# Chapter 1

# Fundamentals

## 1.1. Introduction

In this chapter we present material supplementing the book on fundamental topics. In Sections 1.2 and 1.3 we give detailed proofs of the Prohorov metric properties and the Skorohod representation theorem, stated in Theorems 3.2.1 and 3.2.2 of the book. In Section 1.4 we explain the adjective "weak" in weak convergence from a Banach-space perspective. In Section 1.5 we provide proofs of the continuous mapping theorems, stated in Section 3.4 of the book.

## 1.2. The Prohorov Metric

In this section we prove Theorem 3.2.1 in the book, establishing that the Prohorov (1956) metric is indeed a metric inducing weak convergence $P_n \Rightarrow P$.

Recall that we are considering probability measures on a separable metric space $(S, m)$. In that setting, $P_n \Rightarrow P$ 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 $S$. Recall that the Prohorov metric $\pi$ is defined on the space $\mathcal{P} \equiv \mathcal{P}(S)$ of all probability measures on the separable metric space $(S, m)$ by

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

for $P_1$, $P_2 \in \mathcal{P}(S)$, where $A^\epsilon$ is the *open $\epsilon$-neighborhood* of $A$, i.e.,

$$A^\epsilon \equiv \{y \in S : m(x,y) < \epsilon \quad \text{for some} \quad x \in A\} . \tag{2.3}$$

Here is the result that we wish to prove:

**Theorem 1.2.1.** (the Prohorov metric on $\mathcal{P}$) *For any separable metric space* $(S, m)$*, the function* $\pi$ *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). *Moreover, in* (2.2) *it suffices to let the sets* $A$ *be closed.*

To carry out the proof, we show that weak convergence $P_n \Rightarrow P$ implies uniform convergence of integrals $\int g dP_n$ for an appropriate class of functions $g$.

Consider a class $\mathcal{G}$ real-valued functions on $S$. We say that $\mathcal{G}$ is *uniformly bounded* if

$$\sup_{g \in \mathcal{G}, x \in S} \{|g(x)|\} < \infty .$$

We say that $\mathcal{G}$ is *equicontinuous at* $x$ if, for all $\epsilon > 0$, there is a $\delta > 0$ such that

$$\sup_{g \in \mathcal{G}} |g(x) - g(y)| < \epsilon \quad \text{when} \quad d(x,y) < \delta .$$

We say that $\mathcal{G}$ is *equicontinuous* if it is equicontinuous at all $x \in S$.

**Lemma 1.1.** (uniform convergence for a class of integrals) *Suppose that* $P_n \Rightarrow P$ *on a separable metric space* $(S, m)$*. Let* $\mathcal{G}$ *be a uniformly bounded class of measurable real-valued functions on* $S$ *that is equicontinuous at all* $x \in E^c$*. If* $P(E) = 0$*, then*

$$\sup_{g \in \mathcal{G}} \left| \int g dP_n - \int g dP \right| \to 0 \quad \textit{as} \quad n \to \infty . \tag{2.4}$$

**Proof.** If (2.4) were to fail, then there must exist $\epsilon > 0$ and a sequence $\{g_n : n \geq 1\}$ of functions in $\mathcal{G}$ for which $|\int g_n dP_n - \int g_n dP| > \epsilon$ infinitely often. We will show that cannot happen. Given $P_n \Rightarrow P$, we can apply the Skorohod representation theorem to construct $S$-valued random elements $X_n$ and $X$ with probability laws $P_n$ and $P$ such that $X_n \to X$ w.p.1. By the almost-sure equicontinuity of $\mathcal{G}$ with respect to $P$,

$$\sup_n |g_n(X_n) - g_n(X)| \to 0 \quad \text{w.p.1} .$$

By the uniform-boundedness condition and the bounded convergence theorem,

$$\sup_n |Eg_n(X_n) - Eg_n(X)| \le E\left[\sup_n |g_n(X_n) - g_n(X)|\right] \to 0 \quad \text{as} \quad n \to \infty$$

or, equivalently,

$$\sup_n \left| \int g_n dP_n - \int g_n dP \right| \to 0 \quad \text{as} \quad n \to \infty \, .$$

Since that is a contradiction, (2.4) must actually hold. ∎

We now define a generalization of the Prohorov metric on the space $\mathcal{P}(S)$ of all probability measures on $(S, m)$. We define a family of metrics indexed by the scalar $\gamma$; the standard Prohorov metric is the special case with $\gamma = 1$. For any $P_1$, $P_2 \in \mathcal{P}(S)$ and $\gamma > 0$, let

$$\pi_\gamma(P_1, P_2) \equiv \inf\{\epsilon > 0 : P_1(F) \le P_2(F^\epsilon) + \gamma\epsilon \quad \text{for all closed } F \text{ in } S\} \, , \tag{2.5}$$

where $F^\epsilon$ is the open $\epsilon$-neighborhood of $F$, as in (2.3).

Here is our main result.

**Theorem 1.1.** (generalized Prohorov metric) *Let $(S, m)$ be a separable metric space. For each $\gamma > 0$, $(\mathcal{P}(S), \pi_\gamma)$ for $\pi_\gamma$ in* (2.5) *is a separable metric space. The definition is unchanged if the closed sets $F$ in (2.5) are replaced by general measurable sets $A$. There is convergence $\pi_\gamma(P_n, P) \to 0$ as $n \to \infty$ if and only if $P_n \Rightarrow P$.*

In preparation for the proof, we first establish some preliminary results. We first show that $\pi_\gamma(P_2, P_1) = \pi_\gamma(P_1, P_2)$. For that purpose, use the following elementary lemma. Recall that $A^-$ is the closure of the set $A$.

**Lemma 1.2.** *For any subset $A$ of $S$ and $\alpha > 0$,*

$$A^- = S - (S - A^\alpha)^\alpha \, . \tag{2.6}$$

**Lemma 1.3.** *If $P_1(F) \le P_2(F^\alpha) + \beta$ for all closed $F$ for $\alpha, \beta > 0$, then $P_2(F) \le P_1(F^\alpha) + \beta$ for all closed $F$.*

**Proof.** Since $F^\alpha$ is open, $S - F^\alpha$ is closed. Under the condition,

$$P_1(S - F^\alpha) \le P_2((S - F^\alpha)^\alpha) + \beta \ ,$$

so that

$$P_2(S - (S - F^\alpha)^\alpha) \le P_1(F^\alpha) + \beta \ .$$

By Lemma 1.2, $F = S - (S - F^\alpha)^\alpha$. hence

$$P_2(F) = P_2(S - (S - F^\alpha)^\alpha) \le P_1(F^\alpha) + \beta \ . \quad \blacksquare$$

We now show that closed sets and measurable sets are interchangeable in (2.5).

