Chapter 11

More on the Mathematical Framework

11.1. Introduction

In this chapter we discuss the mathematical framework for stochastic-process limits, expanding upon the introduction in Chapter 3. For more details and discussion, see Billingsley (1968, 1999) and Parthasarathy (1967).

Throughout, we formalize the notion of convergence using topologies. Hence we start in Section 11.2 by reviewing basic topological concepts. In Section 11.3 we discuss the topology on the space $\mathcal{P}$ of all probability measures on a general metric space, expanding upon the introduction in Section 3.2.

In Section 11.4 we review basic properties of product spaces. We describe simple criteria for the joint convergence of random elements and we state the important (even if elementary) convergence-together theorem, which is used in many proofs.

In Section 11.5 we discuss the function space D containing the stochastic-process sample paths, expanding upon the introduction in Section 3.3. We introduce the other two Skorohod (1956) topologies – J_2 and M_2 – and provide additional details.

In Section 11.6 we briefly describe the standard approach to establish stochastic-process limits based on compactness and the convergence of the finite-dimensional distributions. The compactness approach complements the continuous-mapping approach described in Section 3.4.

11.2. Topologies

In this book we focus on stochastic-process limits, i.e., the convergence of a sequence of stochastic processes to a limiting stochastic process. We use *topology* to formalize that notion. Indeed, to a large extent, this is a book about topology. Of course, we use "topology" in the mathematical sense rather than the networking sense: Here we characterize the convergence of sequences of abstract objects (stochastic processes), rather than evaluate alternative configurations of nodes and links in a communication network.

11.2.1. Definitions

In particular, we define a topology on a set of stochastic processes. To explain, we briefly review basic topological concepts. Most of the concepts can be found in any introductory book on topology; e.g., see Simmons (1963) and Dugundji (1967).

A common way to define a topology on a set is via a metric, as defined in Section 3.2. In a metric space, we can regard the topology as a specification of which sequences converge. A direct definition of a topology involves subsets of the set S. We assume familiarity with elementary set theory. It is important to distinguish between *elements* and *subsets* of a set: When x is an element of a set S, we write $x \in S$; then $\{x\}$ is a subset of S and we write $\{x\} \subseteq S$. We write A^c, $A \cap B$, $A \cup B$ and $A - B \equiv A \cap B^c$ for complement, intersection, union and difference, respectively.

So here is the direct definition: A *topological space* is a nonempty set S together with a *topology* $\mathcal{T}$, with the topology $\mathcal{T}$ being a collection (set) of subsets of S called *open sets* satisfying certain axioms. In particular, a topology is any collection of subsets, including the whole set S and the empty set ϕ, that is closed under arbitrary unions and finite intersections. (Thus $S \in \mathcal{T}$.) By *closed under arbitrary unions*, we mean that arbitrary unions of sets in the topology are also in the topology.

Every metric determines a topology generated by (the smallest topology containing) the *open balls*

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

for $x \in S$ and $r > 0$, but not every topology can be induced by a metric. A topology that can be induced by a metric is called *metrizable.* We will primarily be concerned with metrizable topologies.

Given a topology $\mathcal{T}$ on a set S, we identify other sets (subsets of S) of interest. A set is *closed* if its complement is open. Thus the special subsets

S and ϕ in every topology are both open and closed. We often use G to designate an open set and F to designate a closed set. For any subset A, its *closure* A^- is the intersection of all closed sets containing A, which is closed; its *interior* A° is the union of all open sets contained in A, which is open; and its *boundary* $\partial A \equiv A^- - A^\circ$ is the difference between the closure and the interior, which is closed.

The canonical example is the real line $\mathbb{R}$ with the usual distance $m(a,b) = |a-b|$. Let $(a,b] \equiv \{t \in \mathbb{R} : a < t \leq b\}$ and let other intervals be defined similarly. The intervals (a,b), $(-\infty,b)$ and (a,∞) are all open sets (called open intervals), while the intervals $[a,b]$, $(-\infty,b]$ and $[a,\infty)$ are all closed sets (called closed intervals). The intervals (a,b), $(a,b]$ and $[a,b]$ all have boundary the two-point set $\{a,b\}$. The open intervals (a,b) for $-\infty < a < b < \infty$ are the open balls inducing the topology.

A second example is the k-dimensional product space $\mathbb{R}^k$. For any p, $0 < p < \infty$,

$$\|a\|_p \equiv (\sum_{i=1}^{k} |a^i|^p)^{1/p}$$

for $a \equiv (a^1, \ldots, a^k) \in \mathbb{R}^k$ is the L_p norm. The associated metric

$$m_p(a,b) \equiv \|a-b\|_p$$

induces the *Euclidean topology* on $\mathbb{R}^k$. As $p \to \infty$, the L_p norm approaches the L_∞ norm

$$\|a\|_\infty \equiv \max_{1 \leq i \leq k} |a^i| \ ,$$

which also induces the Euclidean topology on $\mathbb{R}^k$.

Finite unions and finite intersections of open (closed) sets are again open (closed). We obtain new kinds of sets when we consider infinite unions or intersections. Because of our interest in probability measures on topological spaces, we are especially interested in countably infinite unions and intersections. A set is a G_δ if it is a countable intersection of open sets, an F_σ if it is a countable union of closed sets, a $G_{\delta\sigma}$ if it is a countable union of G_δ sets, and so forth.

A specific topology on a set is determined by specifying which subsets are open. That can be done by identifying a subbasis or a basis. A *subbasis* for the topology is any family of sets such that the given topology is the smallest topology containing that family. The topology generated by a subbasis contains the whole set S, the empty set ϕ, all finite intersections from the subbasis and all unions from these finite intersections. A *basis* is a family of open sets such that each open set is a union of basis sets. Thus, the family of all finite intersections from a subbasis forms a basis. For example, the collection of all open intervals with rational endpoints is a basis for the real line with the usual topology.

A topology on a set is also determined by specifying which functions from the given set to other topological spaces are continuous. A function f from one metric space (S, m) to another metric space (S', m') is *continuous* if $m'(f(x_n), f(x)) \to 0$ as $n \to \infty$ whenever $m(x_n, x) \to 0$ for a sequence $\{x_n : n \geq 1\}$ in S. A function f from one topological space S to another topological space S' is *continuous* if the inverse image of the open set G, $f^{-1}(G) \equiv \{s \in S : f(s) \in G\}$, is an open set in S for each open set G in S'.

One way to define a topology in terms of functions is to specify a class of functions from the given set to another topological space, and then stipulate that the topology is the smallest topology such that all functions in the designated class are continuous. (The inverse images of open sets for the functions form a subbasis.)

A one-to-one function $f : S \to S'$ mapping one topological space S onto another topological space S' such that both f and its inverse, mapping S' onto S (which we also denote by f^{-1}) are continuous is called a *homeomorphism*. (In the paragraph above, f^{-1} maps subsets of S' into subsets of S; here f^{-1} maps elements of S' into elements of S.) Two homeomorphic spaces are *topologically equivalent*. If we are only concerned about topological concepts, then two homeomorphic spaces can be regarded as two representations of the same space.

An important property held by some topological spaces is compactness. An *open cover* is a collection of open sets whose union is the entire space. A topological space is *compact* if each open cover has a finite subcover. In a metric space, a subset A is compact if every sequence in A has a convergent subsequence with limit in A.

Every subset B of a topological space S becomes a topological space in its own right with the *relative topology*, which contains all intersections of open subsets with B, i.e., all sets of the form $B \cap G$ where G is open in S. A subset of a compact topological space is itself compact if and only if the subset is closed. We often use K to denote a compact subset.

For example, in the real line $\mathbb{R}$ with the usual metric $m(a, b) = |a - b|$, the closed bounded interval $[a, b]$ is compact, but the intervals (a, b), $(a, b]$, $(-\infty, b)$ and $(-\infty, b]$ are not compact.

Every (Cartesian) product of topological spaces $\prod_{i \in I} S_i$ becomes a topological space with the *product topology*, which is defined by letting the subbasis contain all sets of elements $\{x_i : i \in I\}$ such that $x_{i_0} \in G_{i_0}$, where

G_{i_0} is an open set in S_{i_0} for any single index i_0. By Tychonoff's theorem, arbitrary products of compact topological spaces are compact.

