18

Stochastic Convergence in Metric Spaces

This chapter extends the concepts of convergence in distribution, in probability, and almost surely from Euclidean spaces to more abstract metric spaces. We are particularly interested in developing the theory for random functions, or stochastic processes, viewed as elements of the metric space of all bounded functions.

18.1 Metric and Normed Spaces

In this section we recall some basic topological concepts and introduce a number of examples of metric spaces.

A *metric space* is a set $\mathbb{D}$ equipped with a metric. A *metric* or *distance function* is a map $d: \mathbb{D} \times \mathbb{D} \mapsto [0, \infty)$ with the properties

(i) $d(x, y) = d(y, x)$;
(ii) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality);
(iii) $d(x, y) = 0$ if and only if $x = y$.

A *semimetric* satisfies (i) and (ii), but not necessarily (iii). An *open ball* is a set of the form $\{y : d(x, y) < r\}$. A subset of a metric space is *open* if and only if it is the union of open balls; it is *closed* if and only if its complement is open. A sequence x_n *converges* to x if and only if $d(x_n, x) \to 0$; this is denoted by $x_n \to x$. The *closure* $\overline{A}$ of a set $A \subset \mathbb{D}$ consists of all points that are the limit of a sequence in A; it is the smallest closed set containing A. The *interior* $\mathring{A}$ is the collection of all points x such that $x \in G \subset A$ for some open set G; it is the largest open set contained in A. A function $f: \mathbb{D} \mapsto \mathbb{E}$ between two metric spaces is *continuous* at a point x if and only if $f(x_n) \to f(x)$ for every sequence $x_n \to x$; it is continuous at every x if and only if the inverse image $f^{-1}(G)$ of every open set $G \subset \mathbb{E}$ is open in $\mathbb{D}$. A subset of a metric space is *dense* if and only if its closure is the whole space. A metric space is *separable* if and only if it has a countable dense subset. A subset K of a metric space is *compact* if and only if it is closed and every sequence in K has a converging subsequence. A subset K is *totally bounded* if and only if for every $\varepsilon > 0$ it can be covered by finitely many balls of radius ε. A semimetric space is *complete* if every *Cauchy sequence*, a sequence such that $d(x_n, x_m) \to 0$ as $n, m \to \infty$, has a limit. A subset of a complete semimetric space is compact if and only if it is totally bounded and closed.

A *normed space* $\mathbb{D}$ is a vector space equipped with a norm. A *norm* is a map $\|\cdot\|: \mathbb{D} \mapsto [0, \infty)$ such that, for every x, y in $\mathbb{D}$, and $\alpha \in \mathbb{R}$,

(i) $\|x + y\| \le \|x\| + \|y\|$ (triangle inequality);
(ii) $\|\alpha x\| = |\alpha|\,\|x\|$;
(iii) $\|x\| = 0$ if and only if $x = 0$.

A *seminorm* satisfies (i) and (ii), but not necessarily (iii). Given a norm, a metric can be defined by $d(x, y) = \|x - y\|$.

18.1 Definition. The *Borel σ-field* on a metric space $\mathbb{D}$ is the smallest σ-field that contains the open sets (and then also the closed sets). A function defined relative to (one or two) metric spaces is called *Borel-measurable* if it is measurable relative to the Borel σ-field(s). A Borel-measurable map $X : \Omega \mapsto \mathbb{D}$ defined on a probability space $(\Omega, \mathcal{U}, \mathrm{P})$ is referred to as a *random element* with values in $\mathbb{D}$.

For Euclidean spaces, Borel measurability is just the usual measurability. Borel measurability is probably the natural concept to use with metric spaces. It combines well with the topological structure, particularly if the metric space is separable. For instance, continuous maps are Borel-measurable.

18.2 *Lemma.* *A continuous map between metric spaces is Borel-measurable.*

Proof. A map $g : \mathbb{D} \mapsto \mathbb{E}$ is continuous if and only if the inverse image $g^{-1}(G)$ of every open set $G \subset \mathbb{E}$ is open in $\mathbb{D}$. In particular, for every open G the set $g^{-1}(G)$ is a Borel set in $\mathbb{D}$. By definition, the open sets in $\mathbb{E}$ generate the Borel σ-field. Thus, the inverse image of a generator of the Borel sets in $\mathbb{E}$ is contained in the Borel σ-field in $\mathbb{D}$. Because the inverse image $g^{-1}(\mathcal{G})$ of a generator $\mathcal{G}$ of a σ-field $\mathcal{B}$ generates the σ-field $g^{-1}(\mathcal{B})$, it follows that the inverse image of every Borel set is a Borel set. ■

18.3 *Example (Euclidean spaces).* The Euclidean space $\mathbb{R}^k$ is a normed space with respect to the Euclidean norm (whose square is $\|x\|^2 = \sum_{i=1}^k x_i^2$), but also with respect to many other norms, for instance $\|x\| = \max_i |x_i|$, all of which are equivalent. By the Heine-Borel theorem a subset of $\mathbb{R}^k$ is compact if and only if it is closed and bounded. A Euclidean space is separable, with, for instance, the vectors with rational coordinates as a countable dense subset.

