Chapter 3

Building Infinite Processes from Regular Conditional Probability Distributions

Section 3.1 introduces the notion of a probability kernel, which is a useful way of systematizing and extending the treatment of conditional probability distributions you will have seen in 36-752.

Section 3.2 gives an extension theorem (due to Ionescu Tulcea) which lets us build infinite-dimensional distributions from a family of finite-dimensional distributions. Rather than assuming topological regularity of the space, as in Section 2.2, we assume that the FDDs can be derived from one another recursively, through applying probability kernels. This is the same as assuming regularity of the appropriate conditional probabilities.

3.1 Probability Kernels

Definition 30 (Probability Kernel) *A* measure kernel *from a measurable space $\Xi, \mathcal{X}$ to another measurable space $\Upsilon, \mathcal{Y}$ is a function $\kappa : \Xi \times \mathcal{Y} \mapsto \overline{\mathbb{R}}^+$ such that*

1. *for any $Y \in \mathcal{Y}$, $\kappa(x, Y)$ is $\mathcal{X}$-measurable; and*
2. *for any $x \in \Xi$, $\kappa(x, Y) \equiv \kappa_x(Y)$ is a measure on $\Upsilon, \mathcal{Y}$. We will write the integral of a function $f : \Upsilon \mapsto \mathbb{R}$, with respect to this measure, as $\int f(y)\kappa(x, dy)$, $\int f(y)\kappa_x(dy)$, or, most compactly, $\kappa f(x)$.*

If, in addition, κ_x is a probability measure on $\Upsilon, \mathcal{Y}$ for all x, then κ is a probability kernel.

(Some authors use "kernel", unmodified, to mean a probability kernel, and some a measure kernel; read carefully.)

Notice that we can represent any distribution on Υ as a kernel where the first argument is irrelevant: $\kappa(x_1, Y) = \kappa(x_2, Y)$ for all $x_1, x_2 \in \Xi$. The "kernels" in kernel density estimation are probability kernels, as are the stochastic transition matrices of Markov chains. (The kernels in support vector machines, however, generally are not.)

From a previous probability course, like 36-752, you will remember the measure-theoretic definition of the conditional probability of a set A, as the conditional expectation of $\mathbf{1}_A$, i.e., $\mathbb{P}(A|\mathcal{G}) = \mathbf{E}[\mathbf{1}_A(\omega)|\mathcal{G}]$. (I have explicitly included the dependence on ω to make it plain that this is a random set function.) You may also recall a potential bit of trouble, which is that even when these conditional expectations exist for all measurable sets A, it's not necessarily true that they give us a measure, i.e., that $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i|\mathcal{G}) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|\mathcal{G})$ and all the rest of it. A version of the conditional probabilities for which they do form a measure is said to be a *regular conditional probability.* Clearly, regular conditional probabilities are all probability kernels. The ordinary rules for manipulating conditional probabilities suggest how we can define the composition of kernels.

Definition 31 (Composition of probability kernels) *Let κ_1 be a kernel from Ξ to Υ, and κ_2 a kernel from $\Xi \times \Upsilon$ to Γ. Then we define $\kappa_1 \otimes \kappa_2$ as the kernel from Ξ to $\Upsilon \times \Gamma$ such that*

$$(\kappa_1 \otimes \kappa_2)(x, B) = \int \kappa_1(x, dy) \int \kappa_2(x, y, dz) \mathbf{1}_B(y, z)$$

for every measurable $B \subseteq \Upsilon \times \Gamma$ (where z ranges over the space Γ).

Verbally, κ_1 gives us a distribution on Υ, from any starting point $x \in \Xi$. Given a pair of points $(x, y) \in \Xi \times \Upsilon$, κ_2 gives a distribution on Γ. So their composition says, basically, how to chain together conditional distributions, given a starting point.

3.2 Extension via Recursive Conditioning

With the machinery of probability kernels in place, we are in a position to give an alternative extension theorem, i.e., a different way of proving the existence of stochastic processes with specified finite-dimensional marginal distributions. In Section 2.2, we assumed some topological niceness in the sample spaces, namely that they were Borel spaces. Here, instead, we will assume probabilistic niceness in the FDDs themselves, namely that they can be obtained through composing probability kernels. This is the same as assuming that they can be obtained by chaining together regular conditional probabilities. The general form of this result is attributed in the literature to Ionescu Tulcea.