Under regularity conditions, implied by the topology being metrizable, the topology is determined by specifying which sequences of elements from the set converge to limits in the set. A sequence $\{x_n : n \geq 1\}$ in a topological space S *converges* to a limit x in S if, for each open subset G containing x, there is an integer n_0 such that $x_n \in G$ for all $n \geq n_0$. In a metric space (S, m), the sequence converges if, for all ϵ, there exists an integer n_0 such that $x_n \in B_m(x, \epsilon)$ for all $n \geq n_0$.

When sequences are not adequate, we can use nets. A sequence in S can be regarded as a map from the positive integers into S; a net in S is a map from a directed set into S. A *directed set*, say Δ, is a set with an order relation $\prec$ defined on it, so that for any $a, b \in \Delta$ there exists $c \in \Delta$ such that $a \prec c$ and $b \prec c$. Of course the positive integers is a directed set; another directed set that is not totally ordered is the set of all subsets of a given set ordered by set inclusion. A *net* $\{x_\delta : \delta \in \Delta\}$ in S in a topological space S *converges* to a limit x in S if, for each open set G containing x, there is $\delta_0 \in \Delta$ such that $x_\delta \in G$ for all δ with $\delta_0 \prec \delta$. However, as indicated above, in metric spaces it suffices to consider only sequences. We only mention nets when it has not yet been established that the topological space is metrizable.

Thus, in a metrizable topological space the topology can be specified in any of the following ways:

(i) specifying the open subsets, e.g., by specifying a subbasis or a basis,

(ii) specifying a class of functions that must be continuous,

(iii) defining a metric,

(iv) specifying which sequences converge.

11.2.2. Separability and Completeness

In addition to having the topological space be metrizable, we often want to impose two additional regularity properties: separability and completeness. A topological space is *separable* if it has a countable dense subset; a subset A is *dense* in a topological space S if $A^- = S$, i.e., if the closure of A is the whole space S. In metric spaces, separability is equivalent to *second countability*, i.e., the topology having a *countable basis*, which in turn is equivalent to every open cover having a countable subcover. In a separable metric space, the balls $B_m(x, r)$ of rational radius r centered at points x in a countable dense set form a countable basis. Separable metric spaces are quite general, but by the Urysohn embedding theorem, any separable metric space is homeomorphic to a subset (with the relative topology) of the space

$$[0,1]^\infty \equiv [0,1] \times [0,1] \times \ldots$$

(with the product topology), which is metrizable as a compact metric space. The separable metric space itself is in general not compact, however.

A sequence $\{x_n : n \geq 1\}$ in a metric space (S, m) is *fundamental* (or satisfies the *Cauchy property*) if, for all $\epsilon > 0$, there exists $n_0 \equiv n_0(\epsilon)$ such that

$$m(x_n, x_m) < \epsilon \quad \text{for all} \quad n \geq n_0 \quad \text{and} \quad m \geq n_0 \ .$$

A metric space (S, m) is *complete* if each fundamental sequence converges to a limit in S. Completeness is useful for characterizing compactness, because a closed subset of a complete metric space is compact if and only if it is *totally bounded*, i.e., any cover by open balls has a finite subcover. (That is, in verifying that any open cover has a finite subcover, we may restrict attention to covers containing open balls.)

We are primarily concerned about the topology induced by a metric rather than the metric itself. We will often work with metrics that are not complete, but it will usually be possible to construct a topologically equivalent metric that is complete. When a topological space is metrizable as a complete metric space, we call the topological space *topologically complete*. When we are interested in the topology rather than the metric, the important property is topological completeness, not completeness.

A topological space that is metrizable as a complete separable metric space is said to be *Polish*. Closely related to Polish spaces are Lusin spaces. One topology $\mathcal{T}_1$ is a *stronger topology* (or finer topology) than another $\mathcal{T}_2$ if it contains the other as a proper subset, i.e., if $\mathcal{T}_2 \subseteq \mathcal{T}_1$. A set with a metrizable topology (or, more generally, a Hausdorff topology) on which there is a stronger topology that is Polish is called a *Lusin space*. A subset of a complete metric space is itself a complete metric space (with the same metric) if and only if it is closed; a subset of a Polish space is Polish if and only if it is a G_δ; a subset of a Lusin space is Lusin if and only if it is Borel measurable (see the next section). Countable products of Polish (Lusin) spaces are again Polish (Lusin). A nice account of Polish and Lusin spaces, and probability measures on them, is contained in Schwartz (1973). These spaces provide a natural setting for stochastic-process limits; only rarely is greater generality needed.

11.3. The Space $\mathcal{P}$

In this section we supplement the discussion in Section 3.2, in which we described the set $\mathcal{P}(S)$ of probability measures on a separable metric space (S, m), endowed with the topology of weak convergence.

11.3.1. Probability Spaces

In order to define probability measures on S, We make S a *measurable space* by endowing S with a σ-field of measurable sets (subsets of S). We let $\mathcal{S}$ denote a σ-field on S. Like a topology, a *σ-field* (on a set) is a collection of subsets of the designated set satisfying certain axioms. In particular, a σ-field contains the whole set and is closed under complements and countable unions. As usual, the sets in the σ-field are the sets to which we can assign probability. A σ-field generated by a collection of sets is the smallest σ-field containing those sets.

When we define a probability measure on a measurable space, we obtain a probability space. A probability measure on $(S, \mathcal{S})$ is a real-valued function on $\mathcal{S}$ satisfying $0 \leq P(A) \leq 1$ for all $A \in \mathcal{S}$, $P(S) = 1$ and $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ whenever $\{A_n : n \geq 1\}$ is a sequence of mutually disjoint subsets in $\mathcal{S}$ (No two of the subsets have any points in common.).

We will want to consider functions mapping one measurable space $(S, \mathcal{S})$ into another $(S', \mathcal{S}')$. A function $h : (S, \mathcal{S}) \to (S', \mathcal{S}')$ is said to be *measurable* if $h^{-1}(A') \in \mathcal{S}$ for each $A' \in \mathcal{S}'$. A measurable map $h : (S, \mathcal{S}) \to (S, \mathcal{S}')$ induces an *image (probability) measure* Ph^{-1} on $(S', \mathcal{S}')$ associated with each probability measure P on $(S, \mathcal{S})$, defined by

$$Ph^{-1}(A') \equiv P(h^{-1}(A')) \equiv P(\{s \in S : h(s) \in A'\}).$$

forall $A' \in \mathcal{S}'$.

So far, we have not exploited the topology on the space S. For a topological space $(S, \mathcal{T})$, we always use the *Borel σ-field* $\mathcal{B}(S)$ generated by the open subsets of S. In a metric space (S, m), the topology is generated by the metric m, i.e., by the open sets determined by m. Assuming that the σ-fields are Borel σ-fields, all continuous functions are measurable. Indeed, the Borel σ-field can be characterized as the smallest σ-field such that all bounded continuous real-valued functions are measurable.

Since a topology is closed under arbitrary unions, while a σ-field is closed under countable unions, Borel σ-fields tend to be well behaved when the topological space is second countable, i.e., has a countable basis. Thus, as an important regularity condition, we require that the metric space (S, m)

be *separable*. Separability plays an important role in product spaces; see Section 11.4 below.

11.3.2. Characterizing Weak Convergence

We are primarily interested in criteria for the convergence of a sequence of probability measures. As defined in Section 3.2, 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

$$\lim_{n\to\infty} \int_S f \, dP_n = \int_S f \, dP \tag{3.1}$$

for all functions f in $C(S)$, the space of all continuous bounded real-valued functions on S. (In Section 1.4 of the Internet Supplement we give a "Banach-space" explanation for the adjective "weak" in "weak convergence.")