The Borel σ-field is the usual σ-field, generated by the intervals of the type $(-\infty, x]$. □

18.4 *Example (Extended real line).* The extended real line $\overline{\mathbb{R}} = [-\infty, \infty]$ is the set consisting of all real numbers and the additional elements $-\infty$ and ∞. It is a metric space with respect to

$$d(x, y) = \bigl|\Phi(x) - \Phi(y)\bigr|.$$

Here Φ can be any fixed, bounded, strictly increasing continuous function. For instance, the normal distribution function (with $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$). Convergence of a sequence $x_n \to x$ with respect to this metric has the usual meaning, also if the limit x is $-\infty$ or ∞ (normally we would say that x_n "diverges"). Consequently, every sequence has a converging subsequence and hence the extended real line is compact. □

18.5 ***Example (Uniform norm).*** Given an arbitrary set T, let $\ell^\infty(T)$ be the collection of all bounded functions $z : T \mapsto \mathbb{R}$. Define sums $z_1 + z_2$ and products with scalars αz pointwise. For instance, $z_1 + z_2$ is the element of $\ell^\infty(T)$ such that $(z_1 + z_2)(t) = z_1(t) + z_2(t)$ for every t. The *uniform norm* is defined as

$$\|z\|_T = \sup_{t \in T} \bigl|z(t)\bigr|.$$

With this notation the space $\ell^\infty(T)$ consists exactly of all functions $z : T \mapsto \mathbb{R}$ such that $\|z\|_T < \infty$. The space $\ell^\infty(T)$ is separable if and only if T is countable. □

18.6 ***Example (Skorohod space).*** Let $T = [a, b]$ be an interval in the extended real line. We denote by $C[a, b]$ the set of all continuous functions $z : [a, b] \mapsto \mathbb{R}$ and by $D[a, b]$ the set of all functions $z : [a, b] \mapsto \mathbb{R}$ that are right continuous and whose limits from the left exist everywhere in $[a, b]$. (The functions in $D[a, b]$ are called *cadlag*: *continue à droite, limites à gauche*.) It can be shown that $C[a, b] \subset D[a, b] \subset \ell^\infty[a, b]$. We always equip the spaces $C[a, b]$ and $D[a, b]$ with the uniform norm $\|z\|_T$, which they "inherit" from $\ell^\infty[a, b]$.

The space $D[a, b]$ is referred to here as the *Skorohod space*, although Skorohod did not consider the uniform norm but equipped the space with the "Skorohod metric" (which we do not use or discuss).

The space $C[a, b]$ is separable, but the space $D[a, b]$ is not (relative to the uniform norm). □

18.7 ***Example (Uniformly continuous functions).*** Let T be a totally bounded semimetric space with semimetric ρ. We denote by $UC(T, \rho)$ the collection of all uniformly continuous functions $z : T \mapsto \mathbb{R}$. Because a uniformly continuous function on a totally bounded set is necessarily bounded, the space $UC(T, \rho)$ is a subspace of $\ell^\infty(T)$. We equip $UC(T, \rho)$ with the uniform norm.

Because a compact semimetric space is totally bounded, and a continuous function on a compact space is automatically uniformly continuous, the spaces $C(T, \rho)$ for a compact semimetric space T, for instance $C[a, b]$, are special cases of the spaces $UC(T, \rho)$. Actually, every space $UC(T, \rho)$ can be identified with a space $C(\overline{T}, \rho)$, because the *completion* $\overline{T}$ of a totally bounded semimetric T space is compact, and every uniformly continuous function on T has a unique continuous extension to the completion.

