Chapter 8

More on Continuity

Section 8.1 constructs separable modifications of reasonable but non-separable random functions, and explains how separability relates to non-denumerable properties like continuity.

Section 8.2 constructs versions of our favorite one-parameter processes where the sample paths are measurable functions of the parameter.

Section 8.3 gives conditions for the existence of cadlag versions.

Section 8.4 gives some criteria for continuity, and for the existence of "continuous modifications" of discontinuous processes.

Recall the story so far: last time we saw that the existence of processes with given finite-dimensional distributions does not guarantee that they have desirable and natural properties, like continuity, and in fact that one can construct discontinuous versions of processes which *ought* to be continuous. We therefore need extra theorems to guarantee the existence of *continuous* versions of processes with specified FDDs. To get there, we will first prove the existence of *separable* versions. This will require various topological conditions on both the index set T and the value space Ξ.

In the interest of space (or is it time?), Section 8.1 will provide complete and detailed proofs. The other sections will simply state results, and refer proofs to standard sources, mostly Gikhman and Skorokhod (1965/1969). (They in turn follow Doob (1953), but are explicit about what he regarded as obvious generalizations and extensions, *and* they cost about \$20, whereas Doob costs \$120 in paperback.)

8.1 Separable Versions

We can show that separable versions of our favorite stochastic processes exist under quite general conditions, but first we will need some preliminary results, living at the border between topology and measure theory. This starts by recalling some facts about compact spaces.

Definition 82 (Compactness, Compactification) *A set A in a topological space Ξ is* compact *if every covering of A by open sets contains a finite sub-cover. Ξ is a* compact space *if it is itself a compact set. Every non-compact topological space Ξ is a sub-space of some compact topological space $\tilde{\Xi}$. The super-space $\tilde{\Xi}$ is a* compactification *of Ξ. Every compact metric space is separable.*[1]

Example 83 *The real numbers $\mathbb{R}$ are not compact: they have no finite covering by open intervals (or other open sets). The extended reals, $\overline{\mathbb{R}} \equiv \mathbb{R} \cup +\infty \cup -\infty$, are compact, since intervals of the form $(a, \infty]$ and $[-\infty, a)$ are open. This is a* two-point compactification *of the reals. There is also a one-point compactification, with a single point at $\pm\infty$, but this has the undesirable property of making big negative and positive numbers close to each other.*

Recall that a random function is separable if its value at any arbitrary index can be determined almost surely by examining its values on some fixed, countable collection of indices. The next lemma states an alternative characterization of separability. The lemma after that gives conditions under which a weaker property holds — the almost-sure determination of whether $X(t, \omega) \in B$, for a specific t and set B, by the behavior of $X(t_n, \omega)$ at countably many t_n. The final lemma extends this to large collections of sets, and then the proof of the theorem puts all the parts together.

Lemma 84 *Let T be a separable set, Ξ a compact metric space, and D a countable dense subset of T. Define V as the class of all open balls in T centered at points in D and with rational radii. For any $G \subset T$, let*

$$R(G, \omega) \equiv \text{closure}\left(\bigcup_{t \in G \cap D} X(t, \omega) \right) \quad (8.1)$$

$$R(t, \omega) \equiv \bigcap_{S:\ S \in V,\ t \in S} R(S, \omega) \quad (8.2)$$

Then $X(t, \omega)$ is D-separable if and only if there exists a set $N \subset \Omega$ such that

$$\omega \notin N \quad \Rightarrow \quad \forall t,\ X(t, \omega) \in R(t, \omega) \quad (8.3)$$

and $\mathbb{P}(N) = 0$.

PROOF: Roughly speaking, $R(t, \omega)$ is what we'd think the range of the function would be, in the vicinity of t, if it we went just by what it did at points in the separating set D. The actual value of the function falling into this range (almost surely) is necessary and sufficient for the function to be separable. But let's speak less roughly.