**Lemma 1.4.** (closed sets suffice) *For any constants $\alpha > 0$ and $\beta > 0$, the inequality $P_1(A) \le P_2(A^\alpha) + \beta$ holds for all $A \in \mathcal{S}$ if and only if it holds for all $A = F$, where $F$ is closed.*

**Proof.** One direction is immediate. For the non-trivial direction, given any measurable set $A$, choose a sequence of closed sets $\{F_n : n \ge 1\}$ such that $F_n \subseteq F_{n+1}$ and $F_n \uparrow A$. Then $F_n^\alpha \uparrow F^\alpha$, $P_1(F_n) \uparrow P_1(A)$ and $P_2(F_n^\alpha) \uparrow P_2(A^\alpha)$. Hence we have $P_1(A) \le P_2(A^\alpha) + \beta$ when we have $P_1(F_n) \le P_2(F_n^\alpha) + \beta$ for all $n$. $\blacksquare$

**Proof of Theorem 1.1.** Lemma 1.3 establishes the symmetry property. if $\pi_\gamma(P_1, P_2) = 0$, then $P_1(F) = P_2(F)$ for each closed subset $F$. Since the closed sets form a determining class, $P_1 = P_2$. To establish the triangle inequality, suppose that $\pi_\gamma(P_1, P_2) < \epsilon_1 < \pi_\gamma(P_1, P_2) + \delta$ and $\pi_\gamma(P_2, P_3) < \epsilon_2 < \pi_\gamma(P_2, P_3) + \delta$ for some $\delta > 0$. Then for any closed $F$,

$$\begin{aligned} P_1(F) &\le P_2(F^{\epsilon_1}) + \gamma\epsilon_1 \\ &\le P_2((F^{\epsilon_1})^-) + \gamma\epsilon_1 \\ &\le P_3(F^{\epsilon_1+\epsilon_2}) + \gamma(\epsilon_1 + \epsilon_2) \ , \end{aligned}$$

so that

$$\pi_\gamma(P_1, P_3) \le \epsilon_1 + \epsilon_2 \le \pi_\gamma(P_1, P_2) + \pi_\gamma(P_2, P_3) + 2\delta \ .$$

Since $\delta$ was arbitrary, the triangle inequality is established, completing the proof of the metric property.

If $\pi_\gamma(P_n, P) \to 0$, then for any $\epsilon > 0$ there exists $n_0$ such that $P_n(F) \le P(F^\epsilon) + \gamma\epsilon$ for all closed $F$ and $n \ge n_0$. Hence

$$\limsup_{n\to\infty} P_n(F) \le P(F^\epsilon) + \gamma\epsilon \ .$$

However, $F^{\epsilon} \downarrow F$ as $\epsilon \downarrow 0$, so that $P(F^{\epsilon}) \downarrow P(F)$ as $\epsilon \downarrow 0$. Hence,

$$\limsup_{n\to\infty} P_n(F) \le P(F) ,$$

which implies $P_n \Rightarrow P$ by Theorem 11.3.1 in the book.

Next we show that $\pi_{\gamma}(P_n, P) \to 0$ if $P_n \Rightarrow P$. For each $A \in \mathcal{S}$, define

$$g_A(x) \equiv [1 - \epsilon^{-1} m(x, A)]^{+} . \tag{2.7}$$

Notice that $I_A(x) \le g_A(x) \le I_{A^{\epsilon}}(x)$ for all $x$, where $I_B$ is the indicator function of the set $B$. Moreover,

$$|g_A(x) - g_A(y)| \le \epsilon^{-1}|m(x,A) - m(y,A)| \le \epsilon^{-1} m(x,y)$$

for all $A$, so that the class of all such $g_A$ defined in (2.7) is uniformly bounded and equicontinuous. By Lemma 1.1,

$$\Delta_n \equiv \sup_{A\in\mathcal{S}} \left| \int g_A dP_n - \int g_A dP \right| \to 0 \quad \text{as} \quad n \to \infty .$$

Then

$$P(A^{\epsilon}) \ge \int g_A dP \ge \int g_A dP_n - \Delta_n \ge P_n(A) - \Delta_n$$

so that

$$P_n(A) \le P(A^{\epsilon}) + \epsilon$$

when $\Delta_n < \epsilon$.

Finally, we want to show that $(\mathcal{P}(S), \pi_{\gamma})$ is separable. For that purpose, let $S_0$ be a countable dense subset of $(S, m)$, which exists because we have assumed that $(S, m)$ is separable. We will show that the countable family of rational-valued probability measures with finite support in $S_0$ are dense in $\mathcal{P}(S)$.

Given any $P_1 \in \mathcal{P}(S)$ and any $\epsilon > 0$, we show how to construct $P_2$ with finite support in $S_0$ such that $P_1(A) \le P_2(A^{\epsilon})$ for all $A \in \mathcal{S}$, so that $\pi_{\gamma}(P_1, P_2) \le \epsilon$. Let the sequence $\{x_n : n \ge 1\}$ enumerate the elements of $S_0$. We construct a partition of $S$ containing subset of $\epsilon$-balls about points in $S_0$. We start by letting $C_1 = B_m(x_1, \epsilon)$. For $C_1, \ldots, C_n$ given, let $k_{n+1}$ be the index of the first point from $\{x_n : n \ge 0\}$ not contained in $\cup_{i=1}^{n} C_i$. Then let

$$C_{n+1} = B_m(x_{k_{n+1}}, \epsilon) - \cup_{i=1}^{n} C_i .$$

Let $k_1 = 1$. Now let $P_2$ attach mass $P_1(C_n)$ to point $x_{k_n}$ (in $C_n$) for $n \ge 1$. To give $P_2$ finite support, stop when $P_1(\cup_{i=1}^{k} C_i) > 1 - \gamma\epsilon$ and let $P_2$

assign the mass $P_1(\cup_{k+1}^{\infty} C_i)$ to $x_1$. Hence $P_2(\{x_1\}) = P_1(C_1) + P_1(\cup_{k+1}^{\infty} C_i)$. Now consider an arbitrary measurable set $A$. Note that $C_i \subseteq A^{\epsilon}$ whenever $A \cap C_i \neq \phi$. Since $\{C_i\}$ is a partition of $S$,

$$P_1(A) = \sum_{i=1}^{\infty} P_1(A \cap C_i) \leq \sum_{i=1}^{k} P_1(A \cap C_i) + \gamma\epsilon \leq P_2(A^{\epsilon}) + \gamma\epsilon . \quad \blacksquare$$

## 1.3. The Skorohod Representation Theorem

In this section we prove the Skorohod representation theorem, Theorem 3.2.2 in the book. We restate it here:

**Theorem 1.3.1.** (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*

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

We start by giving an elementary proof for the case in which the space $S$ is the real line. Then we give Skorohod's (1956) original proof for the case in which $S$ is a complete separable metric space. Finally, we give a proof for general separable metric spaces due to Wichura (1970). Dudley (1968) first showed that the completeness condition is not needed.

### 1.3.1. Proof for the Real Line

Suppose that $S = \mathbb{R}$. Then we can characterize the probability laws of $X$ and $X_n$, $n \geq 1$, by their cumulative distribution functions (cdf's), i.e.,

$$F(t) \equiv P(X \leq t), \quad t \in \mathbb{R} . \tag{3.1}$$

For any cdf $F$, let $F$ be its right-continuous inverse, defined as in Chapter I by

$$F^{-1}(t) = \inf\{s : F(s) > t\}, \quad 0 < t < 1 . \tag{3.2}$$

The representation is achieved by letting $\Omega = [0,1]$ with Lebesgue measure (the uniform probability distribution), $\tilde{X}(\omega) = F^{-1}(\omega)$ and $\tilde{X}_n(\omega) = F_n^{-1}(\omega)$, $n \geq 1$, with an arbitary definition for $\omega = 0$ and $\omega = 1$. The proof is based on the following four basic lemmas, the first two of which have been discussed in Sections 1.3 and 1.4 of the book.