The space $UC(T, \rho)$ is separable. Furthermore, the Borel σ-field is equal to the σ-field generated by all coordinate projections (see Problem 18.3). The *coordinate projections* are the maps $z \mapsto z(t)$ with t ranging over T. These are continuous and hence always Borel-measurable. □

18.8 ***Example (Product spaces).*** Given a pair of metric spaces $\mathbb{D}$ and $\mathbb{E}$ with metrics d and e, the *Cartesian product* $\mathbb{D} \times \mathbb{E}$ is a metric space with respect to the metric

$$f\bigl((x_1, y_1), (x_2, y_2)\bigr) = d(x_1, x_2) \vee e(y_1, y_2).$$

For this metric, convergence of a sequence $(x_n, y_n) \to (x, y)$ is equivalent to both $x_n \to x$ and $y_n \to y$.

For a product metric space, there exist two natural σ-fields: The product of the Borel σ-fields and the Borel σ-field of the product metric. In general, these are not the same, the second one being bigger. A sufficient condition for them to be equal is that the metric spaces $\mathbb{D}$ and $\mathbb{E}$ are separable (e.g., Chapter 1.4 in [146])).

The possible inequality of the two σ-fields causes an inconvenient problem. If $X:\Omega\mapsto\mathbb{D}$ and $Y:\Omega\mapsto\mathbb{E}$ are Borel-measurable maps, defined on some measurable space $(\Omega,\mathcal{U})$, then $(X,Y):\Omega\mapsto\mathbb{D}\times\mathbb{E}$ is always measurable for the product of the Borel σ-fields. This is an easy fact from measure theory. However, if the two σ-fields are different, then the map (X,Y) need not be Borel-measurable. If they have separable range, then they are. $\square$

18.2 Basic Properties

In Chapter 2 convergence in distribution of random vectors is defined by reference to their distribution functions. Distribution functions do not extend in a natural way to random elements with values in metric spaces. Instead, we define convergence in distribution using one of the characterizations given by the portmanteau lemma.

A sequence of random elements X_n with values in a metric space $\mathbb{D}$ is said to *converge in distribution* to a random element X if $\mathrm{E}f(X_n)\to\mathrm{E}f(X)$ for every bounded, continuous function $f:\mathbb{D}\mapsto\mathbb{R}$. In some applications the "random elements" of interest turn out not to be Borel-measurable. To accomodate this situation, we extend the preceding definition to a sequence of *arbitrary maps* $X_n:\Omega_n\mapsto\mathbb{D}$, defined on probability spaces $(\Omega_n,\mathcal{U}_n,\mathrm{P}_n)$. Because $\mathrm{E}f(X_n)$ need no longer make sense, we replace expectations by outer expectations. For an arbitrary map $X:\Omega\mapsto\mathbb{D}$, define

$$\mathrm{E}^*f(X)=\inf\{\mathrm{E}U:U:\Omega\mapsto\mathbb{R},\text{ measurable},\ U\geq f(X),\ \mathrm{E}U\text{ exists}\}.$$

Then we say that a sequence of arbitrary maps $X_n:\Omega_n\mapsto\mathbb{D}$ *converges in distribution* to a random element X if $\mathrm{E}^*f(X_n)\to\mathrm{E}f(X)$ for every bounded, continuous function $f:\mathbb{D}\mapsto\mathbb{R}$. Here we insist that the limit X be Borel-measurable.

In the following, we do not stress the measurability issues. However, throughout we do write stars, if necessary, as a reminder that there are measurability issues that need to be taken care of. Although Ω_n may depend on n, we do not let this show up in the notation for E^* and P^*.

Next consider convergence in probability and almost surely. An arbitrary sequence of maps $X_n:\Omega_n\mapsto\mathbb{D}$ *converges in probability* to X if $\mathrm{P}^*\bigl(d(X_n,X)>\varepsilon\bigr)\to 0$ for all $\varepsilon>0$. This is denoted by $X_n\xrightarrow{\mathrm{P}}X$. The sequence X_n *converges almost surely* to X if there exists a sequence of (measurable) random variables Δ_n such that $d(X_n,X)\leq\Delta_n$ and $\Delta_n\xrightarrow{\text{as}}0$. This is denoted by $X_n\xrightarrow{\text{as}*}X$.

These definitions also do not require the X_n to be Borel-measurable. In the definition of convergence of probability we solved this by adding a star, for *outer probability*. On the other hand, the definition of almost-sure convergence is unpleasantly complicated. This cannot be avoided easily, because, even for Borel-measurable maps X_n and X, the distance $d(X_n,X)$ need not be a random variable.