Note that we could require more. We could require that (3.1) hold for all bounded measurable real-valued functions or for all indicator functions. That would be equivalent to requiring that

$$P_n(A) \to P(A) \quad \text{for all} \quad A \in \mathcal{B}(S) \ ,$$

but we do not. Indeed, that mode of convergence is often too strong. To see why, let

$$P_n(\{x_n\}) = 1 \quad \text{for} \quad n \geq 1 \quad \text{and} \quad P(\{x\}) = 1 \ , \tag{3.2}$$

where

$$x_n \to x \quad \text{as} \quad n \to \infty \ . \tag{3.3}$$

If $x_n \neq x$ for infinitely many n, then $P_n(A) \not\to P(A)$ for $A = \{x\}$.

We can give a related equivalent characterization of weak convergence $P_n \Rightarrow P$. A measurable subset A for which $P(\partial A) = 0$, where ∂A is the boundary of A, is said to be a *P-continuity set*. Weak convergence $P_n \Rightarrow P$ is equivalent to "pointwise convergence" $P_n(A) \to P(A)$ for all P-continuity sets A in $\mathcal{B}(S)$. The following "Portmanteau theorem" gives several alternative characterizations of weak convergence.

Theorem 11.3.1. (alternative characterizations of weak convergence) *The following are equivalent characterizations of weak convergence $P_n \Rightarrow P$ on a metric space:*

(i) $\lim_{n\to\infty} \int_S f \, dP_n = \int_S f \, dP$ *for all* $f \in C(S)$*;*

(ii) $\lim_{n\to\infty} \int_S f dP_n = \int_S f dP$ *for all uniformly continuous* f *in* $C(S)$;

(iii) $\limsup_{n\to\infty} P_n(F) \le P(F)$ *for all closed* F;

(iv) $\liminf_{n\to\infty} P_n(G) \ge P(G)$ *for all open* G;

(v) $\lim_{n\to\infty} P_n(A) = P(A)$ *for all* P*-continuity sets* A;

(vi) $P_n f^{-1} \Rightarrow P f^{-1}$ *on* $\mathbb{R}$ *for all* $f \in C(S)$;

(vii) $P_n f^{-1} \Rightarrow P f^{-1}$ *on* $\mathbb{R}$ *for all uniformly continuous* f *in* $C(S)$.

The different equivalent criteria in Theorem 11.3.1 can be understood by looking further at the deterministic example in (3.2)–(3.3). A key property used in the proof of Theorem 11.3.1 is the ability to approximate probabilities of measurable sets by the probabilities of open and closed sets.

Theorem 11.3.2. (approximation by closed and open sets) *Let* P *be an arbitary probability measure on a metric space* (S, m) *with the Borel* σ*-field* $\mathcal{B}(S)$. *For all* $A \in \mathcal{B}(S)$ *and all* $\epsilon > 0$, *there exists a closed set* F *and an open set* G *with*

$$F \subseteq A \subseteq G$$

such that

$$P(G - F) < \epsilon \ .$$

It is also useful to be able to approximate probabilities by the probability of compact subsets. A probability measure on a topological space S is said to be *tight* if, for all $\epsilon > 0$, there exists a compact supset K such that

$$P(K) > 1 - \epsilon \ .$$

The following is an important property of Lusin spaces; again see Schwartz (1973).

Theorem 11.3.3. (approximation by compact sets) *In a Lusin space* S *all probability measures are tight. Let* P *be an arbitrary probability measure on a Lusin space* S *with its Borel* σ*-field* $\mathcal{B}(S)$. *For all* $A \in \mathcal{B}(S)$ *and all* $\epsilon > 0$, *there exists a compact set* K *with* $K \subseteq A$ *such that*

$$P(A - K) < \epsilon \ .$$

We put the notion of weak convergence just given in a standard topological framework by observing that it can be characterized by a metric. That can be done in several ways; one is with the Prohorov metric defined in (2.2) in Section 3.2. We remark that in the definition of the Prohorov metric it suffices to restrict attention to A being a closed subset of S. The space $\mathcal{P}(S)$ tends to inherit properties from the underlying space S. Given that (S, m) is a separable metric space, the space $(\mathcal{P}(S), \pi)$ is topologically complete or compact if and only if (S, m) is. Moreover, the subset of probability measures in $\mathcal{P}(S)$ assigning unit mass to individual points in S with the relative topology is homeomorphic to S itself. See Chapter II of Parthasarathy (1967).

On the real line $\mathbb{R}$, we often use metrics applied to cumulative distribution functions (cdf's). The *Lévy metric*, say λ, is defined by (2.2) in Section 3.2 but only considering sets of the form $A = (-\infty, x]$. Clearly, $\lambda \leq \pi$, but both λ and π induce the topology of weak convergence in $\mathcal{P}(\mathbb{R})$.

11.3.3. Random Elements

As indicated in Section 3.2, instead of directly referring to probability measures, we often use random elements. We now restate Theorem 11.3.1 in terms of random elements. We say that a subset A in $\mathcal{B}(S)$ is an X-*continuity set* if $P(X \in \partial A) = 0$.

Theorem 11.3.4. (alternative characterizations of convergence in distribution) *The following are equivalent characterizations of convergence in distribution $X_n \Rightarrow X$ for random elements of a metric space:*

(i) $\lim_{n\to\infty} Ef(X_n) = Ef(X)$ *for all* $f \in C(S)$*;*

(ii) $\lim_{n\to\infty} Ef(X_n) = Ef(X)$ *for all uniformly continuous* f *in* $C(S)$*;*

(iii) $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ *for all closed* F*;*

(iv) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ *for all open* G*;*

(v) $\lim_{n\to\infty} P(X_n \in A)$ *for all* X*-continuity sets* A*;*

(vi) $f(X_n) \Rightarrow f(X)$ *in* $\mathbb{R}$ *for all* $f \in C(S)$*;*

(vii) $f(X_n) \Rightarrow f(X)$ *in* $\mathbb{R}$ *for all uniformly continuous* f *in* $C(S)$*.*

The adjective "weak" in "weak convergence" distinguishes convergence in distribution $X_n \Rightarrow X$ from the stronger convergence $X_n \rightarrow X$ with probability one (w.p.1), which is called a *strong limit.* (The strong limit can hold only when X_n and X are defined on a common probability space.) We will give an alternative explanation for using the adjective "weak" below.

We now elaborate further on the meaning of weak convergence $P_n \Rightarrow P$. First, notice that the definition of the Prohorov metric allows the probability measure P_2 to assign a mass $\pi(P_1, P_2)$ arbitrarily far from where P_1 assigns its mass, allowing for a small chance of a big error. We can better understand the Prohorov metric π by considering a special representation, originally due to Strassen (1965); also see Billingsley (1999) and Pollard (1984).

The Strassen representation theorem relates the Prohorov distance between two probability measures to the distance in probability between two specially constructed random elements with those probability laws. For two random elements X_1 and X_2 of a separable metric space (S, m) defined on the same underlying probability space $(\Omega, \mathcal{F}, P)$, the *in-probability distance* between X_1 and X_2 is defined by

$$p(X_1, X_2) \equiv \inf\{\epsilon > 0 : P(m(X_1, X_2) > \epsilon) < \epsilon\} \ . \tag{3.4}$$

(We need separability for $m(X_1, X_2)$ to be a legitimate random variable; see the next section. The distance p is only a pseudometric because $p(X_1, X_2) = 0$ does not imply that $X_1 = X_2$.)

It is easy to see that

$$\pi(PX_1^{-1}, PX_2^{-1}) \leq p(X_1, X_2)$$

for any random elements mapping an underlying probability space $(\Omega, \mathcal{F}, P)$ into a separable metric space (S, m). The Strassen representation theorem allows us to go the other way for specially constructed random elements.