**Lemma 1.5.** *If $F$ is a cdf on $\mathbb{R}$ and $U$ is a random variable uniformly distributed on $[0,1]$, then $F^{-1}(U)$ is a random variable with cdf $F$.*

**Lemma 1.6.** (weak convergence criterion in terms of cdf's) *Let $X$ and $X_n$ be real-valued random variables with cdf's $F$ and $F_n$ for $n \geq 1$. Then $X_n \Rightarrow X$ as $n \to \infty$ if and only if $F_n(t) \to F(t)$ as $n \to \infty$ for all $t$ that are continuity points of $F$.*

**Lemma 1.7.** *Let $F$ and $F_n$, $n \geq 1$, be cdf's on $\mathbb{R}$. Then $F_n(t) \to F(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ that are continuity points of $F$ if and only if $F_n^{-1}(t) \to F^{-1}(t)$ for all $t \in (0,1)$ that are continuity points of $F^{-1}$.*

**Lemma 1.8.** *For any cdf $F$ on $\mathbb{R}$, the set of discontinuities of $F^{-1}$ in* (3.2) *is at most countably infinite.*

### 1.3.2. Proof for Complete Separable Metric Sspaces

The proof of Theorem 1.3.1 will be based on constructing a special family of subsets of $(S, m)$ and relating these subsets to associated subintervals of the interval $[0,1)$. The length of the subinterval in $[0,1)$ (probability with respect to Lebesgue measure) will match the probability of the corresponding subset of $S$. The proof is a combination of Lemma 1.9 below, which shows the existence of the subsets with the required properties, and Lemma 1.10 below, which shows how to exploit such subsets to establish the Skorohod representation. Lemma 1.9 uses the separability; Lemma 1.10 uses the completeness.

A *partition* of a set $A$ is a collection of disjoint subsets of $A$ whose union is $A$. A *nested family of countable partitions* of a set $A$ is a collection of subsets $A_{i_1,\ldots,i_k}$ of $A$ indexed by $k$-tuples of positive integers such that $\{A_i : i \geq 1\}$ is a partition of $A$ and $\{A_{i_1,\ldots,i_{k+1}} : i_{k+1} \geq 1\}$ is a partition of $A_{i_1,\ldots,i_k}$ for all $k \geq 1$ and $(i_1, \ldots, i_k) \in \mathbb{N}_+^k$. We allow $A_{i_1,\ldots,i_k}$ to be empty for some $(i_1, \ldots, i_k)$. For each $x \in A$, there is one and only one sequence $\{i_k : k \geq 1\}$ such that $x \in A_{i_1,\ldots,i_k}$ for all $k$.

**Example 1.1.** Suppose that $S = \mathbb{R}^+$. We can obtain a nested family of countably partitions of $S$ by letting $A_i$ be $[i-1, i)$ and $A_{i_1,\ldots,i_k}$ be the set of all positive numbers with decimal expansion beginning $(i_1-1).(i_2-1),(i_3-1),\ldots,(i_k-1)$. Let $A_{i_1,\ldots,i_k} = \phi$ if $i_j > 10$ for any $j \geq 2$. ■

We say that the *radius* of a set $A$ in $S$ is less than $r$, and write $rad(A) < r$ if $A \subseteq B_m(x,r)$ for some $x \in S$, where $B_m(x,r)$ is the open ball of radius $r$ about $x$ in $(S,m)$. As before, let $\partial A$ be the boundary of $A$.

**Lemma 1.9.** *If $P$ is a probability measure on a separable metric space $(S,m)$, then there exists a nested family of countably partitions $\{S_{i_1,\ldots,i_k}\}$ of $S$ such that, for all $k$ and $(i_1,\ldots,i_k)$,*

$$(i) \quad rad(S_{i_1,\ldots,i_k}) < 2^{-k} \tag{3.3}$$

*and*

$$(ii) \quad P(\partial S_{i_1,\ldots,i_k}) = 0 \ . \tag{3.4}$$

**Proof.** Since $(S,m)$ is a separable metric space, there exists a countable dense subset, which we can express as a sequence $\{x_i : i \geq 1\}$. For each $k$, we can choose an $r_k$ such that $2^{-(k+1)} < r_k < 2^{-k}$ and

$$P(\partial B_m(x_i, r_k)) = 0 \quad \text{for all} \quad i \ , \tag{3.5}$$

because there are at most countably many $(r,i)$ such that $P(\partial B_m(x_i,r) > 0)$. Now write

$$D_i^k = B_m(x_i, r_k) - \cup_{j=1}^{i-1} B_m(x_j, r_k) \tag{3.6}$$

and

$$S_{i_1,\ldots,i_k} = D_{i_1}^1 \cap D_{i_2}^2 \cap \cdots \cap D_{i_k}^k \ . \tag{3.7}$$

Since

$$S_{i_1,\ldots,i_k} \subseteq D_{i_k}^k \subseteq B_m(x_{i_k}, r_k) \subseteq B_m(x_{i_k}, 2^{-k}) \ , \tag{3.8}$$

(3.3) holds. Since

$$\partial D_i^k \subseteq \cup_{j=1}^{i} \partial B_m(x_j, r_k) \tag{3.9}$$

and

$$\partial S_{i_1,\ldots,i_k} \subseteq \partial D_{i_1}^1 \cup \cdots \cup \partial D_{i_k}^k \subseteq \cup_{j=1}^{k} \cup_{l=1}^{i_j} \partial B_m(x_l, r_j) \ , \tag{3.10}$$

(3.5) implies that (3.4) holds. ■

**Lemma 1.10.** *Suppose that $P_0$ is a probability measure on a complete metric space $(S, m)$ with a nested family of countable partitions $\{S_{i_1,\ldots,i_k}\}$ satisfying* (3.3) *and* (3.4). *If $P_n \Rightarrow P_0$ as $n \to \infty$ on $(S, m)$, then there exist $\tilde{X}_n$, $n \geq 0$, defined on $[0,1]$ with Lebesgue measure, denoted by $P$, such that $P\tilde{X}_n^{-1} = P_n$, $n \geq 0$, and*

$$P\left(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}_0\right) = 1 \ . \tag{3.11}$$

**Proof.** We construct nested sequences of countably partitions of $[0,1)$ corresponding to the given nested sequence $\{S_{i_1,\ldots,i_k}\}$ of $(S, m)$. For $n \geq 0$, we construct subintervals $I^n_{i_1,\ldots,i_k}$ corresponding to $X_n$. We make each subinterval closed on the left and open on the right. Let $I^n_1 = [0, P_n(S_1))$ and

$$I^n_i = \left[\sum_{j=1}^{i-1} P_n(S_j), \sum_{j=1}^{i} P_n(S_j)\right], \quad i > 1 \ . \tag{3.12}$$

Let $\{I^n_{i_1,\ldots,i_{k+1}} : i_{k+1} \geq 1\}$ be a countable partition of subintervals of $I^n_{i_1,\ldots,i_k}$. If $I^n_{i_1,\ldots,i_k} = [a_n, b_n)$, then

$$I^n_{i_1,\ldots,i_{k+1}} = \left[a_n + \sum_{j=1}^{i_{k+1}-1} P_n(S_{i_1,\ldots,i_k,j}), a_n + \sum_{j=1}^{i_{k+1}} P_n(S_{i_1,\ldots,i_k,j})\right) \ . \tag{3.13}$$

The length of each subinterval $I^n_{i_1,\ldots,i_k}$ is the probability $P_n(S_{i_1,\ldots,i_k})$. Now from each nonempty subset $S_{i_1,\ldots,i_k}$ we choose one point $x_{i_1,\ldots,i_k}$. For each $n \geq 0$ and $k \geq 1$, we define functions $x^k_n : [0,1) \to S$ by letting $x^k_n(\omega) = x_{i_1,\ldots,i_k}$ for $\omega \in I^n_{i_1,\ldots,i_k}$. By the nested partition property and (3.3),

$$m(x^k_n(\omega), x^{k+j}_n(\omega)) < 2^{-k} \quad \text{for all} \quad j, k, n \tag{3.14}$$

and $\omega \in [0,1)$. Since $(S, m)$ is a complete metric space, (3.14) implies that there is $x_n \in S$ for all $n \geq 0$ such that

$$m(x^k_n(\omega), x_n(\omega)) \to 0 \quad \text{as} \quad k \to \infty \ . \tag{3.15}$$