The portmanteau lemma, the continuous-mapping theorem and the relations among the three modes of stochastic convergence extend without essential changes to the present definitions. Even the proofs, as given in Chapter 2, do not need essential modifications. However, we seize the opportunity to formulate and prove a refinement of the continuous-mapping theorem. The continuous-mapping theorem furnishes a more intuitive interpretation of weak convergence in terms of weak convergence of random vectors: $X_n \rightsquigarrow X$ in the metric space $\mathbb{D}$ if and only if $g(X_n) \rightsquigarrow g(X)$ for every continuous map $g : \mathbb{D} \mapsto \mathbb{R}^k$.

18.9 ***Lemma (Portmanteau).*** *For arbitrary maps $X_n : \Omega_n \mapsto \mathbb{D}$ and every random element X with values in $\mathbb{D}$, the following statements are equivalent.*

(i) $\mathrm{E}^* f(X_n) \to \mathrm{E} f(X)$ *for all bounded, continuous functions* f.
(ii) $\mathrm{E}^* f(X_n) \to \mathrm{E} f(X)$ *for all bounded, Lipschitz functions* f.
(iii) $\liminf \mathrm{P}_*(X_n \in G) \geq \mathrm{P}(X \in G)$ *for every open set* G.
(iv) $\limsup \mathrm{P}^*(X_n \in F) \leq \mathrm{P}(X \in F)$ *for every closed set* F.
(v) $\mathrm{P}^*(X_n \in B) \to \mathrm{P}(X \in B)$ *for all Borel sets B with* $\mathrm{P}(X \in \delta B) = 0$.

18.10 ***Theorem.*** *For arbitrary maps $X_n, Y_n : \Omega_n \mapsto \mathbb{D}$ and every random element X with values in $\mathbb{D}$:*

(i) $X_n \xrightarrow{\text{as}*} X$ *implies* $X_n \xrightarrow{\mathrm{P}} X$.
(ii) $X_n \xrightarrow{\mathrm{P}} X$ *implies* $X_n \rightsquigarrow X$.
(iii) $X_n \xrightarrow{\mathrm{P}} c$ *for a constant c if and only if* $X_n \rightsquigarrow c$.
(iv) if $X_n \rightsquigarrow X$ *and* $d(X_n, Y_n) \xrightarrow{\mathrm{P}} 0$, *then* $Y_n \rightsquigarrow X$.
(v) if $X_n \rightsquigarrow X$ *and* $Y_n \xrightarrow{\mathrm{P}} c$ *for a constant c, then* $(X_n, Y_n) \rightsquigarrow (X, c)$.
(vi) if $X_n \xrightarrow{\mathrm{P}} X$ *and* $Y_n \xrightarrow{\mathrm{P}} Y$, *then* $(X_n, Y_n) \xrightarrow{\mathrm{P}} (X, Y)$.

18.11 ***Theorem (Continuous mapping).*** *Let $\mathbb{D}_n \subset \mathbb{D}$ be arbitrary subsets and $g_n : \mathbb{D}_n \mapsto \mathbb{E}$ be arbitrary maps ($n \geq 0$) such that for every sequence $x_n \in \mathbb{D}_n$: if $x_{n'} \to x$ along a subsequence and $x \in \mathbb{D}_0$, then $g_{n'}(x_{n'}) \to g_0(x)$. Then, for arbitrary maps $X_n : \Omega_n \mapsto \mathbb{D}_n$ and every random element X with values in $\mathbb{D}_0$ such that $g_0(X)$ is a random element in $\mathbb{E}$:*

(i) If $X_n \rightsquigarrow X$, *then* $g_n(X_n) \rightsquigarrow g_0(X)$.
(ii) If $X_n \xrightarrow{\mathrm{P}} X$, *then* $g_n(X_n) \xrightarrow{\mathrm{P}} g_0(X)$.
(iii) If $X_n \xrightarrow{\text{as}*} X$, *then* $g_n(X_n) \xrightarrow{\text{as}*} g_0(X)$.

Proof. The proofs for $\mathbb{D}_n = \mathbb{D}$ and $g_n = g$ fixed, where g is continuous at every point of $\mathbb{D}_0$, are the same as in the case of Euclidean spaces. We prove the refinement only for (i). The other refinements are not needed in the following.

For every closed set F, we have the inclusion

$$\bigcap_{k=1}^{\infty} \overline{\bigcup_{m=k}^{\infty} \bigl\{x \in \mathbb{D}_m : g_m(x) \in F\bigr\}} \subset g_0^{-1}(F) \cup (\mathbb{D} - \mathbb{D}_0).$$

Indeed, suppose that x is in the set on the left side. Then for every k there is an $m_k \geq k$ and an element $x_{m_k} \in g_{m_k}^{-1}(F)$ with $d(x_{m_k}, x) < 1/k$. Thus, there exist a sequence $m_k \to \infty$ and elements $x_{m_k} \in \mathbb{D}_{m_k}$ with $x_{m_k} \to x$. Then either $g_{m_k}(x_{m_k}) \to g_0(x)$ or $x \notin \mathbb{D}_0$. Because the set F is closed, this implies that $g_0(x) \in F$ or $x \notin \mathbb{D}_0$.