Just as proving the Kolmogorov Extension Theorem needed a measure-theoretic result, the Carathéodory Extension Theorem, our proof of the Ionescu

Tulcea Extension Theorem will require a different measure-theoretic result, which is not, so far as I know, named after anyone.

Proposition 32 (Set functions continuous at $\emptyset$) *Suppose μ is a finite, non-negative, additive set function on a field $\mathcal{A}$. If, for any sequence of sets $A_n \in \mathcal{A}$, $A_n \downarrow \emptyset \implies \mu A_n \to 0$, then (1) μ is countably additive on $\mathcal{A}$, and (2) μ extends uniquely to a measure on $\sigma(A)$.*

PROOF: Part (1) is a weaker version of Theorem F in Chapter 2, §9 of Halmos, *Measure Theory* (p. 39). (When reading his proof, remember that every field of sets is also a ring of sets.) Part (2) follows from part (1) and the Carathéodory Extension Theorem (Proposition 28). □

With this preliminary out of the way, let's turn to the main event.

Theorem 33 (Ionescu Tulcea Extension Theorem) *Consider a sequence of measurable spaces $\Xi_n, \mathcal{X}_n$, $n \in \mathbb{N}$. Suppose that for each n, there exists a probability kernel κ_n from $\prod_{i=1}^{n-1} \Xi_i$ to Ξ_n (taking κ_1 to be a kernel insensitive to its first argument, i.e., a probability measure). Then there exists a sequence of random variables X_n, $n \in \mathbb{N}$, taking values in the corresponding Ξ_n, such that $\mathcal{L}(X_1, X_2, \ldots X_n) = \bigotimes_{i=1}^{n} \kappa_i$.*

PROOF: As before, we'll be working with the cylinder sets, but now we'll make our life simpler if we consider cylinders where the base set rests in the first n spaces $\Xi_1, \ldots \Xi_n$. More specifically, set $\mathcal{B}_n = \bigotimes_{i=1}^{n} \mathcal{X}_i$ (these are the base sets), and $\mathcal{C}_n = \mathcal{B}_n \times \prod_{i=n+1}^{\infty} \Xi_i$ (these are the cylinder sets), and $\mathcal{C} = \bigcup_n \mathcal{C}_n$. $\mathcal{C}$ clearly contains all the finite cylinders, so it generates the product σ-field on infinite sequences. We will use it as the field in Proposition 32. (Checking that $\mathcal{C}$ is a field is entirely parallel to checking that the $\mathcal{D}$ appearing in the proof of Theorem 29 was a field.)

For each base set $A \in \mathcal{B}_n$, let $[A]$ be the corresponding cylinder, $[A] = A \times \prod_{i=n+1}^{\infty} \Xi_i$. Notice that for every set $C \in \mathcal{C}$, there is at least one A, in some $\mathcal{B}_n$, such that $C = [A]$. Now we define a set function μ on $\mathcal{C}$.

$$\mu([A]) \quad = \quad \left(\bigotimes_{i=1}^{n} \kappa_i \right) A \tag{3.1}$$

(Checking that this is well-defined is left as an exercise, 3.2.) Clearly, this is a finite, and finitely-additive, set function defined on a field. So to use Proposition 32, we just need to check continuity from above at $\emptyset$. Let A_n be any sequence of sets such that $[A_n] \downarrow \emptyset$ and $A_n \in \mathcal{B}_n$. (Any sequence of sets in $\mathcal{C} \downarrow \emptyset$ can be massaged into this form.) We wish to show that $\mu([A_n]) \downarrow 0$. We'll get this to work by considering functions which are (pretty much) conditional probabilities for these sets:

$$p_{n|k} \quad = \quad \left(\bigotimes_{i=k+1}^{n} \kappa_i \right) \mathbf{1}_{A_n}, \ k \leq n \tag{3.2}$$

$$p_{n|n} \quad = \quad \mathbf{1}_{A_n} \tag{3.3}$$

Two facts follow immediately from the definitions:

$$p_{n|0} = \left(\bigotimes_{i=1}^{n} \kappa_i\right) \mathbf{1}_{A_n} = \mu([A_n]) \quad (3.4)$$