Theorem 11.3.5. (Strassen representation theorem) *For any $\epsilon > 0$ and any two probability measures P_1 and P_2 on a separable metric space (S, m), there exist special S-valued random elements X_1 and X_2 on some common underlying probability space such that*

$$PX_i^{-1} = P_i \quad \textit{for} \quad i = 1, 2$$

and

$$p(X_1, X_2) < \pi(P_1, P_2) + \epsilon \ , \tag{3.5}$$

where p is the in-probability distance in (3.4). *If the two probability measures* P_1 *and* P_2 *are tight, which always holds if* (S, m) *is also a Lusin space, then the random elements* X_1 *and* X_2 *can be constructed so that*

$$p(X_1, X_2) = \pi(P_1, P_2) \ . \tag{3.6}$$

The Strassen representation theorem says that the Prohorov distance between probability measures can be realized (possibly only via an infimum) as the distance in probability between two specially constructed random elements on a common probability space that have the given probability measures as their probability laws. It suffices to let the underlying probability space be the product space $(S, m) \times (S, m)$ and the random elements be the coordinate projections. The problem then is to construct the probability measure on $(S, m) \times (S, m)$ with the specified marginal probability laws satisfying (3.5) or (3.6).

Example 11.3.1. *A simple example.* To fix ideas it is useful to consider an example. In Table 11.1 we specify three different random variables defined on a simple probability space $(\Omega, \mathcal{F}, P)$. The sample space Ω contains only four elements; the σ-field $\mathcal{F}$ contains all subsets; and the probability measure P assigns equal probabilities to each set containing a single point.

Ω	$P(\{\omega\})$	$X_1(\omega)$	$X_2(\omega)$	$X_3(\omega)$
ω_1	1/4	1/4	3/4	1/4
ω_2	1/4	2/4	4/4	2/4
ω_3	1/4	3/4	1/4	3/4
ω_4	1/4	4/4	2/4	100

Table 11.1: Three possible random variables

In Table 11.2 we display the distances $\pi(PX_1^{-1}, PX_2^{-1})$ and $p(X_1, X_2)$,

distance	$i = 1,\ j = 2$	$i = 1,\ j = 3$	$i = 2,\ j = 3$
$\pi(PX_i^{-1}, PX_j^{-1})$	0	1/4	1/4
$p(X_i, X_j)$	1/2	1/4	1/2
$\|X_i - X_j\|$	1/2	99	99 1/4

Table 11.2: Distances between the random variables

along with the uniform distance between the random variables $\| X_1 - X_2 \|$, where

$$\| X \| \equiv \sup_{\omega \in \Omega} |X(\omega)| \ .$$

Note that the random variables X_1 and X_2 are different, but they have the same distribution. The distances always increase as we go down in Table 11.2, but generalizations going sideways are hard to make. ∎

The Skorohod representation theorem, Theorem 3.2.2, also helps to understand the topology of weak convergence.

11.4. Product Spaces

We are often interested in joint convergence of random elements: We want to go beyond $X_n \Rightarrow X$ and $Y_n \Rightarrow Y$ to obtain $(X_n, Y_n) \Rightarrow (X, Y)$. We consider such joint limits because we want to understand the joint distribution of X_n and Y_n. We also often require the joint convergence in order to apply the continuous-mapping approach. Just li

First, to have a vector random element (X, Y) well defined, we need X and Y to be defined on a common underlying probability space. Then, given random elements X of a separable metric space (S', m') and Y of a separable metric space (S'', m''), we can regard (X, Y) as a random element of the *product space* associated with two metric spaces (S', m') and (S'', m''), i.e.,

$$S \equiv S' \times S'' \equiv \{(x, y) : x \in S', y \in S''\} \ .$$

The *open rectangles* $G' \times G''$ with G' open in S' and G'' open in S'' are a basis for the *product topology* on $S' \times S''$. The product topology is characterized by having convergence $(x'_n, x''_n) \to (x', x'')$ if and only if $x'_n \to x'$ and $x''_n \to x''$. The product topology can be induced by several different metrics, one being the *maximum metric*

$$m((x_1, y_1), (x_2, y_2)) \equiv \max\{m'(x_1, x_2), m''(y_1, y_2)\} \ .$$

Similarly, the *measurable rectangles* $A' \times A''$ with A' measurable in S' and A'' measurable in S'' *generate* the *product σ-field* on the product space $S' \times S''$; i.e., the product σ-field is the smallest σ-field on $S' \times S''$ containing the measurable rectangles.

The following basic theorems explain why we need our metric spaces to be separable, i.e., to have countable dense subsets.

Theorem 11.4.1. (separability of product spaces) *The product space $S' \times S''$ with the product topology is separable if and only if the component spaces S' and S'' are separable.*

Theorem 11.4.2. (the Borel σ-field in product spaces) *The Borel σ-field $\mathcal{B}(S)$ associated with the product space $S = S' \times S''$ with product topology is the product σ-field $\mathcal{B}(S') \times \mathcal{B}(S'')$ if and only if the metric spaces (S', m') and (S'', m'') are separable.*

For a probability measure P on $S \equiv S' \times S''$, *marginal probability measures* P' and P'' are defined on $(S', \mathcal{S}')$ and $(S'', \mathcal{S}'')$ by setting

$$P'(A') \equiv P(A' \times S'') \quad \text{and} \quad P''(A'') \equiv P(S' \times A'')$$

for every $A' \in \mathcal{S}'$ and $A'' \in \mathcal{S}''$. Thus, if (X, Y) is a random element of $S = S' \times S''$ with probability law P, then P' and P'' are the probability laws of X and Y, respectively.

Theorem 11.4.3. (criteria for joint convergence) *Suppose that the product space $S \equiv S' \times S''$ with the product topology is separable. Then the following are each necessary and sufficient conditions for convergence in distribution $(X_n, Y_n) \Rightarrow (X, Y)$ in S:*

$$(i) \quad P(X_n \in A', Y_n \in A'') \to P(X \in A', Y \in A'')$$

for every X-continuity set A' and every Y-continuity set A''.

$$(ii) \quad \overline{\lim_{n \to \infty}} \; P(X_n \in F', Y_n \in F'') \le P(X \in F', Y \in F'')$$

for every closed set F' in S' and every closed set F'' in S''.

$$(iii) \quad \underline{\lim_{n \to \infty}} \; P(X_n \in G', Y_n \in G'') \ge P(X \in G', Y \in G'')$$

for every open set G' in S' and every open set G'' in S''.

In general, we must verify one of the limits in Theorem 11.4.3 (i) – (iii) in order to establish convergence in distribution for vector random elements, but there are two special cases in which we easily get convergence in distribution for vector random elements. One case involves independence and the other involves a deterministic limit.

For given probability measures P' on $(S', \mathcal{S}')$ and P'' on $(S'', \mathcal{S}'')$, the *product probability measure* $P' \times P''$ on the product space $S' \times S''$ with the product σ-field $\mathcal{S}' \times \mathcal{S}''$ is defined by

$$(P' \times P'')(A' \times A'') = P'(A')P''(A'')$$

for every $A' \in \mathcal{S}'$ and $A'' \in \mathcal{S}''$. By Theorem 11.4.2, the product measure is defined on the Borel field of $S' \times S''$ when (S', m') and (S'', m'') are separable. When X and Y are independent random elements of S' and S'', the probability law of (X, Y) is the product probability law $P_1X^{-1} \times P_2Y^{-1}$, where P_i are the probability measures in the underlying probability spaces.

Theorem 11.4.4. (joint convergence for independent random elements) *Let X_n and Y_n be independent random elements of separable metric spaces (S', m') and (S'', m'') for each $n \geq 1$. Then there is joint convergence in distribution*

$$(X_n, Y_n) \Rightarrow (X, Y) \quad \text{in} \quad S' \times S''$$

if and only if $X_n \Rightarrow X$ in S' and $Y_n \Rightarrow Y$ in S''.

Theorem 11.4.5. (joint convergence when one limit is deterministic) *Suppose that $X_n \Rightarrow X$ in a separable metric space (S', m') and $Y_n \Rightarrow y$ in a separable metric space (S'', m''), where y is deterministic. Then*

$$(X_n, Y_n) \Rightarrow (X, y) \quad \text{in} \quad S' \times S'' .$$

Proof. By Theorem 11.4.3, it suffices to show that

$$P(X_n \in A, Y_n \in B) \to P(X \in A, y \in B) \tag{4.1}$$

for each X-continuity set A and y-continuity set B (i.e., where $y \notin \partial B$). First suppose that $y \in B$, which implies that $P(Y_n \notin B) \to 0$. Then (4.1) holds because

$$P(X_n \in A) - P(Y_n \notin B) \leq P(X_n \in A, Y_n \in B) \leq P(X_n \in A) .$$

Now suppose that $y \notin B$. Then (4.1) again holds because

$$P(X_n \in A, Y_n \in B) \leq P(Y_n \in B) \to 0. \quad \blacksquare$$

Given two random elements X and Y of a common metric space (S, m), we can speak of the random distance $m(X, Y)$. Such a random distance is a legitimate real-valued random variable when (S, m) is a separable metric space, but not otherwise; see p. 225 of Billingsley (1968).