Now, for every fixed k, by the portmanteau lemma,

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{P}^*\bigl(g_n(X_n) \in F\bigr) &\leq \limsup_{n\to\infty} \mathrm{P}^*\Bigl(X_n \in \overline{\bigcup_{m=k}^{\infty} \bigl\{x \in \mathbb{D}_m : g_m(x) \in F\bigr\}}\Bigr) \\ &\leq \mathrm{P}\bigl(X \in \overline{\bigcup_{m=k}^{\infty} g_m^{-1}(F)}\bigr).\end{aligned}$$

As $k \to \infty$, the last probability converges to $\mathrm{P}\big(X \in \bigcap_{k=1}^{\infty} \overline{\bigcup_{m=k}^{\infty} g_m^{-1}(F)}\big)$, which is smaller than or equal to $\mathrm{P}\big(g_0(X) \in F\big)$, by the preceding paragraph. Thus, $g_n(X_n) \rightsquigarrow g_0(X)$ by the portmanteau lemma in the other direction. ∎

The extension of Prohorov's theorem requires more care.† In a Euclidean space, a set is compact if and only if it is closed and bounded. In general metric spaces, a compact set is closed and bounded, but a closed, bounded set is not necessarily compact. It is the compactness that we employ in the definition of tightness. A Borel-measurable random element X into a metric space is *tight* if for every $\varepsilon > 0$ there exists a compact set K such that $\mathrm{P}(X \notin K) < \varepsilon$. A sequence of arbitrary maps $X_n : \Omega_n \mapsto \mathbb{D}$ is called *asymptotically tight* if for every $\varepsilon > 0$ there exists a compact set K such that

$$\limsup_{n\to\infty} \mathrm{P}^*(X_n \notin K^\delta) < \varepsilon, \qquad \text{every } \delta > 0.$$

Here K^δ is the δ-enlargement $\{y : d(y, K) < \delta\}$ of the set K. It can be shown that, for Borel-measurable maps in $\mathbb{R}^k$, this is identical to "uniformly tight," as defined in Chapter 2. In order to obtain a theory that applies to a sufficient number of applications, again we do not wish to assume that the X_n are Borel-measurable. However, Prohorov's theorem is true only under, at least, "measurability in the limit." An arbitrary sequence of maps X_n is called *asymptotically measurable* if

$$E^* f(X_n) - \mathrm{E}_* f(X_n) \to 0, \qquad \text{every } f \in C_b(D).$$

Here E_* denotes the inner expectation, which is defined in analogy with the outer expectation, and $C_b(\mathbb{D})$ is the collection of all bounded, continuous functions $f : \mathbb{D} \mapsto \mathbb{R}$. A Borel-measurable sequence of random elements X_n is certainly asymptotically measurable, because then both the outer and the inner expectations in the preceding display are equal to the expectation, and the difference is identically zero.

18.12 ***Theorem (Prohorov's theorem).*** *Let $X_n : \Omega_n \to \mathbb{D}$ be arbitrary maps into a metric space.*

(i) If $X_n \rightsquigarrow X$ for some tight random element X, then $\{X_n : n \in \mathbb{N}\}$ is asymptotically tight and asymptotically measurable.

(ii) If X_n is asymptotically tight and asymptotically measurable, then there is a subsequence and a tight random element X such that $X_{n_j} \rightsquigarrow X$ as $j \to \infty$.

18.3 Bounded Stochastic Processes

A *stochastic process* $X = \{X_t : t \in T\}$ is a collection of random variables $X_t : \Omega \mapsto \mathbb{R}$, indexed by an arbitrary set T and defined on the same probability space $(\Omega, \mathcal{U}, \mathrm{P})$. For a fixed ω, the map $t \mapsto X_t(\omega)$ is called a *sample path*, and it is helpful to think of X as a random function, whose realizations are the sample paths, rather than as a collection of random variables. If every sample path is a bounded function, then X can be viewed as a

† The following Prohorov's theorem is not used in this book. For a proof see, for instance, [146].

map $X:\Omega \mapsto \ell^\infty(T)$. If $T=[a,b]$ and the sample paths are continuous or cadlag, then X is also a map with values in $C[a,b]$ or $D[a,b]$.