"Only if": Since $X(t, \omega)$ is D-separable, for almost all ω, for any t there is some sequence $t_n \in D$ such that $t_n \to t$ and $X(t_n, \omega) \to X(t, \omega)$. For any

[1] This last statement requires the axiom of choice.

ball S centered at t, there is some N such that $t_n \in S$ if $n \geq N$. Hence the values of $x(t_n)$ are eventually confined to the set $\bigcup_{t \in S \cap D} X(t,\omega)$. Recall that the closure of a set A consists of the points x such that, for some sequence $x_n \in A$, $x_n \rightarrow x$. As $X(t_n,\omega) \rightarrow X(t,\omega)$, it must be the case that $X(t,\omega) \in \text{closure}\left(\bigcup_{t \in S \cap D} X(t,\omega)\right)$. Since this applies to all S, $X(t,\omega)$ must be in the intersection of all those closures, hence $X(t,\omega) \in R(t,\omega)$ — unless we are on one of the probability-zero bad sample paths, i.e., unless $\omega \in N$.

"If": Assume that, with probability 1, $X(t,\omega) \in R(t,\omega)$. Thus, for any $S \in V$, we know that there exists a sequence of points $t_n \in S \cap D$ such that $X(t_n,\omega) \rightarrow X(t,\omega)$. However, this doesn't say that $t_n \rightarrow t$, which is what we need for separability. We will now build such a sequence. Consider a series of spheres $S_k \in V$ such that (i) every point in S_k is within a distance 2^{-k} of t and (ii) $S_{k+1} \subset S_k$. For each S_k, there is a sequence $t_n^{(k)} \in S_k$ such that $X(t_n^{(k)},\omega) \rightarrow X(t,\omega)$. In fact, for any $m > 0$, $|X(t_n^{(k)},\omega) - X(t,\omega)| < 2^{-m}$ if $n \geq N(k,m)$, for some $N(k,m)$. Our final sequence of indices t_i then consists of the following points: $t_n^{(1)}$ for n from $N(1,1)$ to $N(1,2)$; $t_n^{(2)}$ for n from $N(2,2)$ to $N(2,3)$; and in general $t_n^{(k)}$ for n from $N(k,k)$ to $N(k,k+1)$. Clearly, $t_i \rightarrow t$, and $X(t_i,\omega) \rightarrow X(t,\omega)$. Since every $t_i \in D$, we have shown that $X(t,\omega)$ is D-separable. □

Lemma 85 *Let T be a separable index set, Ξ a compact space, X a random function from T to Ξ, and B be an arbitrary Borel set of Ξ. Then there exists a denumerable set of points $t_n \in T$ such that, for any $t \in T$, the set*

$$N(t,B) \equiv \{\omega : X(t,\omega) \not\in B\} \cap \left(\bigcap_{n=1}^{\infty} \{\omega : X(t_n,\omega) \in B\}\right) \tag{8.4}$$

has probability 0.

PROOF: We proceed recursively. The first point, t_1, can be whatever we like. Suppose $t_1, t_2, \ldots t_n$ are already found, and define the following:

$$M_n \equiv \bigcap_{k=1}^{n} \{\omega : X(t_k,\omega) \in B\} \tag{8.5}$$

$$L_n(t) \equiv M_n \cap \{\omega : X(t,\omega) \not\in B\} \tag{8.6}$$

$$p_n \equiv \sup_t \mathbb{P}(L_n(t)) \tag{8.7}$$

M_n is the set where the random function, evaluated at the first n indices, gives a value in our favorite set; it's clearly measurable. $L_n(t)$, also clearly measurable, gives the collection of points in Ω where, if we chose t for the next point in the collection, this will break down. p_n is the worst-case probability of this happening. For each t, $L_{n+1}(t) \subseteq L_n(t)$, so $p_{n+1} \leq p_n$. Suppose $p_n = 0$; then we've found the promised denumerable sequence, and we're done. Suppose instead that $p_n > 0$. Pick any t such that $\mathbb{P}(L_n(t)) \geq \frac{1}{2}p_n$, and call it t_{n+1}. (There has to be such a point, or else p_n wouldn't be the supremum.) Now notice that $L_1(t_2)$, $L_2(t_3)$, $\ldots L_n(t_{n+1})$ are all mutually exclusive, but not necessarily jointly exhaustive. So

$$\begin{aligned}
1 &= \mathbb{P}(\Omega) && (8.8)\\
&\geq \mathbb{P}\left(\bigcup_n L_n(t_{n+1})\right) && (8.9)\\
&= \sum_n \mathbb{P}(L_n(t_{n+1})) && (8.10)\\
&\geq \sum_n \frac{1}{2} p_n > 0 && (8.11)
\end{aligned}$$

so $p_n \to 0$ as $n \to \infty$.