Theorem 11.4.6. (measurability of the distance between random elements) *If (S, m) is a separable metric space and X and Y are random elements of S defined on a common domain, then $m(X, Y)$ is a legitimate measurable real-valued random variable.*

Note that convergence in distribution $Y_n \Rightarrow y$ to a deterministic limit in Theorem 11.4.5 (where (S'', m'') is a separable metric space) is equivalent to *convergence in probability*; i.e., $Y_n \Rightarrow y$ above if and only if

$$P(m''(Y_n, y) > \epsilon) \to 0 \quad \text{as} \quad n \to \infty$$

for all $\epsilon > 0$.

We now give a useful way to establish new weak convergence limits from given ones. We already used this result in our treatment of the Kolmogorov-Smirnov statistic in Section **??**.

Theorem 11.4.7. (convergence-together theorem) *Suppose that X_n and Y_n are random elements of a separable metric space (S, m) defined on a common domain. If $X_n \Rightarrow X$ in S and $m(X_n, Y_n) \Rightarrow 0$ in $\mathbb{R}$, then*

$$(X_n, Y_n) \Rightarrow (X, X) \quad \textit{in} \quad (S, m) \times (S, m) \ .$$

Proof. For any closed subset F of S, let $F^{\bar{\epsilon}}$ be its closed ϵ-neighborhood, defined by

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

For any two closed subsets F_1 and F_2 of S,

$$P(X_n \in F_1, Y_n \in F_2) \le P(X_n \in (F_1 \cap F_2)^{\bar{\epsilon}}) + P(m(X_n, Y_n) \ge \epsilon) \ ,$$

so that

$$\overline{\lim_{n \to \infty}} \ P(X_n \in F_1, Y_n \in F_2) \le \overline{\lim_{n \to \infty}} \ P(X_n \in (F_1 \cap F_2)^{\bar{\epsilon}}) \le P(X \in (F_1 \cap F_2)^{\bar{\epsilon}})$$

by Theorem 11.3.4 (iii), since $X_n \Rightarrow X$. Letting $\epsilon \downarrow 0$, we have $P(X \in (F_1 \cap F_2)^{\bar{\epsilon}}) \downarrow P(X \in (F_1 \cap F_2))$. Hence,

$$\overline{\lim_{n \to \infty}} \ P(X_n \in F_1, Y_n \in F_2) \le P(X \in F_1, X \in F_2) \ ,$$

which implies the conclusion by Theorem 11.4.3. ■

We usually use Theorem 11.4.7 in proofs to obtain a desired limit $Y_n \Rightarrow X$ in S (one coordinate only) by treating a closely related sequence $\{X_n\}$ that is easier to analyze. There is a converse to Theorem 11.4.7 that adds insight into the significance of joint convergence to a common limit. We not only get the two marginal limits $X_n \Rightarrow X$ and $Y_n \Rightarrow X$, but we also get asymptotic equivalence of X_n and Y_n.

Theorem 11.4.8. (asymptotic equivalence from joint convergence) *Suppose that X_n and Y_n are random elements of a separable metric space (S, m) with a common domain. If $(X_n, Y_n) \Rightarrow (X, X)$ in $S \times S$, then*

$$m(X_n, Y_n) \Rightarrow 0 \quad in \quad \mathbb{R} .$$

Proof. Apply the Skorohod representation theorem to replace the convergence in distribution $(X_n, Y_n) \Rightarrow (X, X)$ in $S \times S$ by convergence w.p.1 for the special versions $\tilde{X}_n$ and $\tilde{Y}_n$. Then apply the triangle inequality in (S, m) to deduce that $m(\tilde{X}_n, \tilde{Y}_n) \to 0$ w.p.1. Finally, note that $m(X_n, Y_n)$ has the same distribution as $m(\tilde{X}_n, \tilde{Y}_n)$, so that we obtain the desired conclusion. ■

11.5. The Space *D*

In this section we supplement the discussion of the functions space D in Section 3.3, primarily by introducing the Skorohod (1956) J_2 and M_2 topologies. For the discussion here, we assume that the functions are real-valued and that the function domain is $[0, 1]$.

We start by making some observations about the J_1 metric defined in (3.2) of Section 3.3. The metric d_{J_1} is incomplete, but the topology is topologically complete; there exists a topologically equivalent metric that is complete; see Billingsley (1968). The space (D, J_1) is separable; a countable dense set is made up of the rational-valued piecewise-constant functions with only finitely many discontinuities, all at rational time points in the domain $[0, 1]$. Thus the space (D, J_1) is Polish. The J_1 topology can also be defined on D spaces with more general ranges. We can let the range be $\mathbb{R}^k$ or any Polish space.

11.5.1. J_2 and M_2 Metrics

A metric inducing the J_2 topology on $D([0, 1], \mathbb{R})$ is defined by replacing the set of functions Λ by the larger set Λ' of all one-to-one maps of $[0, 1]$

onto $[0, 1]$, without requiring any continuity, i.e.,

$$d_{J_2}(x_1, x_2) \equiv \inf_{\lambda \in \Lambda'} \{\|x_1 \circ \lambda - x_2\| \vee \|\lambda - e\|\} \ . \tag{5.1}$$

Since $\Lambda \subseteq \Lambda'$, we obviously have

$$d_{J_2}(x_1, x_2) \le d_{J_1}(x_1, x_2) \le \|x_1 - x_2\| \ .$$

We will have little to say about the J_2 topology; see Bass and Pyke (1987) for an application.

As noted in Section 3.3, we need a different topology on D if we want the jump in a limit function to be unmatched in the converging functions. In order to establish limits with unmatched jumps in the limit function (or process), we use the Skorohod (1956) M topologies. We define the M topologies using the completed graphs of the functions, defined in (3.3) in Section 3.3.

The completed graph It is thus natural to consider established metrics defined on the set of all compact subsets of $\mathbb{R}^k$. Perhaps the best known such metric is the Hausdorff metric. Given compact subsets K_1 and K_2 of $\mathbb{R}^k$, the *Hausdorff metric* m_H is defined by

$$m_H(K_1, K_2) \equiv \sup_{x_1 \in K_1} m(x_1, K_2) \vee \sup_{x_2 \in K_2} m(x_2, K_1) \ , \tag{5.2}$$

where $m(x, A)$ is the *distance between the point x and the set A*, defined by

$$m(x, A) \equiv m(A, x) \equiv \inf_{y \in A} m(x, y) \tag{5.3}$$

and m is the metric used on $\mathbb{R}^k$; e.g., see Section 1.4 of Matheron (1975).

To illustrate, the Hausdorff distance between two compact subsets in $\mathbb{R}^k$ is depicted in Figure 11.1. The compact sets are the set of points inside the oval and the set of points inside the rectangle, including the boundaries. Note that the two sets overlap. The dashed lines identify the point in the oval furthest away from the rectangle and the point in the rectangle furthest away from the oval. The Hausdorff distance is the greater of these two distances. Thus, even if one compact subset is a proper subset of another, they are a positive distance apart.