Because $C[a,b] \subset D[a,b] \subset \ell^\infty[a,b]$, we can consider the weak convergence of a sequence of maps with values in $C[a,b]$ relative to $C[a,b]$, but also relative to $D[a,b]$, or $\ell^\infty[a,b]$. The following lemma shows that this does not make a difference, as long as we use the uniform norm for all three spaces.

18.13 ***Lemma.*** *Let $\mathbb{D}_0 \subset \mathbb{D}$ be arbitrary metric spaces equipped with the same metric. If X and every X_n take their values in $\mathbb{D}_0$, then $X_n \rightsquigarrow X$ as maps in $\mathbb{D}_0$ if and only if $X_n \rightsquigarrow X$ as maps in $\mathbb{D}$.*

Proof. Because a set G_0 in $\mathbb{D}_0$ is open if and only if it is of the form $G \cap \mathbb{D}_0$ for an open set G in $\mathbb{D}$, this is an easy corollary of (iii) of the portmanteau lemma. ∎

Thus, we may concentrate on weak convergence in the space $\ell^\infty(T)$, and automatically obtain characterizations of weak convergence in $C[a,b]$ or $D[a,b]$. The next theorem gives a characterization by *finite approximation.* It is required that, for any $\varepsilon > 0$, the index set T can be partitioned into finitely many sets $T_1, \dots, T_k$ such that (asymptotically) the variation of the sample paths $t \mapsto X_{n,t}$ is less than ε on every one of the sets T_i, with large probability. Then the behavior of the process can be described, within a small error margin, by the behavior of the *marginal vectors* $(X_{n,t_1}, \dots, X_{n,t_k})$ for arbitrary fixed points $t_i \in T_i$. If these marginals converge, then the processes converge.

18.14 ***Theorem.*** *A sequence of arbitrary maps $X_n : \Omega_n \mapsto \ell^\infty(T)$ converges weakly to a tight random element if and only if both of the following conditions hold:*

(i) The sequence $(X_{n,t_1}, \dots, X_{n,t_k})$ converges in distribution in $\mathbb{R}^k$ for every finite set of points $t_1, \dots, t_k$ in T;

(ii) for every $\varepsilon, \eta > 0$ there exists a partition of T into finitely many sets $T_1, \dots, T_k$ such that

$$\limsup_{n\to\infty} \mathrm{P}^*\Bigl(\sup_i \sup_{s,t\in T_i} |X_{n,s} - X_{n,t}| \geq \varepsilon\Bigr) \leq \eta.$$

Proof. We only give the proof of the more constructive part, the sufficiency of (i) and (ii). For each natural number m, partition T into sets $T_1^m, \dots, T_{k_m}^m$, as in (ii) corresponding to $\varepsilon = \eta = 2^{-m}$. Because the probabilities in (ii) decrease if the partition is refined, we can assume without loss of generality that the partitions are successive refinements as m increases. For fixed m define a semimetric ρ_m on T by $\rho_m(s,t) = 0$ if s and t belong to the same partioning set T_j^m, and by $\rho_m(s,t) = 1$ otherwise. Every ρ_m-ball of radius $0 < \varepsilon < 1$ coincides with a partitioning set. In particular, T is totally bounded for ρ_m, and the ρ_m-diameter of a set T_j^m is zero. By the nesting of the partitions, $\rho_1 \leq \rho_2 \leq \cdots$. Define $\rho(s,t) = \sum_{m=1}^\infty 2^{-m}\rho_m(s,t)$. Then ρ is a semimetric such that the ρ-diameter of T_j^m is smaller than $\sum_{k>m} 2^{-k} = 2^{-m}$, and hence T is totally bounded for ρ. Let T_0 be the countable ρ-dense subset constructed by choosing an arbitrary point t_j^m from every T_j^m.