We let $\tilde{X}_n = x_n$ on $[0,1)$ for $n \geq 0$. Since $P_n \Rightarrow P_0$ as $n \to \infty$, $P_n(A) \to P_0(A)$ as $n \to \infty$ for all $A$ for which $P_0(\partial A) = 0$ by Theorem 11.3.1 of the book. Hence, $P_n(S_{i_1,\ldots,i_k}) \to P_0(S_{i_1,\ldots,i_k})$ by (3.4). Consequently, the length of the intervals $I^n_{i_1,\ldots,i_k}$ converge to the length of the intervals $I^0_{i_1,\ldots,i_k}$

as $n \to \infty$. Since

$$\begin{aligned} m(\tilde{X}_n(\omega), \tilde{X}_0(\omega)) &\leq m(\tilde{X}_n(\omega), x_n^k(\omega)) + m(x_n^k(\omega), x_0^k(\omega)) \\ &\quad + m(x_0^k(\omega), \tilde{X}_0(\omega)) \\ &\leq 2^{-(k-1)} + m(x_n^k(\omega), x_0^k(\omega)) \;, \end{aligned} \tag{3.16}$$

for all $\omega$ in the interior of $I^0_{i_1,\ldots,i_k}$,

$$\lim_{n\to\infty} m(\tilde{X}_n(\omega), \tilde{X}_0(\omega)) \leq 2^{-(k-1)} \;. \tag{3.17}$$

Since $k$ is arbitrary, we must have $\tilde{X}_n(\omega) \to \tilde{X}_0(\omega)$ as $n \to \infty$ for all but at most countably many $\omega \in [0, \infty)$.

It remains to show that $\tilde{X}_n$ has the probability law $P_n$ for $n \geq 0$. It suffices to show that $P(\tilde{X}_n \in A) = P_n(A)$ for each $A$ such that $P_n(\partial A) = 0$. Let $A$ be such a set. Let $A^k$ be the union of the sets $S_{i_1,\ldots,i_k}$ such that $S_{i_1,\ldots,i_k} \subseteq A$ and let $A'^k$ be the union of the sets $S_{i_1,\ldots,i_k}$ such that $S_{i_1,\ldots,i_k} \cap A \neq \phi$. Then $A^k \subseteq A \subseteq A'^k$ and, by construction above,

$$P(\tilde{X}_n \in A^k) = P_n(A^k) \quad \text{and} \quad P(\tilde{X}_n \in A'^k) = P_n(A'^k) \;. \tag{3.18}$$

Now let

$$C^k = \{x \in S : m(x, \partial A) \leq 2^{-k}\} \;. \tag{3.19}$$

Then $A'^k - A^k \subseteq C^k \downarrow \partial A$ as $k \to \infty$. Since $P_n(\partial A) = 0$ by assumption, $P_n(C^k) \downarrow 0$ as $k \to \infty$. Hence

$$P(\tilde{X}_n \in A) = \lim_{k\to\infty} P(\tilde{X}_n \in A^k) = \lim_{k\to\infty} P_n(A^k) = P_n(A) \;. \quad \blacksquare \tag{3.20}$$

### 1.3.3. Proof for Separable Metric Spaces

We now do the proof of Theorem 1.3.1 without assuming completeness. Start by letting $P_n$ be the probability distribution of $X_n$ on $S$ for $n \geq 0$. Let the underlying probability space be the product space $\Omega \equiv S^\infty$ with elements $\omega \equiv \{s_k : k \geq 0\}$. Let $\tilde{X}_n$ be the coordinate mapping, e.g., $\tilde{X}_n(\{s_k : k \geq 0\}) = s_n$, $n \geq 0$. To quickly get the idea, first suppose that $P_n(\{s\}) = 1$ for all $n \geq 0$. In this special case we can let the probability measure $P$ on $\Omega$ be the product measure $P = \delta_s \times \delta_s \times \cdots$, where $\delta_s$ is the Dirac measure assigning probability 1 to the point $s \in S$. Then $P$ assigns probability 1 to the sequence $\{s_n : n \geq 0\}$ where $s_n = s$ for all $n$. Since $P(\tilde{X}_n = s) = 1$ for all $n$,

$$P(\tilde{X}_n = \tilde{X}_0 \quad \text{for all} \quad n) = P\left(\cap_{n=0}^{\infty}\{\tilde{X}_n = s\}\right) = 1 \;. \tag{3.21}$$

To continue to develop the idea of the approach, now suppose that each probability measure $P_n$, $n \geq 0$, concentrates all probability on a common finite subset of $S$. Thus it suffices to assume that $S$ is finite. For a sequence $\{k_n : n \geq 1\}$ with $k_n \to \infty$ as $n \to \infty$ to be defined later, let

$$U_k = \cap_{n:k_n \geq k}\{\tilde{X}_n = \tilde{X}_0\} \ . \tag{3.22}$$

(Note that we have a strong form of convergence on $U_k$.) Also let $\{Q_n : n \geq 1\}$ be a sequence of probability measures on $S$ to be defined later. Now let $P_{j,s}$ be the product measure

$$P_{j,s} = \delta_s \times \prod_{n=1}^{\infty} P_{j,s,n} \ , \tag{3.23}$$

where $P_{j,s,n}$ is a probability measure on $S$ defined by

$$P_{j,s,n} = \begin{cases} Q_n & \text{if} \quad 0 \leq k_n < j \\ \delta_s & \text{if} \quad j \leq k_n \leq \infty \ . \end{cases} \tag{3.24}$$

Then let $P'_j$ be a mixture of the probabilities $P_{j,s}$ in (3.23) with respect to $P_0$, in particular,

$$P'_j = \sum_{s \in S} P_0(\{s\}) P_{j,s} \ . \tag{3.25}$$

Next let $\{w_k : k \geq 1\}$ and $\{q_k : k \geq 0\}$ be sequences of numbers with

$$w_k \geq 0, \quad \sum_{k=1}^{\infty} w_k = 1, \quad q_0 = 0, \quad q_k = \sum_{j=1}^{k} w_j < 1, \quad 1 \leq k < \infty \ . \tag{3.26}$$

Then let $P$ be a mixture of the probabilities $P'_j$ in (3.25) using the weights $w_j$ in (3.26), i.e.,

$$P = \sum_{j=1}^{\infty} w_j P'_j \ . \tag{3.27}$$

We will show that this construction does the job with an appropriate choice of the sequences $\{k_n : n \geq 1\}$ and $\{Q_n : n \geq 1\}$. (The weights $w_k$ in (3.26) can be arbitrary subject to the conditions in (3.26).)

Note that $P_{j,s}$ in (3.23) attaches positive probability only to sets of sequences $\{s_n : n \geq 0\}$ such that $s_n = s$ for all but a finite number of $n$ (those $n$ for which $0 \leq k_n < j$). Thus even though $S^\infty$ is uncountably infinite, $P_{j,s}$ has finite support. Since $S$ is finite, $P'_j$ in (3.25) also has finite support. All sequences $\{s_n : n \geq 0\}$ in $S^\infty$ with positive $P$-measure have $s_n = s$ for all sufficiently large $n$ for some $s$.

By (3.23), $P_{j,s}(\tilde{X}_0 = s) = 1$. Thus, by (3.25) and (3.27), $P(\tilde{X}_0 = s) = P_j'(\tilde{X}_0 = s) = P_0(\{s\})$ for all $s \in S$. Hence $P\tilde{X}_0^{-1} = P_0$ or, equivalently, $\tilde{X}_0 \stackrel{\text{d}}{=} X_0$.