$$p_{n|k} = \kappa_{k+1} p_{n|k+1} \quad (3.5)$$

From the fact that the $[A_n] \downarrow \emptyset$, we know that $p_{n+1|k} \leq p_{n|k}$, for all k. This implies that $\lim_n p_{n|k} = m_k$ exists, for each k, and is approached from above. Applied to $p_{n|0}$, we see from 3.5 that $\mu([A_n]) \rightarrow m_0$. We would like $m_0 = 0$. Assume the contrary, that $m_0 > 0$. From 3.5 and the dominated convergence theorem, we can see that $m_k = \kappa_{k+1} m_{k+1}$. Hence if $m_0 > 0$, $\kappa_1 m_1 > 0$, which means (since that last expression is really an integral) that there is at least one point $x_1 \in \Xi_1$ such that $m_1(s_1) > 0$. Recursing our way down the line, we get a sequence $x = x_1, x_2, \ldots \in \Xi^{\mathbb{N}}$ such that $m_n(x_1, \ldots x_n) > 0$ for all n. But now look what we've done: for each n,

$$0 < m_n(x_1, \ldots x_n) \quad (3.6)$$

$$\leq p_{n|n}(x_1, \ldots x_n) \quad (3.7)$$

$$= \mathbf{1}_{A_n}(x_1, \ldots x_n) \quad (3.8)$$

$$= \mathbf{1}_{[A_n]}(x) \quad (3.9)$$

$$x \in [A_n] \quad (3.10)$$

This is the same as saying that $x \in \bigcap_n [A_n]$. But $[A_n] \downarrow \emptyset$, so there can be no such x. Hence $m_0 = 0$, meaning that $\mu([A_n]) \rightarrow 0$, and μ is continuous at the empty set.

Since μ is finite, finitely-additive, non-negative and continuous at $\emptyset$, by Proposition 32 it extends uniquely to a measure on the product σ-field. □

Notes on the proof: It would seem natural that one could show $m_0 = 0$ directly, rather than by contradiction, but I can't think of a way to do it, and every book I've consulted does it in exactly this way.

To appreciate the simplification made possible by the notion of probability kernels, compare this proof to the one given by Fristedt and Gray (1997, §22.1).

Notice that the Daniell, Kolmogorov and Ionescu Tulcea Extension Theorems all give *sufficient* conditions for the existence of stochastic processes, not necessary ones. The necessary and sufficient condition for extending the FDDs to a process probability measure is something called σ-smoothness. (See Pollard (2002) for details.) Generally speaking, we will deal with processes which satisfy both the Kolmogorov and the Ionescu Tulcea type conditions, e.g., real-valued Markov process.

3.3 Exercises

Exercise 3.1 (Łomnick-Ulam Theorem on infinite product measures) *Let T be an uncountable index set, and $(\Xi_t, \mathcal{X}_t, \mu_t)$ a collection of probability* *spaces. Show that there exist independent random variables X_t in Ξ_t with distributions μ_t.* Hint: *use the Ionescu Tulcea theorem on countable subsets of T, and then imitate the proof of the Kolmogorov extension theorem.*

Exercise 3.2 (Measures of cylinder sets) *In the proof of the Ionescu Tulcea Theorem, we employed a set function on the finite cylinder sets, where the measure of an infinite-dimensional cylinder set $[A]$ is taken to be the measure of its finite-dimensional base set A. However, the same cylinder set can be specified by different base sets, so it is necessary to show that Equation 3.1 has a unique value on its right-hand side. In what follows, C is an arbitrary member of the class $\mathcal{C}$.*

(i) Show that, when $A, B \in \mathcal{B}_n$, $[A] = [B]$ iff $A = B$. That is, two cylinders generated by bases of equal dimensionality are equal iff their bases are equal.

(ii) Show that there is a smallest n such that $C = [A]$ for an $A \in \mathcal{B}_n$. Conclude that the right-hand side of Equation 3.1 could be made well-defined if we took n there to be this least possible n.

(iii) Suppose that $m < n$, $A \in \mathcal{B}_m$, $B \in \mathcal{B}_n$, and $[A] = [B]$. Show that $B = A \times \prod_{i=m+1}^{n} \Xi_i$.

(iv) Continuing the situation in (iii), show that

$$\left(\bigotimes_{i=1}^{m} \kappa_i\right) A = \left(\bigotimes_{i=1}^{n} \kappa_i\right) B$$

Conclude that the right-hand side of Equation 3.1 is well-defined, as promised.