By assumption (i) and Kolmogorov's consistency theorem (e.g., [133, p. 244] or [42, p. 347]), we can construct a stochastic process $\{X_t : t \in T_0\}$ on some probability space such that $(X_{n,t_1}, \dots, X_{n,t_k}) \rightsquigarrow (X_{t_1}, \dots, X_{t_k})$ for every finite set of points $t_1, \dots, t_k$ in T_0. By the portmanteau lemma and assumption (ii), for every finite set $S \subset T_0$,

$$\mathrm{P}\left(\sup_j \sup_{\substack{s,t \in T_j^m \\ s,t \in S}} |X_s - X_t| > 2^{-m}\right) \le 2^{-m}.$$

By the monotone convergence theorem this remains true if S is replaced by T_0. If $\rho(s,t) < 2^{-m}$, then $\rho_m(s,t) < 1$ and hence s and t belong to the same partitioning set T_j^m. Consequently, the event in the preceding display with $S = T_0$ contains the event in the following display, and

$$\mathrm{P}\left(\sup_{\substack{\rho(s,t)<2^{-m} \\ s,t \in T_0}} |X_s - X_t| > 2^{-m}\right) \le 2^{-m}.$$

This sums to a finite number over $m \in \mathbb{N}$. Hence, by the Borel-Cantelli lemma, for almost all ω, $\left|X_s(\omega) - X_t(\omega)\right| \le 2^{-m}$ for all $\rho(s,t) < 2^{-m}$ and all sufficiently large m. This implies that almost all sample paths of $\{X_t : t \in T_0\}$ are contained in $UC(T_0, \rho)$. Extend the process by continuity to a process $\{X_t : t \in T\}$ with almost all sample paths in $UC(T, \rho)$.

Define $\pi_m : T \mapsto T$ as the map that maps every partioning set T_j^m onto the point $t_j^m \in T_j^m$. Then, by the uniform continuity of X, and the fact that the ρ-diameter of T_j^m is smaller than 2^{-m}, $X \circ \pi_m \rightsquigarrow X$ in $\ell^\infty(T)$ as $m \to \infty$ (even almost surely). The processes $\left\{X_n \circ \pi_m(t) : t \in T\right\}$ are essentially k_m-dimensional vectors. By (i), $X_n \circ \pi_m \rightsquigarrow X \circ \pi_m$ in $\ell^\infty(T)$ as $n \to \infty$, for every fixed m. Consequently, for every Lipschitz function $f : \ell^\infty(T) \mapsto [0,1]$, $\mathrm{E}^* f(X_n \circ \pi_m) \to \mathrm{E} f(X)$ as $n \to \infty$, followed by $m \to \infty$. Conclude that, for every $\varepsilon > 0$,

$$\begin{aligned}\left|\mathrm{E}^* f(X_n) - \mathrm{E} f(X)\right| &\le \left|\mathrm{E}^* f(X_n) - \mathrm{E}^* f(X_n \circ \pi_m)\right| + o(1) \\ &\le \|f\|_{\mathrm{lip}}\varepsilon + \mathrm{P}^*\left(\|X_n - X_n \circ \pi_m\|_T > \varepsilon\right) + o(1).\end{aligned}$$

For $\varepsilon = 2^{-m}$ this is bounded by $\|f\|_{\mathrm{lip}} 2^{-m} + 2^{-m} + o(1)$, by the construction of the partitions. The proof is complete. ■

In the course of the proof of the preceding theorem a semimetric ρ is constructed such that the weak limit X has uniformly ρ-continuous sample paths, and such that (T, ρ) is totally bounded. This is surprising: even though we are discussing stochastic processes with values in the very large space $\ell^\infty(T)$, the limit is concentrated on a much smaller space of continuous functions. Actually, this is a consequence of imposing the condition (ii), which can be shown to be equivalent to asymptotic tightness. It can be shown, more generally, that every tight random element X in $\ell^\infty(T)$ necessarily concentrates on $UC(T, \rho)$ for some semimetric ρ (depending on X) that makes T totally bounded.

In view of this connection between the partitioning condition (ii), continuity, and tightness, we shall sometimes refer to this condition as the condition of *asymptotic tightness* or *asymptotic equicontinuity*.

We record the existence of the semimetric for later reference and note that, for a Gaussian limit process, this can always be taken equal to the "intrinsic" standard deviation semimetric.

18.15 ***Lemma.*** *Under the conditions (i) and (ii) of the preceding theorem there exists a semimetric ρ on T for which T is totally bounded, and such that the weak limit of the* *sequence X_n can be constructed to have almost all sample paths in $UC(T, \rho)$. Furthermore, if the weak limit X is zero-mean Gaussian, then this semimetric can be taken equal to $\rho(s,t) = \mathrm{sd}(X_s - X_t)$.*

Proof. A semimetric ρ is constructed explicitly in the proof of the preceding theorem. It suffices to prove the statement concerning Gaussian limits X.

Let ρ be the semimetric obtained in the proof of the theorem and let ρ_2 be the standard deviation semimetric. Because every uniformly ρ-continuous function has a unique continuous extension to the ρ-completion of T, which is compact, it is no loss of generality to assume that T is ρ-compact. Furthermore, assume that *every* sample path of X is ρ-continuous.

