection map used for single queues, the multidimensional reflection map is not simply continuous in the M_1 topology. However, it is continuous, using the product M_1 topology, at all limit functions without simultaneous jumps of opposite sign in its coordinate functions.

The *fourth part*, containing only the final chapter, Chapter 15, introduces new function spaces larger than D. These spaces, called E and F, are intended to express limits for sequences of stochastic processes with oscillations in their sample paths so great that there is no limit in D. The names are chosen because of the ordering

$$C \subset D \subset E \subset F .$$

Example 3.6.1. *Motivation for the spaces E and F.* Suppose that the n^{th} function in a sequence of continuous functions takes the value 4 in the interval $[0, 2^{-1} - n^{-1}]$, the value 5 in the interval $[2^{-1} + n^{-1}, 1]$ and has oscillations in the subinterval $[2^{-1} - n^{-1}, 2^{-1} + n^{-1}]$ for all $n \geq 3$. Specifically, within the subinterval $[2^{-1} - n^{-1}, 2^{-1} + n^{-1}]$, let this n^{th} function first increase from the value 4 at the left endpoint $2^{-1} - n^{-1}$ to 7, then decrease to 1, and then increase again to 5 at the right endpoint $2^{-1} + n^{-1}$, as shown in Figure 3.5.

That sequence of continuous functions converges pointwise to the limit function $x = 4I_{[0,2^{-1})} + 5I_{[2^{-1},1]}$ everywhere except possibly at $t = 1/2$, but it does not converge in D with any of the Skorohod topologies. Nevertheless, we might want to say that convergence does in fact occur, with the limit somehow revealing the oscillations of the functions in the neighborhood of $t = 1/2$. The spaces E and F allow for such limits

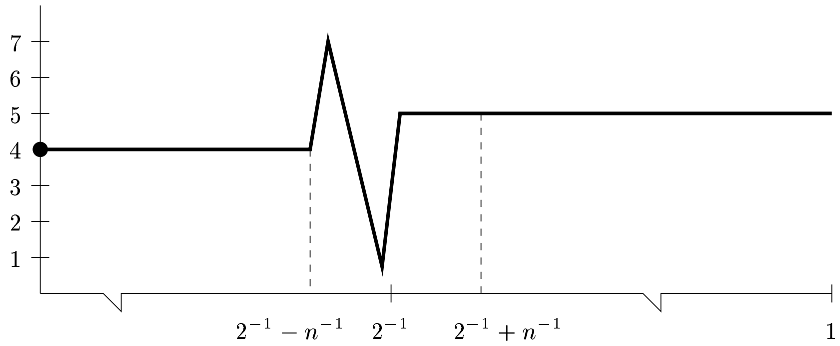

Figure 3.5: The n^{th} function in $C[0,1]$ in a sequence of functions that converges to a proper limit in the space E but not in the space D.

In E, the limit corresponds to the set-valued function that is the one-point set $\{4\}$ for $t \in [0, 2^{-1})$, the one-point set $\{5\}$ for $t \in (2^{-1}, 1])$ and is the interval $[1, 7]$ at $t = 1/2$.

In E, the limit fails to capture the order in which the points are visited in the neighborhood of $t = 1/2$. The space F exploits parametric representations to also capture the order in which the points are visited. The larger spaces E and F are given topologies similar to the M_2 and M_1 topologies on D. Thus Chapter 15 draws heavily upon the development of the M topologies in Chapter 12. However, Chapter 15 only begins to develop the theory of E and F. Further development is a topic for future research. ■

At the end of the book there are two appendices. Appendix A gives basic facts about regularly varying functions, while Appendix B gives the intial contents of the Internet Supplement.