Thus, for any $x_1, x_2 \in D$, the M_2 metric on D is defined by

$$d_{M_2}(x_1, x_2) \equiv m_H(\Gamma_{x_1}, \Gamma_{x_2}) \ , \tag{5.4}$$

where m_H is the Hausdorff metric in (5.2) and Γ_x is the completed graph of x, defined in (3.3) of Section 3.3. The topology on the space of completed

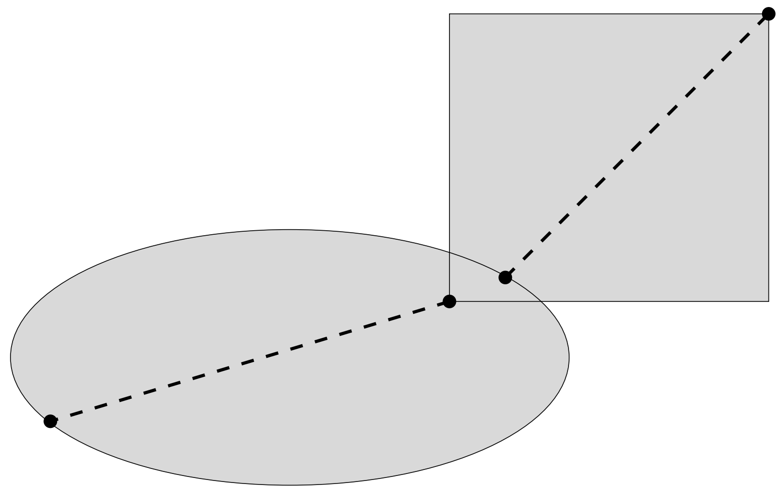

Figure 11.1: The Hausdorff distance between two compact subsets of the plane. The Hausdorff distance is the length of the longer dashed line.

graphs induced by the Hausdorff metric is the Skorohod M_2 topology (although that is not the way it was originally defined). The M_1 topology is stronger than the M_2 topology. It pays closer attention to order.

From the definitions above, it is not obvious how the M_1 and M_2 topologies are related. For understanding the relation beween the two topologies, it is significant that the M_2 topology can also be expressed via parametric representations. Indeed, an alternative M_2 metric (inducing the M_2 topology) can be defined by (3.4) in Section 3.3, after changing the definition of a parametric representation: Instead of requiring that (u, r) be nondecreasing, using the order on the completed graphs, we only require that the time component function r be nondecreasing. With that definition, it is evident that the M_1 topology is stronger than the M_2 topology, i.e., M_1 convergence implies M_2 convergence.

Unlike the J topologies, it is not possible to extend the M topologies by allowing the range to be an arbitrary Polish space, because the completed graphs require linear structure. However, the range can be a separable Banach space with the M topologies. That generalization is used to obtain heavy-traffic stochastic-process limits for the workload process in an infinite-server queue in Section 10.3.

11.5.2. The Four Skorohod Topologies

A unified approach to the four Skorohod topologies via graphs was provided in the thesis by Pomarede (1976). In that approach, the M_2 and J_2 topologies are generated by the Hausdorff metric applied to the completed and uncompleted graphs, respectively. Similarly, the M_1 and J_1 topologies are defined in terms of parametric representations of the completed and uncompleted graphs. That approach to the J_1 topology draws upon Kolmogorov (1956).

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 one of the other Skorohod (1956) non-uniform topologies instead, because the J_1 topology is stronger (or finer) than the other topologies. The four non-uniform Skorohod topologies are ordered by

$$J_1 > J_2 > M_2 \quad \text{and} \quad J_1 > M_1 > M_2 \ , \tag{5.5}$$

where $>$ means stronger than, with M_1 and J_2 not being comparable. Examples of functions x_n converging to the indicator function $x \equiv I_{[2^{-1},1]}$ in $D([0,1],\mathbb{R})$ in the different topologies are given in Figure 11.2. We contend that the M_1 topology is often the most appropriate one; we discuss this point further in Chapter 6.

We have indicated that all four non-uniform Skorohod topologies reduce to uniform convergence over $[0,1]$ when the limit function is continuous. More generally, convergence in all four of these topologies implies local uniform convergence at any continuity function of a limit function. Thus all four Skorohod topologies are stronger than the L_p topologies on D induced by the norms

$$\| x \|_p \equiv (\int_0^1 |x(t)|^p dt)^{1/p} \ . \tag{5.6}$$

More generally, we have the following continuity result.

Theorem 11.5.1. (continuity of integrals) *Suppose that $g : \mathbb{R}^k \to \mathbb{R}$ is a continuous function and let $f : D^k \to (C, U)$ be defined by*

$$f(x)(t) \equiv \int_0^t g(x)(s)ds, \quad t \geq 0 \ .$$

If D^k is endowed with any of the Skorohod non-uniform topologies, then f is continuous.

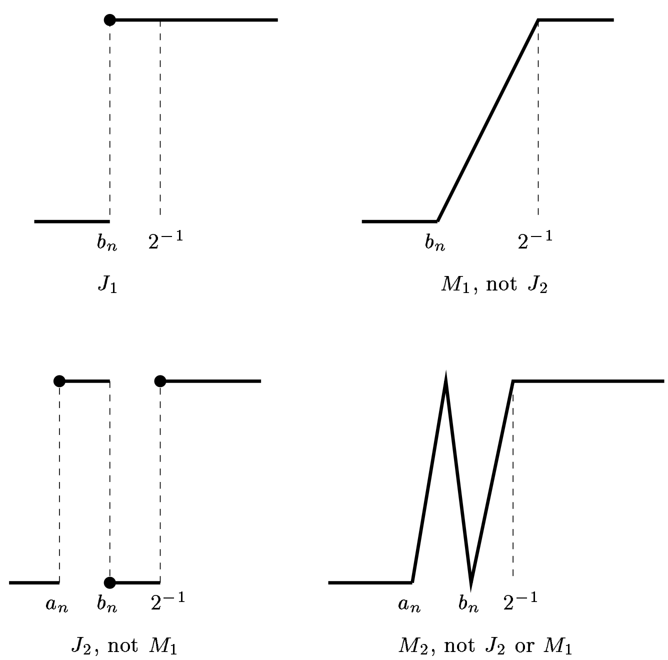

Figure 11.2: Four candidate sequences of functions $\{x_n : n \geq 1\}$ that might converge to $x \equiv I_{[1/2,1]}$ in $D([0,1],\mathbb{R})$, where $a_n = 2^{-1} - 2n^{-1}$ and $b_n = 2^{-1} - n^{-1}$.

Proof. First note that

$$\int_0^t |x(s)|ds \leq \int_0^T |x(s)|ds$$

for all real-valued x and t with $0 \leq t \leq T$. Then use the bounded convergence theorem: Convergence $x_n \to x$ imples that $\sup_n \|x_n\| < \infty$ and that $x_n(t) \to x(t)$ at almost all t. ■

Since the J_1 topology on D is Polish, the J_2, M_1, M_2 and L_p topologies on D are automatically Lusin. In Chapter 12 we will show that the M_1 topology is Polish.

As with the J_1 and M_1 topologies on D, the domain of the functions can be changed to $[0,\infty)$. for the J_2 and M_2 topologies. In all cases $x_n \to x$ is understood to mean that there is convergence of the restrictions of x_n to the restriction of x in $D([0,t],\mathbb{R}^k)$ for all t that are continuity points of the limit function x. However, it sometimes is desirable to obtain stronger control of the convergence at the end of the domain. By exploiting the SLLN, we can often assume that the sample paths x satisfy $\|x\|_w < \infty$ w.p.1, where $\|x\|_w$ is the *weighted-supremum norm*, i.e.,

$$\|x\|_w \equiv \sup_{0 \le t < \infty} \{|x(t)|/(1+t)\} ,$$

so that we can use the metric $\|x_1 - x_2\|_w$ on the subset of functions in $D([0,\infty),\mathbb{R})$ with finite weighted-supremum norm. Weak convergence with the weighted-supremum norm was established by Müller (1968) and applied by Whitt (1972). Related weighted distances have been used extensively in the study of empirical processes, as can be seen from Shorack and Wellner (1986) and Csörgő and Horvath (1993).

In Section 3.3 we noted that addition is not continuous everywhere on $D \times D$, but that it is continuous at all pairs (x,y) in $D \times D$ that have no common discontinuity points (and, for the M_1 topology, at all pairs (x,y) that have no common discontinuity points with jumps of opposite sign). In many applications, we are able to show that the two-dimensional limiting stochastic process has sample paths in one of those subsets w.p.1., so that we can apply the continuous-mapping approach with addition.