An arbitrary sequence t_n in T has a ρ-converging subsequence $t_{n'} \to t$. By the ρ-continuity of the sample paths, $X_{t_{n'}} \to X_t$ almost surely. Because every X_t is Gaussian, this implies convergence of means and variances, whence $\rho_2(t_{n'}, t)^2 = \mathrm{E}(X_{t_{n'}} - X_t)^2 \to 0$ by Proposition 2.29. Thus $t_{n'} \to t$ also for ρ_2 and hence T is ρ_2-compact.

Suppose that a sample path $t \mapsto X_t(\omega)$ is not ρ_2-continuous. Then there exists an $\varepsilon > 0$ and a $t \in T$ such that $\rho_2(t_n, t) \to 0$, but $\left|X_{t_n}(\omega) - X_t(\omega)\right| \geq \varepsilon$ for every n. By the ρ-compactness and continuity, there exists a subsequence such that $\rho(t_{n'}, s) \to 0$ and $X_{t_{n'}}(\omega) \to X_s(\omega)$ for some s. By the argument of the preceding paragraph, $\rho_2(t_{n'}, s) \to 0$, so that $\rho_2(s,t) = 0$ and $\left|X_s(\omega) - X_t(\omega)\right| \geq \varepsilon$. Conclude that the path $t \mapsto X_t(\omega)$ can only fail to be ρ_2-continuous for ω for which there exist $s, t \in T$ with $\rho_2(s,t) = 0$, but $X_s(\omega) \neq X_t(\omega)$. Let N be the set of ω for which there do exist such s, t. Take a countable, ρ-dense subset A of $\left\{(s,t) \in T \times T : \rho_2(s,t) = 0\right\}$. Because $t \mapsto X_t(\omega)$ is ρ-continuous, N is also the set of all ω such that there exist $(s,t) \in A$ with $X_s(\omega) \neq X_t(\omega)$. From the definition of ρ_2, it is clear that for every fixed (s,t), the set of ω such that $X_s(\omega) \neq X_t(\omega)$ is a null set. Conclude that N is a null set. Hence, almost all paths of X are ρ_2-continuous. ∎

Notes

The theory in this chapter was developed in increasing generality over the course of many years. Work by Donsker around 1950 on the approximation of the empirical process and the partial sum process by the Brownian bridge and Brownian motion processes was an important motivation. The first type of approximation is discussed in Chapter 19. For further details and references concerning the material in this chapter, see, for example, [76] or [146].

PROBLEMS

1. (i) Show that a compact set is totally bounded.
(ii) Show that a compact set is separable.

2. Show that a function $f : \mathbb{D} \mapsto \mathbb{E}$ is continuous at every $x \in \mathbb{D}$ if and only if $f^{-1}(G)$ is open in $\mathbb{D}$ for every open $G \in \mathbb{E}$.

3. **(Projection σ-field.)** Show that the σ-field generated by the coordinate projections $z \mapsto z(t)$ on $C[a,b]$ is equal to the Borel σ-field generated by the uniform norm. (First, show that the space

$C[a, b]$ is separable. Next show that every open set in a separable metric space is a *countable* union of open balls. Next, it suffices to prove that every open ball is measurable for the projection σ-field.)

4. Show that $D[a, b]$ is not separable for the uniform norm.

5. Show that every function in $D[a, b]$ is bounded.

6. Let h be an arbitrary element of $D[-\infty, \infty]$ and let $\varepsilon > 0$. Show that there exists a grid $u_0 = -\infty < u_1 < \cdots u_m = \infty$ such that h varies at most ε on every interval $[u_i, u_{i+1})$. Here "varies at most ε" means that $|h(u) - h(v)|$ is less than ε for every u, v in the interval. (Make sure that all points at which h jumps more than ε are grid points.)

7. Suppose that H_n and H_0 are subsets of a semimetric space H such that $H_n \to H_0$ in the sense that

(i) Every $h \in H_0$ is the limit of a sequence $h_n \in H_n$;

(ii) If a subsequence h_{n_j} converges to a limit h, then $h \in H_0$.

Suppose that Λ_n are stochastic processes indexed by H that converge in distribution in the space $\ell^\infty(H)$ to a stochastic process Λ that has uniformly continuous sample paths. Show that $\sup_{h\in H_n} \Lambda_n(h) \rightsquigarrow \sup_{h\in H_0} \Lambda(h)$.