Next $P_{j,s}(\tilde{X}_n = s) = P_{j,s,n}(\{s\})$ for $n \geq 1$. Note that $P_{j,s}(U_k) = 1$ for $j \leq k$, where $U_k$ is given in (3.22), so that $P_j'(U_k) = 1$ if $j \leq k$ and $P(U_k) \geq q_k$. Since $q_k \to 1$ as $k \to \infty$ by (3.26), $\tilde{X}_n \to \tilde{X}_0$ as $n \to \infty$ almost uniformly on $\Omega$ with respect to $P$, i.e., for any $\epsilon > 0$, there exists a subset $U_k$ of $S$ with $P(U_k) > 1 - \epsilon$ such that $\tilde{X}_n$ converges uniformly to $\tilde{X}_0$ as $n \to \infty$ on $U_k$. (In our finite-state-space setting, we actually have $\tilde{X}_n = \tilde{X}_0$ on $U_k$ for all $n$ such that $k_n \geq k_0$ by (3.22) and (3.26).) For $\epsilon$ given, choose $k$ so that $q_k > 1 - \epsilon$. By Egoroff's theorem, p. 89 of Halmos (1950), that implies that

$$P\left(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}_0\right) = 1 \; . \tag{3.28}$$

The difficult part is to obtain $\tilde{X}_n \stackrel{\text{d}}{=} X_n$ for $n \geq 1$. The construction above yields

$$P(\tilde{X}_n = s) = q_{k_n} P_0(\{s\}) + (1 - q_{k_n}) Q_n(\{s\}) \quad \text{for all} \quad n \; . \tag{3.29}$$

We now choose the sequences $\{k_n : n \geq 1\}$ and $\{Q_k : k \geq 1\}$ to achieve $P\tilde{X}_n^{-1} = P_n$ for all $n$. Note that (3.29) is equivalent to

$$Q_n(\{s\}) = P_n(\{s\}) + \frac{q_{k_n}}{1 - q_{k_n}} (P_n(\{s\}) - P_0(\{s\}) \tag{3.30}$$

provided $k_n < \infty$. If $k_n = \infty$, then $q_{k_n} = 1$, so that we must have $P_n(\{s\}) = P_0(\{s\})$, and then any $Q_n(\{s\})$ will do.

Thus, let

$$\begin{aligned} Q(k, s, n) &\equiv P_n(\{s\}) + \frac{q_k}{1 - q_k} (P_n(\{s\}) - P_0(\{s\}) \; , && (3.31) \\ m_{k,n} &\equiv \min_{s \in S} Q(k, s, n) \; , && (3.32) \\ k_n &= \sup\{j \geq 0 : m_{j,n} \geq 0\} && (3.33) \end{aligned}$$

and

$$Q_n(\{s\}) = Q(k_n, s, n) \quad \text{for} \quad k_n < \infty \; . \tag{3.34}$$

Note that $m_{0,n} \geq 0$, so that $k_n \leq \infty$ is well defined in (3.33). Note that $\sum_{s \in S} Q(k, s, n) = 1$ for all $k$, $0 \leq k < \infty$, and $Q(k_n, s, n) \geq 0$ by (3.32) and (3.33). Thus, under (3.31)–(3.34), $Q_n$ is a probability measure on $S$ satisfying (3.29) provided that $k_n < \infty$.

Since $\sum_{s \in S} Q(k,s,n) = 1$, we must have $0 \le Q(k,s,n) \le 1$ for $Q(k,s,n)$ in (3.31). Since $q_k \to 1$ as $k \to \infty$, $q_k/(1-q_k) \to \infty$ as $k \to \infty$. Hence, we must have $P_n(\{s\}) = P_0(\{s\})$ for all $n$ if $k_n = \infty$, under which (3.29) has been shown to hold for any probability measure $Q_n$.

We now show that $k_n \to \infty$ as $n \to \infty$. Since $P_n(\{s\}) \to P_0(\{s\})$ as $n \to \infty$ for each $s$, $Q(k,s,n) \to P_0(\{s\})$ as $n \to \infty$ for each $s$ and $k$, $1 \le k < \infty$. This, together with the fact that $Q(k,s,n) \ge 0$ if $P_0(\{s\}) = 0$, implies that $Q(k,s,n)$ is ultimately nonnegative for all sufficiently large $n$ depending upon $k$. Thus, for each $k$, there is an index $n_k$ such that $Q(k,s,n) \ge 0$, and thus $m_{k,n} \ge 0$, for all $n \ge n_k$. Since $m_{k,n} \ge 0$ implies $k_n \ge k$, we can conclude that, for all $n \ge n_k$, $k_n \ge k$. Hence, $k_n \to \infty$ as $n \to \infty$.

We now turn to the general case: We now assume that $S$ is a separable metric space. We start by constructing a finite collection of subsets appropriately approximating $S$. This step is a minor modification of Lemma 1.9.

**Lemma 1.11.** *If $P$ is a probability measure on separable metric space $(S,m)$, then for any $\delta$, $\epsilon > 0$ there exist disjoint subsets $S_{i_1,\ldots,i_k}$ of $S$, $1 \le i_j \le i'_j$, $1 \le j \le k$, such that, for all $k$ and $(i_1,\ldots,i_k)$, (3.3) holds for $2^{-k} < \delta$, (3.4) holds and*

$$P\left(\cup_{i_1=1}^{i'_1} \cdots \cup_{i_k=1}^{i'_k} S_{i_1,\ldots,i_k}\right) > 1-\epsilon \ . \tag{3.35}$$

**Proof.** We use the construction in Lemma 1.9. Choose $i'_1$ such that $P(S_1 \cup \cdots \cup S_{i'_1}) > 1 - \epsilon 2^{-1}$; choose $i'_2$ such that

$$P(S_{i_1,1} \cup \cdots \cup S_{i_1,i'_2}) > 1 - P(S_{i_1})\epsilon 2^{-2} \tag{3.36}$$

for all $i_1$, $1 \le i_1 \le i'_1$; choose $i'_{j+1}$ such that

$$P(S_{i_1,\ldots,i_j,1} \cup \cdots \cup S_{i_1,\ldots,i_j,i'_{j+1}}) > 1 - P(S_{i_1,\ldots,i_j})\epsilon 2^{-j} \tag{3.37}$$

for all $(i_1,\ldots,i_j) \le (i'_1,\ldots,i'_j)$. Stop at $k$ with $2^{-k} < \delta$, so that (3.3) holds. Then

$$P\left(\cup_{i_1=1}^{i'_1} \cdots \cup_{i_k=1}^{i'_k} S_{i_1,\ldots,i_k}\right) > 1 - \epsilon(2^{-1} + \cdots + 2^{-k}) > 1-\epsilon \ , \tag{3.38}$$

so that (3.35) holds. ∎

We now return to the proof of the theorem. Let $\{\delta_k : k \geq 1\}$ and $\{\epsilon_k : k \geq 1\}$ be sequences of positive numbers such that $\delta_k \to 0$, $\epsilon_k \to 0$ and $\sum_{k=1}^{\infty} \epsilon_k < \infty$. For each $k$, let $\{C_{k,j} : 0 \leq j \leq n_k\}$ be the finite collection of subsets $S_{i_1,\ldots,i_k}$ in Lemma 1.11 constructed with respect to $P_0$, where $\delta$ and $\epsilon$ for $k$ are required to be $\delta_k$ and $\epsilon_k$. Let $C_{k,0} = S - \cup_{j=1}^{n_k} C_{k,j}$. By (3.35), $P_0(C_{k,0}) < \epsilon_k$.