11.5.3. Measurability Issues

Since addition is not continuous everywhere, we most not only show that it is continuous almost everywhere, but we must show that it is measurable. It is thus important to know more about the Borel σ-fields associated with the non-uniform Skorohod topologies. Fortunately, in each case, the Borel σ-field coincides with the usual σ-field, namely, the *Kolmorogov σ-field*, which is generated by the projection maps $\pi_{t_1,\ldots,t_k} : D \to \mathbb{R}^k$, defined by

$$\pi_{t_1,\ldots,t_k}(x) = [x(t_1),\ldots,x(t_k)] ,$$

so that measurability in $(D,\mathcal{B}(D))$ with any of the non-uniform topologies is consistent with the standard notion. The Kolmogorov σ-field generated by the coordinate projections (or "cylinder sets") is usually associated with the product space $\mathbb{R}^\infty$ and the Kolmogorov (1950) extension theorem; see Neveu (1965).

Theorem 11.5.2. (the Borel σ-fields on D) *The Borel σ-fields on D with any of the non-uniform Skorohod topologies coincides with the Kolmogorov σ-field generated by the coordinate projections.*

Theorem 11.5.2 can be proved by direct verification in each case, as done for the J_1 topology in Billingsley (1968). For the non-J_1 topologies, we can also exploit the established J_1 result and properties of Lusin spaces; see p. 101 of Schwartz (1973).

Theorem 11.5.3. (Borel σ-fields for comparable Lusin spaces) *Any two comparable Lusin topologies on a set have identical Borel σ-fields.*

It often happens that we have a limit for a sequence of stochastic processes with sample paths in D, where the limit process has continuous sample paths. Then there is considerable flexibility on the choice of the topology. In that case, the four non-uniform topologies on D reduce to uniform convergence over all bounded intervals, and all four topologies have the same Borel σ-field. Clearly, then it does not matter which of these topologies is used.

We might naturally try to simplify matters even further in such a situation. We might choose to work directly with the space (D, U), where U denotes the topology of uniform convergence (over closed bounded subintervals) on D, induced by the uniform metric in Section 3.3 when the function domain is $[0, 1]$. There is a complication, however. Even though convergence $x_n \to x$ in D with the various non-uniform Skorohod topologies is equivalent to uniform convergence over all bounded intervals when the limit function x is continuous, in general we cannot simply work with the space (D, U), because there are measurability problems. The Borel σ-fields on D (generated by the open subsets) of all the non-uniform Skorohod topologies on D coincide with the usual Kolmogorov σ-field on D generated by the coordinate projections, but that is not true for (D, U). The Borel σ-field on (D, U) is much larger than the Kolmogorov σ-field. As a consequence, familiar stochastic processes such as the Poisson process cannot be regarded as random elements of (D, U) with the Borel σ-field; see Section 18 of Billingsley (1968).

The measurability problems arise because the space (D, U) is nonseparable. However, everything works well if we use D with a non-uniform Skorohod topology. The Borel σ-field then is the familiar one and, if a limit function is continuous, convergence is equivalent to uniform convergence.

If we want to consider only convergence of stochastic processes where the limiting stochastic process has continuous sample paths, then there is an alternative approach. We can then use the space (D, U) with the uniform topology, but use a smaller σ-field than the Borel σ-field, in particular, the σ-field generated by the open balls, called the em ball σ-field $B_m(x, r)$. Such a theory was developed by Dudley (1966, 1967) and is explained and used by Pollard (1984).

In contrast, in this book we are interested in the convergence of stochastic processes where the limiting stochastic process has discontinuous sample paths. Thus, for both measurability and convergence, we want to use a non-uniform topology on D. The particular non-uniform topology on D becomes important when the limit functions become discontinuous. The J_1 topology is useful because it allows some flexibility in the location of jumps, but it requires that the converging functions have jumps corresponding to each jump in the limit function. Thus, here we focus on the M_1 topology.

11.6. The Compactness Approach

In this book we focus on the continuous-mapping approach to establish stochastic-process limits. To put the continuous-mapping approach in perspective, we now describe the standard approach to establish stochastic-process limits based on compactness. Our accounts of both the compactness approach and the continuous-mapping approach are abridged versions of the excellent accounts in Billingsley (1968).

The compactness-approach applies to probability measures on a general metric space (S, m). In Section 3.2 we indicated that the space $(\mathcal{P}(S), \pi)$ of probability measures on (S, m) with the Prohorov metric π is a metric space. As in any metric space, we have convergence $\pi(P_n, P) \to 0$ as $n \to \infty$ for a sequence $\{P_n : n \geq 1\}$ in $(\mathcal{P}(S), \pi)$ if and only if every subsequence $\{P_{n'} : n' \geq 1\}$ contains a further subsequence $\{P_{n''} : n'' \geq 1\}$ with $P_{n''} \Rightarrow P$. We exploit a version of sequential compactness to provide conditions under which that characterization of convergence is satisfied. The compactness approach has been used to establish most of the initial stochastic-process limits we will use in the continuous-mapping approach.

As in any metric space, a subset A of $(\mathcal{P}(S), \pi)$ has compact closure A^- if and only if the set A is *relatively compact*, i.e., if every sequence $\{P_n : n \geq 1\}$ in A has a subsequence $\{P_{n'} : n' \geq 1\}$ with $P_{n'} \Rightarrow P'$, where the limit P' is necessarily in the closure A^-. Thus, given a sequence $\{P_n : n \geq 1\}$, we can establish convergence $P_n \Rightarrow P$ in $\mathcal{P}(S)$ by showing, first, that the sequence $\{P_n : n \geq 1\}$ is relatively compact and, second, by showing that the limit of any convergent subsequence must be P. The second step can be established by establishing a weaker form of convergence, which is not strong enough to imply weak convergence (i.e., $\pi(P_n, P) \to 0$), but which is strong enough to uniquely determine the limit P. The two steps together imply that $P_n \Rightarrow P$.

A key step in the compactness-approach to limits for sequences of probability measures on a metric space is to relate compact subsets in the space $(\mathcal{P}(S), \pi)$ to compact subsets of the underlying space (S, m). That can be done by applying *Prohorov's theorem*, from Prohorov (1956). The key concept is tightness, which we now extend from a single probability measure to a set of probability measures. A subset A of probability measures in $\mathcal{P}(S)$ is said to be *tight* if, for all ϵ, there exists a compact subset K of (S, m) such that

$$P(K) > 1 - \epsilon \quad \text{for all} \quad P \in A \ .$$

The compact set K depends upon ϵ, but it must do the job for all P in A. Theorem 11.3.3 implies that every single probability measure on a Lusin space is tight.

Theorem 11.6.1. (Prohorov's theorem) *Let (S, m) be a metric space. If a subset A in $\mathcal{P}(S)$ is tight, then it is relatively compact. On the other hand, if the subset A is relatively compact and the topological space S is Polish, then A is tight.*

Thus, in Polish spaces tightness is necessary and sufficient for relative compactness. That implies that nothing is lost by focusing on tightness when we want to establish relative compactness. (That is the primary basis for interest in knowing whether a topological space is Polish when we are concerned about weak convergence of probability measures.)

From Prohorov's theorem, we obtain a useful way to establish convergence $P_n \Rightarrow P$.

Corollary 11.6.1. (tightness criterion for weak convergence) *Let $\{P_n : n \geq 1\}$ be a sequence of probability measures on a metric space (S, m). If the sequence $\{P_n\}$ is tight and the limit of any convergent subsequence from $\{P_n\}$ must be P, then $P_n \Rightarrow P$.*

These two conditions apply very naturally to establish criteria for convergence $X_n \Rightarrow X$ for stochastic processes $\{X_n(t) : 0 \leq t \leq 1\}$ in the function space $C \equiv C([0,1], \mathbb{R})$ of continuous real-valued functions on the interval $[0,1]$ (or any other closed bounded interval) with the uniform metric. First, it is natural to require convergence of all the finite-dimensional distributions, i.e., to show that

$$(X_n(t_1), \ldots, X_n(t_k)) \Rightarrow (X(t_1), \ldots, X(t_k)) \quad \text{in} \quad \mathbb{R}^k \tag{6.1}$$

for all positive integers k and all k time points $t_1, \ldots, t_k$ with $0 \leq t_1 < \cdots < t_k \leq 1$. By the Kolmogorov extension theorem, it is known that the finite-dimensional distributions uniquely determine a probability distribution on the larger product space $\mathbb{R}^{[0,1]}$, endowed with the Kolmogorov σ-field generated by the coordinate projections. Since the Borel σ-field on C with the uniform norm coincides with the Kolmogorov σ-field generated by the coordinate projections, the finite-dimensional distributions also determine the distribution of a stochastic process X with sample paths in C.