We saw that $L_n(t)$ is a monotone-decreasing sequence of sets, for each t, so a limiting set exists, and in fact $\lim_n L_n(t) = N(t, B)$. So, by monotone convergence,

$$\begin{aligned}
\mathbb{P}(N(t,B)) &= \mathbb{P}\left(\lim_n L_n(t)\right) && (8.12)\\
&= \lim_n \mathbb{P}(L_n(t)) && (8.13)\\
&\leq \lim_n p_n && (8.14)\\
&= 0 && (8.15)
\end{aligned}$$

as was to be shown. □

Lemma 86 *Let $\mathcal{B}_0$ be any countable class of Borel sets in Ξ, and $\mathcal{B}$ the closure of $\mathcal{B}_0$ under countable intersection. Under the hypotheses of the previous lemma, there is a denumerable sequence t_n such that, for every $t \in T$, there exists a set $N(t) \subset \Omega$ with $\mathbb{P}(N(t)) = 0$, and, for all $B \in \mathcal{B}$,*

$$\{\omega : X(t,\omega) \notin A\} \cap \left(\bigcap_{n=1}^{\infty} \{\omega : X(t_n,\omega) \in A\}\right) \subseteq N(t) \quad (8.16)$$

Proof: For each $B \in \mathcal{B}_0$, construct the sequence of indices as in the previous lemma. Since there only countably many sets in B, if we take the union of all of these sequences, we will get another countable sequence, call it t_n. Then we have that, $\forall B \in \mathcal{B}_0$, $\forall t \in T$, $\mathbb{P}(X(t_n,\omega) \in B, n \geq 1,\ X(t,\omega) \notin B) = 0$. Take this set to be $N(t,B)$, and define $N(t) \equiv \bigcup_{B \in \mathcal{B}_0} N(t,B)$. Since $N(t)$ is a countable union of probability-zero events, it is itself a probability-zero event. Now, take any $B \in \mathcal{B}$, and any $B_0 \in \mathcal{B}_0$ such that $B \subseteq B_0$. Then

$$\begin{aligned}
&\{X(t,\omega) \notin B_0\} \cap \left(\bigcap_{n=1}^{\infty} \{X(t_n,\omega) \in B\}\right) && (8.17)\\
&\subseteq \{X(t,\omega) \notin B_0\} \cap \left(\bigcap_{n=1}^{\infty} \{X(t_n,\omega) \in B_0\}\right) &&\\
&\subseteq N(t) && (8.18)
\end{aligned}$$

Since $B = \bigcap_k B_0^{(k)}$ for some sequence of sets $B_0^{(k)} \in \mathcal{B}_0$, it follows (via De Morgan's laws and the distributive law) that

$$\{X(t,\omega) \not\in B\} = \bigcup_{k=1}^{\infty} \left\{X(t,\omega) \not\in B_0^{(k)}\right\} \tag{8.19}$$

$$\begin{aligned}
&\{X(t,\omega) \not\in B\} \cap \left(\bigcap_{n=1}^{\infty} \{X(t_n,\omega) \in B\}\right) \\
&= \bigcup_{k=1}^{\infty} \left\{X(t,\omega) \not\in B_0^{(k)}\right\} \cap \left(\bigcap_{n=1}^{\infty} \{X(t_n,\omega) \in B\}\right) && (8.20)\\
&\subseteq \bigcup_{n=1}^{\infty} N(t) && (8.21)\\
&= N(t) && (8.22)
\end{aligned}$$

which was to be shown. □

Theorem 87 (Separable Versions, Separable Modifications) *Suppose that Ξ is a compact metric space and T is a separable metric space. Then, for any Ξ-valued stochastic process X on T, there exists a separable version $\tilde{X}$. This is called a* separable modification *of X.*