With $\tilde{X}_n$ the coordinate projections on $S^{\infty}$ as before, instead of (3.22), let

$$U_k = \cap_{n:k_n \geq k} \{m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}\} \tag{3.39}$$

where $\delta_{\infty} = 0$. (The separability of $(S, m)$ is used to have $\{m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}\}$ and thus $U_k$ be measurable.) Given that $k_n \to \infty$ as $n \to \infty$, $\tilde{X}_n \to \tilde{X}_0$ uniformly on $U_k$. To apply Egoroff's theorem, we will need to show that $P(U_k) \to 1$ as $k \to \infty$.

Let $\Pi_k$ be the collection of sets $C_{k,j}$, $1 \leq j \leq n_k$, and let $\Pi_0 = S$. We now modify the finite-state-space proof above, letting $C_{k,j}$ play the role of $s$. Let the weights $w_k$ and their partial sums $q_k$ be defined by (3.26). Paralleling (3.31)–(3.34), for $0 \leq k < \infty$, let

$$\begin{aligned} Q(k, C, n) &= P_n(C) + \frac{q_k}{1 - q_k}(P_n(C) - P_0(C)) \ , && (3.40) \\ m_{k,n} &= \min_{C \in \Pi_k} \{Q(k, C, n)\} \ , && (3.41) \\ k_n &= \sup\{j \geq 0 : m_{j,n} \geq 0\} && (3.42) \end{aligned}$$

and

$$Q_n(C) = Q(k_n, C, n) \ . \tag{3.43}$$

Since $P_n(C) \to P_0(C)$ as $n \to \infty$ for all $C \in \Pi_k$, $k_n \to \infty$ as $n \to \infty$ by the same argument as before.

Paralleling (3.23), let $P_{j,s}$ be the product measure

$$P_{j,s} = \delta_s \times \prod_{n=1}^{\infty} P_{j,s,n} \ , \tag{3.44}$$

where $P_{j,s,n}$ is a probability measure on $S$ defined by

$$P_{j,s,n} = \begin{cases} Q_n & \text{if } \ 0 \leq k_n < j \\ P_n(\cdot | C_{k_n,s}) & \text{if } \ j \leq k_n < \infty \\ \delta_s & \text{if } \ k_n = \infty \ , \end{cases} \tag{3.45}$$

where $P_n(\cdot|C_{k_n,s})$ is the conditional probability measure with $C_{k,s}$ being the element of $\Pi_k$ containing $s \in S$. Note that $P_{j,s,n}$ in (3.45) has three possibilities instead of only the two in (3.24). Unlike the case of finite $S$, $P_{j,s}$ in (3.44) does not have finite support, but if $s \in C_{k_n,i}$, then $P_{j,s}$ has support on the set of sequences $\{s_n : n \geq 0\}$ such that $s_n \in C_{k_n,i}$ for all but finitely many $n$, in particular, for all $n$ such that $k_n \geq j$. On this subset of sequences, $m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}$ for all $n$ such that $k_n \geq j$.

Paralleling (3.25), let

$$P'_j(A) = \int_S P_0(ds) P_{j,s}(A) \ . \tag{3.46}$$

The integral in (3.46) is well defined since $P_{j,s}(A)$ is a measurable function on $S$ for each $A$ a cylinder set with finite base in the $\sigma$-field on $S^\infty$; see pp. 74–76 of Neveu (1965). Note that $P'_j$ has support on the set of sequences $\{s_n : n \geq 0\}$ such that $s_n \in C_{k_n,i}$ for all but finitely many $n$, for some $i$. Thus

$$P'_j(m(\tilde{X}_n, \tilde{X}_0) \leq \delta_{k_n}) \geq 1 - P(C_{k_n,0}) > 1 - \epsilon_{k_n} \ . \tag{3.47}$$

Paralleling (3.27), let

$$P = \sum_{j=1}^{\infty} w_j P'_j \ . \tag{3.48}$$

As before, the construction yields $P\tilde{X}_0^{-1} = P_0$. The probability distribution of $\tilde{X}_n$ is

$$P\tilde{X}_n^{-1} = \begin{cases} q_{k_n} \sum_{C \in \Pi_{k_n}} P_n(\cdot|C) P_0(C) + (1 - q_{k_n}) Q_n & \text{if} \quad k_n < \infty \\ P_0 & \text{if} \quad k_n = \infty \ . \end{cases} \tag{3.49}$$

For $n$ such that $k_n < \infty$, let

$$Q_n = \sum_{C \in \Pi_k} Q(k_n, C, n) P_n(\cdot|C) \ . \tag{3.50}$$

Combining (3.40), (3.49) and (3.50), we see that $P\tilde{X}_n^{-1} = P_n$ if $k_n < \infty$. On the other hand, as before, if $k_n = \infty$, then we are forced to have $P_n(C) = P_0(C)$ for all $C \in \Pi_k$ for any $k \geq 1$, but that implies that $P_n = P_0$. (We can apply the reasoning in the proof of Lemma 1.10 using (3.18) and (3.19).)

Finally, it remains to show that $P(U_k) \to 1$ as $k \to \infty$ for $U_k$ in (3.39). However,

$$\begin{aligned} P(U_k) &= \sum_{j=1}^{\infty} w_j P'_j(U_k) \geq \sum_{j=1}^{k} w_j P'_j(U_k) \\ &\geq \left(\sum_{j=1}^{k} w_k\right) \left(1 - \sum_{j=k}^{\infty} \epsilon_j\right) \to 1 \quad \text{as} \quad k \to \infty \ , \end{aligned} \tag{3.51}$$

since, for $j \leq k \leq k_n$,

$$1 - P_j'(U_k) \leq P_0\left(\cup_{l=k}^{\infty} C_{l,0}\right) \leq \sum_{l=k}^{\infty} \epsilon_l \ , \tag{3.52}$$

because $P_{j,s}'$ assigns probability 1 to product sets in which all coordinates are in common sets $C_{i,k}$.

## 1.4. The "Weak" in Weak Convergence

This section is devoted, not to a proof of a theorem, but to an expanation of a term – the adjective "weak" in "weak convergence." The term "weak" can be understood from a Banach-space perspective.

The starting point is the definition of convergence $P_n \Rightarrow P$; i.e., $P_n \Rightarrow P$, if

$$\lim_{n\to\infty} \int_S f dP_n = \int_S f dP \tag{4.1}$$

for all functions $f$ in $C(S)$, the space of all continuous bounded real-valued functions on $S$.

The space $C(S)$ of continuous bounded real-valued functions $h$ on $S$ used in definition (4.1) is a Banach space (a complete normed linear topological space) with the uniform norm

$$\|h\| \equiv \sup_{s\in S} |h(s)| \ .$$

The adjoint or conjugate space of $C(S)$, the space of all continuous linear real-valued functions $L$ on $C(S)$, denoted by $C^*(S)$, turns out to be the space $Z(S)$ of all finite signed measures $\mu$ on $S$, defined via

$$L(h) \equiv \int_S h d\mu \ ;$$

e.g., see pp. 262, 419 of Dunford and Schwartz (1958) or Chapter 9 of Simmons (1963).

The adjoint space $B^*$ of any Banach space $B$ is itself a Banach space with the norm

$$\|L\| \equiv \sup\{\|L(b)\| : b \in B, \quad \|b\| \leq 1\} \ .$$

Since $B^*$ is a Banach space, one can consider its adjoint space $B^{**}$. There is a natural embedding of $B$ in $B^{**}$ so that we can regard $B$ as a subset of

$B^{**}$. (Just let $L_b(f) = f(b)$ for $b \in B$ and $f \in B^*$.) When $B = B^{**}$, $B$ is said to be *reflexive*. However, $C(S)$ is reflexive only when $S$ is finite. So, in our setting with infinite $S$, $C(S)$ is a proper subset of $C^{**}(S)$.