However, convergence of the finite-dimensional distributions is *not* strong enough to imply convergence in distribution $X_n \Rightarrow X$ of the random elements of C.

Example 11.6.1. *Convergence of finite-dimensional distributions is not enough.* To see that convergence of the finite-dimensional distributions does not imply convergence in distribution $X_n \Rightarrow X$ for random elements of C, it suffices to consider a deterministic example. Let $P(X = x) = 1$ and $P(X_n = x_n) = 1$ for all $n \geq 1$, where $x(t) = 0, 0 \leq t \leq 1$, and

$$x_n(0) = 0, \quad x_n(n^{-1}) = 1 \quad \text{and} \quad x_n(2n^{-1}) = x_n(1) = 0 \ ,$$

with x_n defined by linear interpolation elsewhere. Clearly, $x_n(t) \to x(t)$ pointwise as $n \to \infty$, but $\| x_n - x \| = 1$ for all n. Hence, (6.1) holds for all positive integers k and all k-tuples $(t_1, \ldots, t_k)$ with $0 \leq t_1 < \cdots < t_k \leq 1$, but $P(\| X_n - X \|) = 1) = 1$. ■

The observations above yield a simple criterion for convergence in distribution in C. We say that a set of random elements is tight if the associated set of image measures is tight,

Corollary 11.6.2. (criteria for convergence in distribution in C) *There is convergence $X_n \Rightarrow X$ in C if and only if the sequence $\{X_n : n \geq 1\}$ is tight and there is convergence of all the finite-dimensional distributions of X_n to those of X.*

Hence, in addition to the convergence of the finite-dimensional distributions in (6.1), we need to establish tightness of the sequence $\{\{X_n(t) : t \geq 0\} : n \geq 1\}$.

Fortunately compact subsets of the space C can be conveniently characterized by the *Arzelà-Ascoli theorem.* To state it, let $v(x;\delta)$ be a *modulus of continuity*, defined for any function x in C by

$$v(x, \delta) \equiv \sup\{|x(t_1) - x(t_2)| : \quad 0 \leq t_1 < t_2 \leq 1, \quad |t_1 - t_2| < \delta\} \ . \qquad (6.2)$$

Theorem 11.6.2. (Arzelà-Ascoli theorem) *A subset A of C has compact closure if and only if*

$$\sup_{x \in A} x(0) < \infty$$

and

$$\lim_{\delta \to 0} \quad \sup_{x \in A} v(x, \delta) = 0 \ .$$

From the Arzelà-Ascoli theorem we easily obtain criteria for a sequence of probability measures on C to be tight.

Theorem 11.6.3. (tightness criterion for random elements of C) *A sequence $\{X_n : n \geq 1\}$ of random elements of C is tight if and only if, for every $\epsilon > 0$, there exists a constant c such that*

$$P(|X_n(0)| > c) < \epsilon \quad \textit{for all} \quad n \geq 1, \tag{6.3}$$

and, for every $\epsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ and n_0 such that

$$P(v(X_n, \delta) \geq \epsilon) \leq \eta \quad \textit{for all} \quad n \geq n_0. \tag{6.4}$$

The tightness criterion in Theorem 11.6.3 in turn give us convenient necessary and sufficient conditions for convergence in distribution for random elements of C.

Theorem 11.6.4. (criteria for convergence in distribution in C) *There is convergence in distribution $X_n \Rightarrow X$ in C if and only if* (6.1), (6.3) *and* (6.4) *all hold.*

The modulus inequality in (6.4) can in turn be translated into various probability and moment inequalities. The following is a consequence of Theorem 12.3 of Billingsley (1968).

Theorem 11.6.5. (moment criterion for tightness in C) *A sequence $\{X_n : n \geq 1\}$ of random elements of $C \equiv C([0,1], \mathbb{R})$ is tight if $\{X_n(0)\}$ is tight in $\mathbb{R}$ and there exist constants $\gamma \geq 0$ and $\alpha > 1$ and a nondecreasing continuous function g on $[0,1]$ such that*

$$E[|X_n(t) - X_n(s)|^{\gamma}] \leq |g(t) - g(s)|^{\alpha} \tag{6.5}$$

for $0 \leq s \leq t \leq 1$.

The compactness approach to establish stochastic-process limits in (C, U) and (D, J_1) is developed in detail in Billingsley (1968). As illustrated by the treatment of the J_1 topology in Billingsley (1968), there are related criteria for convergence $X_n \Rightarrow X$ in $D \equiv D([0,1], \mathbb{R})$ with the non-uniform Skorohod topologies. Because of the discontinuities, we want to require convergence of the finite-dimensional distributions only for time points t that are almost surely continuity points of the limit process X, i.e., for which $P(t \in Disc(X)) = 0$, where $Disc(x)$ is the set of discontinuity points of x, and that suffices. Let

$$T_X \equiv \{t > 0 : P(t \in Disc(X)) = 0\} \cup \{1\} .$$

Theorem 11.6.6. (criteria for convergence in distribution in D) *There is convergence in distribution $X_n \Rightarrow X$ in D with one of the Skorohod non-uniform topologies if* (6.1) *holds for all $t_i \in T_X$ and $\{X_n : n \geq 1\}$ is tight with respect to the topology. The conditions are necessary for the J_1 and M_1 topologies.*

From Theorem 11.6.1, we know that necessity in Theorem 11.6.6 depends on the space being Polish. Since we have separability, it remains to establish topological completeness. Topological completeness for J_1 was demonstrated by Kolmogorov (1956) and Prohorov (1956); see Billingsley (1968) for a different approach. For the M_1 topology, we establish topological completeness in Section 12.8 by using Prohorov's argument.

We exploit analogs of the Arzelà-Ascoli theorem characterizing compact subsets of D with the relevant non-uniform Skorohod topology. For that purpose, we exploit generalizations of the modulus of continuity $v(x, \delta)$ in (6.2). The compactness-approach to stochastic-process limits via Prohorov's theorem explains our interest in characterizing compact subsets of D in Section 12.12.

We conclude this section by establishing an elementary result about tightness on product spaces. The result applies to finite or countably infinite products.

Theorem 11.6.7. (tightness on product spaces) *Let $S \equiv \prod_{i=1}^{\infty} S_i$ be a product of separable metric spaces with the product topology. A set A of probability measures on S is tight if and only if the sets $A_i \equiv \{P\pi_i^{-1} : P \in A\}$ of marginal probability measures on S_i, where π_i is the i^{th} coordinate projection map, are tight for all i.*

Proof. First suppose that $A \in \mathcal{P}(S)$ is tight. Let ϵ be given. Thus there is a compact subst K in S with $P(K) > 1 - \epsilon$ for all $P \in A$. Then, for each i, $\pi_i(K)$ is compact in S_i and $K \subseteq \pi_i^{-1}(\pi_i(K))$, so that

$$P\pi_i^{-1}(\pi_i(K)) \geq P(K) > 1 - \epsilon \ .$$

Second, suppose that $P\pi_i^{-1}$ is tight for each i. For $\epsilon > 0$ given, choose compact K_i in S_i such that $P\pi_i^{-1}(K_i) > 1 - \epsilon 2^{-i}$ for all i. By Tychonoff's theorem, $K \equiv \prod_{i=1}^{\infty} K_i$ is compact in S. Moreover,

$$P(K^c) \leq \sum_{i=1}^{\infty} P\pi_i^{-1}(K_i^c) \leq \epsilon \sum_{i=1}^{\infty} 2^{-i} = \epsilon \ . \quad \blacksquare$$