Proof: Let D be a countable dense subset of T, and V the class of open spheres of rational radius centered at points in D. Any open subset of T is a union of countably many sets from V, which is itself countable. Similarly, let C be a countable dense subset of Ξ, and let $\mathcal{B}_0$ consist of the complements of spheres centers at points in D with rational radii, and (as in the previous lemma) let $\mathcal{B}$ be the closure of $\mathcal{B}_0$ under countable intersection. Every closed set in Ξ belongs to $\mathcal{B}$.[2] For every $S \in V$, consider the restriction of $X(t,\omega)$ to $t \in S$, and apply Lemma 86 to the random function $X(t,\omega)$ to get a sequence of indices $I(S) \subset T$, and, for every $t \in S$, a measure-zero set $N_S(t) \subset \Omega$ where things can go wrong. Set $I = \bigcup_{S \in V} I(S)$ and $N(t) = \bigcup_{S \in V} N_S(t)$. Because V is countable, I is still a countable set of indices, and $N(t)$ is still of measure zero. I is going to be our separating set, and we're going to show that we have uncountably many sets $N(t)$ won't be a problem.

Define $\tilde{X}(t,\omega) = X(t,\omega)$ if $t \in I$ or $\omega \not\in N(t)$ — if we're at a time in the separating set, or we're at some other time but have avoided the bad set, we'll just copy our original random function. What to do otherwise, when $t \not\in I$ and $\omega \in N(t)$? Construct $R(t,\omega)$, as in the proof of Lemma 84, and let $\tilde{X}(t,\omega)$ take *any* value in this set. Since $R(t,\omega)$ depends only on the value of the function at indices in the separating set, it doesn't matter whether we build it from X or from $\tilde{X}$. In fact, for all t and ω, $\tilde{X}(t,\omega) \in R(t,\omega)$, so, by Lemma 84,

[2] *You* show this.

$\tilde{X}(t,\omega)$ is separable. Finally, for every t, $\left\{\tilde{X}(t,\omega) = X(t,\omega)\right\} \subseteq N(t)$, so $\forall t$, $\mathbb{P}\left(\tilde{X}(t) = X(t)\right)$, and $\tilde{X}$ is a version of X (Definition 74). □

Corollary 88 *If the situation is as in the previous theorem, but Ξ is not compact, there exists a separable version of X in some compactification $\tilde{\Xi}$ of Ξ.*

PROOF: Because Ξ is a sub-space of any of its compactifications $\tilde{\Xi}$, X is also a process with values in $\tilde{\Xi}$.[3] Since $\tilde{\Xi}$ is compact, X has a separable modification $\tilde{X}$ with values in $\tilde{\Xi}$, but (with probability 1) $\tilde{X}(t) \in \Xi$. □

Corollary 89 *Let Ξ be a compact metric space, T a separable index set, and μ_J, $J \in \mathrm{Fin}(T)$ a projective family of probability distributions. Then there is a separable stochastic process with finite-dimensional distributions given by μ_J.*

PROOF: Combine Theorem 87 with the Kolmogorov Extension Theorem 29. □

8.2 Measurable Versions

It would be nice for us if $X(t)$ is a measurable function *of* t, because we are going to want to write down things like

$$\int_{t=a}^{t=b} X(t)dt$$

and have them mean something. Irritatingly, this will require another modification.

Definition 90 (Measurable sample paths) *Let $T, \mathcal{T}, \tau$ be a measurable space, its σ-field and a measure defined thereon. A random function X on T with values in $\Xi, \mathcal{X}$ has* measurable sample paths *or* is measurable *if X is measurable with respect to the product σ-field $\mathcal{T} \times \mathcal{X}$, completed by null sets of the product measure $\tau \times \mathbb{P}$.*

It would seem more natural to simply define measurable sample paths by saying that $X(\cdot,\omega)$ is a $\mathcal{T}$-measurable function of t for $\mathbb{P}$-almost-all ω. However, Definition 90 implies this version, via Fubini's Theorem, and facilitates the proofs of the two following theorems.

Theorem 91 *If $X(t)$ is measurable, and $\mathbf{E}[X(t)]$ is integrable (with respect to the measure τ on T), then for any set $I \in \mathcal{T}$,*

$$\int_I \mathbf{E}[X(t)]\,\tau(dt) \quad = \quad \mathbf{E}\left[\int_I X(t)\tau(dt)\right] \tag{8.23}$$

[3] If you want to be really picky, define a 1-1 function $h : \Xi \mapsto \tilde{\Xi}$ taking points to their counterparts. Then X and $h^{-1}(X)$ are indistinguishable. Do I need to go on?