Instead of the topology induced on a Banach space $B$ by its norm, it is sometimes of interest to consider a weaker topology on $B$ called the *weak topology*, which is the weakest topology such that all the functions in $B^*$ remain continuous; i.e., $b_n \to b$ in $B$ with the weak topology if and only if $L(b_n) \to L(b)$ for all $L$ in $B^*$. Furthermore, on the adjoint space $B^*$ one can also consider a still weaker topology called the weak* topology, which is the weakest topology such that all the functions in $B$, regarded as a subset of $B^{**}$, remain continuous. Thus the weak* topology on $Z(S) = C^*(S)$ relativized to the subset $P(S)$ is what is characterized by (4.1). (The discussion also implies that the weak topology on $Z(S)$ is stronger than the weak* topology on $Z(S)$, so the terminology "weak convergence" is something of a misnomer. From this Banach-space perspective, we should actually call weak convergence $P_n \Rightarrow P$ *weak** convergence.) ∎

## 1.5. Continuous-Mapping Theorems

In this section we supplement the discussion of the continuous-mapping approach in Section 3.4 of the book by providing proofs for the unproved theorems. We first prove the Lipschitz mapping theorem, which comes from Whitt (1974).

### 1.5.1. Proof of the Lipschitz Mapping Theorem

We now prove the Lipschitz mapping theorem, Theorem 2.4.2 in the book. First suppose that $(S, m)$ is a separable metric space and $B = S$. Then we can employ the Strassen representation theorem, Theorem 11.3.5 in the book. It is elementary that the Lipschitz property is inherited by the in-probability distance $p$: Given $P(m(X,Y) > \delta) < \delta$, the Lipschitz property of $g$ implies that $P(m'(g(X), g(Y)) > K\delta) < \delta$, so that $p(g(X), g(Y)) \le (K \vee 1)p(X,Y)$. By the Strassen representation theorem, for $X, Y$ and positive $\epsilon$ given, we can find $\tilde{X}, \tilde{Y}$ on a common probability space so that $\tilde{X} \stackrel{d}{=} X$, $\tilde{Y} \stackrel{d}{=} Y$ and

$$p(\tilde{X}, \tilde{Y}) \le \pi(X, Y) + \epsilon \ .$$

Hence,

$$\pi(g(X), g(Y)) = \pi(g(\tilde{X}), g(\tilde{Y})) \le p(g(\tilde{X}), g(\tilde{Y}))$$

and

$$p(g(\tilde{X}), g(\tilde{Y})) \leq (K \vee 1) p(\tilde{X}, \tilde{Y}) \leq (K \vee 1)(\pi(X, Y) + \epsilon) .$$

Since $\epsilon$ was arbitrary, we have the desired conclusion.

Now we consider the general case, for which we argue directly. Let $B$ be the subset for which $P(Y \in B) = 1$. The Lipschitz property implies that

$$B \cap g^{-1}(A)^{\delta} \subseteq g^{-1}(A^{\epsilon}) \quad \text{in} \quad S \quad \text{for} \quad \delta \leq \epsilon/K$$

and any $A \in \mathcal{S}'$. Hence,

$$\begin{aligned}
&\pi(g(X), g(Y)) \\
&= \inf\left\{\epsilon > 0 : P(g(X) \in A) \leq \epsilon + P(g(Y) \in A^{\epsilon}) \text{ for all } A \in \mathcal{S}'\right\} \\
&= \inf\left\{\epsilon > 0 : P(X \in g^{-1}(A)) \leq \epsilon + P(Y \in g^{-1}(A^{\epsilon})) \text{ for all } A \in \mathcal{S}'\right\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in g^{-1}(A) \leq \epsilon + P(Y \in B \cap g^{-1}(A)^{\delta}) \text{ for all } A \in \mathcal{S}'\right\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in g^{-1}(A) \leq \epsilon + P(Y \in g^{-1}(A)^{\delta}) \text{ for all } A \in \mathcal{S}'\right\} \\
&\leq \inf\left\{\epsilon > 0 : P(X \in A) \leq \epsilon + P(Y \in A^{\delta}) \text{ for all } A \in \mathcal{S}\right\} \\
&\leq (1 \vee K)\pi(X, Y) . \quad \blacksquare
\end{aligned}$$

**Example 1.5.1.** *The advantage of the Prohorov metric on $\mathcal{P}(\mathbb{R})$.* Even on the real line $\mathbb{R}$, the Prohorov metric is useful to establish rate of convergence results, because the Lipschitz mapping theorem does not apply to two other metrics commonly used. On $\mathcal{P}(\mathbb{R})$ one often uses the *Lévy metric* $\lambda$, which is defined just as the Prohorov metric $\pi$ in (2.2) except that only sets of the form $A = (-\infty, x]$ are used. The *uniform metric for cdf's* is also sometimes used; i.e.,

$$\|F_1 - F_2\| \equiv \mu(P_1, P_2) \equiv \sup\{(|P_1(A) - P_2(A)| : A = (-\infty, x]\} ,$$

where $F_i(x) \equiv P((-\infty, x])$. The uniform-cdf metric $\mu$ also induces weak convergence at limiting probability measures without atoms. However, the Lipschitz theorem is not valid for $\lambda$ and $\mu$. To see that, for $n \geq 1$, let

$$P(X_n = 2j) = P(Y_n = 2j + 1) = 1/n \quad \text{for} \quad 1 \leq j \leq n ,$$

and let $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$g(t) = \sin(\pi x/2) \quad \text{for} \quad t \in \mathbb{R} .$$

Clearly, $g$ is Lipschitz with Lipschitz constant 1, but

$$\lambda(X_n, Y_n) \leq \mu(X_n, Y_n) = 1/n ,$$

while

$$P(g(X_n) = 0) = P(g(Y_n) \in \{-1, 1\}) = 1 \quad \text{for all} \quad n ,$$

so that

$$\mu(g(X_n), g(Y_n)) \geq \lambda(g(X_n), g(Y_n)) = 1/2 \quad \text{for all} \quad n .$$

Given a bound with the Prohorov metric $\pi$ in $\mathcal{P}(\mathbb{R})$, we can obtain corresponding bounds with the metrics $\lambda$ and $\mu$. First we use the inequality $\lambda \leq \pi$. In many cases we can relate $\lambda$ and $\mu$: When a probability measure $P_1$ on $\mathbb{R}$ has a Lipschitz cdf $F_1$ with Lipschitz constant $c$, i.e., when

$$|F_1(t_1) - F_1(t_2)| \leq c|t_1 - t_2| ,$$

then we have the ordering

$$\mu(P_1, P_2) \leq (1 + c)\lambda(P_1, P_2) \quad \text{for all} \quad P_2 \in \mathcal{P}(\mathbb{R}) . \quad \blacksquare \tag{5.1}$$

### 1.5.2. Proof of the Continuous-Mapping Theorems

We now turn to Theorem 3.4.3 of the book, following Billingsley (1968, Section 5), which we restate here. Let $Disc(g)$ be the set of discontinuity points of the function $g$.

**Theorem 1.5.1.** (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 $g(X_n) \Rightarrow g(X)$.*

We first establish the measurability of $Disc(g)$ (even if $g$ is not measurable).

**Lemma 1.5.1.** (measurability of the set of discontinuity points) *For $g : (S, m) \to (S', m')$, $Disc(g) \in \mathcal{S}$.*

**Proof.** For any $y, z \in S$ with $m'(g(y), g(z)) \geq \epsilon$ and $\epsilon > 0$, let

$$A_{\epsilon,\delta}(y, z) \equiv \{x \in S : m(x, y) < \delta \text{ and } m(x, z) < \delta\} .$$

Then the complement is

$$A^c_{\epsilon,\delta}(y, z) \equiv \{x \in S : m(x, y) \geq \delta \text{ or } m(x, z) \geq \delta\} .$$

It is easy to see that $A^c_{\epsilon,\delta}(y,z)$ is closed, so that $A_{\epsilon,\delta}(y,z)$ is open, as necessarily is

$$A_{\epsilon,\delta} \equiv \bigcup_y \bigcup_z A_{\epsilon,\delta}(y,z) \ .$$

Since

$$Disc(g) = \bigcup_\epsilon \bigcap_\delta A_{\epsilon,\delta} \ ,$$

where $\epsilon$ and $\delta$ run over the positive rationals, $Disc(g)$ is a $G_{\delta\sigma}$, implying that $Disc(g) \in \mathcal{S}$. ■

**Proof of Theorem 1.5.1.** By Theorem 11.3.4 (iii) in the book, it suffices to show that

$$\overline{\lim_{n\to\infty}} \ P(g(X_n) \in F) \le P(g(X) \in F)$$

for each closed subset $F \in \mathcal{S}'$. Given that $X_n \Rightarrow X$, we have

$$\begin{aligned} \overline{\lim_{n\to\infty}} \ P(g(X_n) \in F) &= \overline{\lim_{n\to\infty}} \ P(X_n \in g^{-1}(F)) \\ &\le \overline{\lim_{n\to\infty}} \ P(X_n \in g^{-1}(F)^-) \\ &\le P(X \in g^{-1}(F)^-) \ . \end{aligned}$$

However, $P(X \in g^{-1}(F)^-) = P(X \in g^{-1}(F))$ because $P(Disc(g)) = 0$ and $g^{-1}(F)^- \subseteq g^{-1}(F) \cup Disc(g)$. ■

Finally, we treat Theorem 3.4.4 of the book, involving a sequence of measurable mappings:

**Theorem 1.5.2.** (generalized continuous-mapping theorem) *Let $g$ and $g_n$, $n \ge 1$, 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 $\{x_n : n \ge 1\}$ 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 $(S',m')$.*

Here we need to assume that the range is a separable metric space. Again we follow Billingsley (1968, Section 5).

**Lemma 1.5.2.** (measurability of the bad set) *Suppose that $g_n$, $n \ge 1$, and $g$ are measurable functions from a metric space $(S,m)$ into a separable metric space $(S',m')$. Let $E$ be the set of $x$ in $S$ such that $g_n(x_n) \to g(x)$ fails for some sequence $\{x_n : n \ge 1\}$ with $m(x_n,x) \to 0$ as $n \to \infty$. Then $E$ is a measurable subset of $S$.*

**Proof.** Let $B_{\epsilon,\delta,i}$ be the set of $x$ in $S$ such that $m'(g(x), g_i(y)) \geq \epsilon$ for some $y$ with $m(x, y) < \delta$. Note that

$$E = \cup_\epsilon \cap_\delta \cap_{k\geq 1} \cup_{i\geq k} B_{\epsilon,\delta,i} \tag{5.2}$$

where $\epsilon$ and $\delta$ range over the positive rationals. We would be done if we could conclude that $B_{\epsilon,\delta,i}$ is measurable, but we do not know that. Note that $B_{\epsilon,\delta,i}$ is decreasing in $\epsilon$. Hence (5.2) remains true if $B_{\epsilon,\delta,i}$ is replaced by $B_{\epsilon/2,\delta,i}$. It thus suffices to show that, for all $(\epsilon, \delta, i)$, there are sets $C_{\epsilon,\delta,i} \in \mathcal{S}$ such that

$$B_{\epsilon,\delta,i} \subseteq C_{\epsilon,\delta,i} \subseteq B_{\epsilon/2,\delta,i} \ . \tag{5.3}$$

Since $(S', m')$ is separable, we can find a sequence $\{u_k : k \geq 1\}$ dense in $S'$. Let $A_{\epsilon,k} = \{x : m'(g(x), u_k) < \epsilon/4\}$ and note that $A_{\epsilon,k} \in \mathcal{S}$ and $S = \cup_k A_{\epsilon,k}$. Then (5.3) holds if

$$C_{\epsilon,\delta,i} = \cup_k (A_{\epsilon,k} \cap J_{\epsilon,\delta,i,k}) \ ,$$

where $J_{\epsilon,\delta,i,k}$ is the set of $x$ such that $m'(g_i(y), g(z)) \geq \epsilon$ for some pair of points $y, z$ in $S$ with $m(x, y) < \delta$, $m(x, z) < \delta$ and $z \in A_{\epsilon,k}$. It is not difficult to see that $J^c_{\epsilon,\delta,i,k}$ is closed, so that $J_{\epsilon,\delta,i,k}$ is open and $C_{\epsilon,\delta,i} \in \mathcal{S}$. ∎

**Proof of Theorem 1.5.2.** By Lemma 1.5.2, $E \in \mathcal{S}$. From Theorem 11.3.4 (iv) in the book, it suffices to show that

$$P(g(X) \in G) \leq \varliminf_{n\to\infty} P(g_n(X_n) \in G)$$

for every open $G$ in $S'$. If $x \in E^c$ and $g(x) \in G$, then there must exist $k$ and $\delta$ such that $g_i(y) \in G$ if $i \geq k$ and $m(x, y) < \delta$, so that $x \in T_k^o$, the interior of $T_k$, where

$$T_k = \cap_{i\geq k} g_i^{-1}(G) \ .$$

Consequently,

$$g^{-1}(G) \subseteq E \bigcup \cup_{k=1}^{\infty} T_k^o \ .$$

Since $P(X \in E) = 0$ and $T_k^o \subseteq T_{k+1}^o$, for any given $\epsilon$ there is a $k$ such that

$$P(X \in g^{-1}(G)) \leq P(X \in \cup_k T_k^o) \leq P(X \in T_k^o) + \epsilon$$

for $k \geq k_0$. Since $X_n \Rightarrow X$ and $T_k \subseteq g_n^{-1}(G)$ for $n \geq k$,

$$P(X \in T_k^o) \leq \varliminf_{n\to\infty} P(X_n \in T_k^\circ) \leq \varliminf_{n\to\infty} P(X_n \in g_n^{-1}(G)) \ .$$

Since $\epsilon$ was arbitrary, the proof is completed by combining these two strings of inequalities. ■

The continuous-mapping approach to stochastic-process limits leads us to focus on the underlying sample paths of the stochastic processes. Thus the continuous-mapping approach is a *sample-path method.* In recent years, probabilists have tended to favor sample-path methods over more traditional analytic methods, because they are less removed from the phenomenon under study. However, the two approaches often can be fruitfully combined.

Many traditional analytic results are based on transforms, such as the characteristic function (version of the Fourier transform), probability generating function (or $z$ transform) and the Laplace transform, as can be seen from Feller (1971). Fortunately, the analytic approach has been applied with great success over the years to yield explicit expressions for many probability distributions of interest in the form of transforms. That is true for many of the limit processes that we will consider. Thus we can use previous analytic results to obtain explicit transforms for approximating distributions. We then can apply numerical transform inversion to compute the probability distribution itself; e.g., see Abate and Whitt (1992, 1995), Choudhury, Lucantoni and Whitt (1994) and Abate, Choudhury and Whitt (1999).

For example, as shown in Section 7.5 of the book and Section 5.2 here, the heavy-traffic limit for a queue with heavy-tailed distributions is often a reflection of a stable Levy motion or more general Levy process without negative jumps. These limit processes are somewhat complicated, but fortunately the analytic approach has shown that the steady-state distribution has a relatively simple expression via its Laplace transform, which is known as the generalized Pollaczek-Khintchine transform. Thus we can calculate the steady-state distribution of the limit process by applying numerical transform inversion.