Proof: This is just Fubini's Theorem! □

Theorem 92 (Measurable Separable Modifications) *Suppose that T and Ξ are both compact. If $X(t,\omega)$ is continuous in probability at τ-almost-all t, then it has a version which is both separable and measurable, its* measurable separable modification.

Proof: See Gikhman and Skorokhod (1965/1969, ch. IV, sec. 3, thm. 1, p. 157). □

8.3 Cadlag Versions

Theorem 93 *Let X be a separable random process with $T = [a,b] \subseteq \mathbb{R}$, and Ξ a complete metric space with metric ρ. Suppose that $X(t)$ is continuous in probability on T, and there are real constants $p, q, C \geq 0$, $r > 1$ such that, for any three indices $t_1 < t_2 < t_3 \in T$,*

$$\mathbf{E}\left[\rho^p(X(t_1), X(t_2))\rho^q(X(t_2), X(t_3))\right] \leq C|t_3 - t_1|^r \tag{8.24}$$

The there is a version of X whose sample paths are cadlag (a.s.).

Proof: Combine Theorem 1 and Theorem 3 of Gikhman and Skorokhod (1965/1969, ch. IV, sec. 4, pp. 159–169). □

8.4 Continuous Modifications

Theorem 94 *Let X be a separable stochastic process with $T = [a,b] \subseteq \mathbb{R}$, and Ξ a complete metric space with metric ρ. Suppose that there are constants $C, p > 0$, $r > 1$ such that, for any $t_1 < t_2 \in T$,*

$$\mathbf{E}\left[\rho^p(X(t_1), X(t_2))\right] \leq C|t_2 - t_1|^r \tag{8.25}$$

Then $X(t)$ has a continuous version.

Proof: See Gikhman and Skorokhod (1965/1969, ch. IV, sec. 5, thm. 2, p. 170), and the first remark following the theorem. □

A slightly more refined result requires two preliminary definitions.

Definition 95 (Modulus of continuity) *For any function x from a metric space T, d to a metric space Ξ, ρ, the* modulus of continuity *is the function $m_x(r) : \mathbb{R}^+ \mapsto \mathbb{R}^+$ given by*

$$m_x(r) = \sup\left\{\rho(x(s), x(t)) : s, t \in T, \ d(s,t) \leq r\right\} \tag{8.26}$$

Lemma 96 *x is uniformly continuous if and only if its modulus of continuity $\to 0$ as $r \to 0$.*

PROOF: Obvious from Definition 95 and the definition of uniform continuity.

Definition 97 (Hölder-continuous) *Continuing the notation of Definition 95, we say that x is* Hölder-continuous with exponent c *if there are positive constants c, γ such that $m_x(r) \leq \gamma r^c$ for all sufficiently small r; i.e., $m_x(r) = O(r^c)$. If this holds on every bounded subset of T, then the function is* locally Hölder-continuous.

Theorem 98 *Let T be $\mathbb{R}^d$ and Ξ a complete metric space with metric ρ. If there are constants $p, q, \gamma > 0$, such that, for any $t_1, t_2 \in T$,*

$$\mathbf{E}\left[\rho^p(X(t_1), X(t_2))\right] \leq \gamma |t_1 - t_2|^{d+q} \tag{8.27}$$

then X has a continuous version $\tilde{X}$, and almost all sample paths of $\tilde{X}$ are locally Hölder-continuous for any exponent between 0 and q/p exclusive.

PROOF: See Kallenberg, theorem 3.23 (pp. 57–58). Note that part of Kallenberg's proof is a restricted case of what we've already done in prove the existence of a separable version! □

This lecture, the last, and even a lot of the one before have all been pretty hard and abstract. As a reward for our labor, however, we now have a collection of very important tools — operator representations, filtrations and optional times, recurrence times, and finally existence theorems for continuous processes. These are the devices which will let us take the familiar theory of elementary Markov chains, with finitely many states in discrete time, and produce the general theory of Markov processes with continuous states and/or continuous time. The next lecture will begin this work, starting with the operators.