Chapter 2

Building Infinite Processes from Finite-Dimensional Distributions

Section 2.1 introduces the finite-dimensional distributions of a stochastic process, and shows how they determine its infinite-dimensional distribution.

Section 2.2 considers the consistency conditions satisfied by the finite-dimensional distributions of a stochastic process, and the extension theorems (due to Daniell and Kolmogorov) which prove the existence of stochastic processes with specified, consistent finite-dimensional distributions.

2.1 Finite-Dimensional Distributions

So, we now have X, our favorite Ξ-valued stochastic process on T with paths in U. Like any other random variable, it has a probability law or distribution, which is defined over the entire set U. Generally, this is infinite-dimensional. Since it is inconvenient to specify distributions over infinite-dimensional spaces all in a block, we consider the *finite-dimensional distributions*.

Definition 22 (Finite-dimensional distributions) *The finite-dimensional distributions of X are the the joint distributions of $X_{t_1}, X_{t_2}, \ldots X_{t_n}$, $t_1, t_2, \ldots t_n \in T$, $n \in \mathbb{N}$.*

You will sometimes see "FDDs" and "fidis" as abbreviations for "finite-dimensional distributions". Please do not use "fidis".

We can at least hope to specify the finite-dimensional distributions. But we are going to want to ask a lot of questions about asymptotics, and global properties of sample paths, which go beyond any *finite* dimension, so you might worry that we'll still need to deal directly with the infinite-dimensional distribution. The next theorem says that this worry is unfounded; the finite-dimensional distributions specify the infinite-dimensional distribution (pretty much) uniquely.

Theorem 23 (Finite-dimensional distributions determine process distributions) *Let X and Y be two Ξ-valued processes on T with paths in U. Then X and Y have the same distribution iff all their finite-dimensional distributions agree.*

Proof: "Only if": Since X and Y have the same distribution, applying the any given set of coordinate mappings will result in identically-distributed random vectors, hence all the finite-dimensional distributions will agree.

"If": We'll use the π-λ theorem.

Let $\mathcal{C}$ be the finite cylinder sets, i.e., all sets of the form

$$C = \left\{x \in \Xi^T | (x_{t_1}, x_{t_2}, \ldots x_{t_n}) \in B\right\}$$

where $n \in \mathbb{N}$, $B \in \mathcal{X}^n$, $t_1, t_2, \ldots t_n \in T$. Clearly, this is a π-system, since it is closed under intersection.

Now let $\mathcal{L}$ consist of all the sets $L \in \mathcal{X}^T$ where $\mathbb{P}(X \in L) = \mathbb{P}(Y \in L)$. We need to show that this is a λ-system, i.e., that it (i) includes Ξ^T, (ii) is closed under complementation, and (iii) is closed under monotone increasing limits. (i) is clearly true: $\mathbb{P}\left(X \in \Xi^T\right) = \mathbb{P}\left(Y \in \Xi^T\right) = 1$. (ii) is true because we're looking at a probability: if $L \in \mathcal{L}$, then $\mathbb{P}(X \in L^c) = 1 - \mathbb{P}(X \in L) = 1 - \mathbb{P}(Y \in L) = \mathbb{P}(Y \in L^c)$. To see (iii), let $L_n \uparrow L$ be a monotone-increasing sequence of sets in $\mathcal{L}$, and recall that, for any measure, $L_n \uparrow L$ implies $\mu L_n \uparrow \mu L$. So $\mathbb{P}(X \in L_n) \uparrow \mathbb{P}(X \in L)$, $\mathbb{P}(Y \in L_n) \uparrow \mathbb{P}(Y \in L)$, and (since $\mathbb{P}(X \in L_n) = \mathbb{P}(Y \in L_n)$), $\mathbb{P}(X \in L_n) \uparrow \mathbb{P}(Y \in L)$ as well. A sequence cannot have two limits, so $\mathbb{P}(X \in L) = \mathbb{P}(Y \in L)$, and $L \in \mathcal{L}$.

Since the finite-dimensional distributions match, $\mathbb{P}(X \in C) = \mathbb{P}(Y \in C)$ for all $C \in \mathcal{C}$, which means that $\mathcal{C} \subset \mathcal{L}$. Also, from the definition of the product σ-field, $\sigma(\mathcal{C}) = \mathcal{X}^T$. Hence, by the π-λ theorem, $\mathcal{X}^T \subseteq \mathcal{L}$. □

A note of caution is in order here. If X is a Ξ-valued process on T whose paths are constrained to line in U, and Y is a similar process that is not so constrained, it is nonetheless possible that X and Y agree in all their finite-dimensional distributions. The trick comes if U is not, itself, an element of $\mathcal{X}^T$. The most prominent instance of this is when $\Xi = \mathbb{R}$, $T = \mathbb{R}$, and the constraint is continuity of the sample paths: we will see that $U \notin \mathcal{B}^{\mathbb{R}}$. (This is the point of Exercise 1.1.)

2.2 Consistency and Extension

The finite-dimensional distributions of a given stochastic process are related to one another in the usual way of joint and marginal distributions. Take some collection of indices $t_1, t_2 \ldots t_n \in T$, and corresponding measurable sets

$B_1 \in \mathcal{X}_1, B_2 \in \mathcal{X}_2, \ldots B_n \in \mathcal{X}_n$. Then, for any $m > n$, and any further indices $t_{n+1}, t_{n_2}, \ldots t_m$, it must be the case that

$$\begin{aligned} & \mathbb{P}\left(X_{t_1} \in B_1, X_{t_2} \in B_2, \ldots X_{t_n} \in B_n\right) \quad (2.1) \\ = \quad & \mathbb{P}\left(X_{t_1} \in B_1, X_{t_2} \in B_2, \ldots X_{t_n} \in B_n, X_{t_{n+1}} \in \Xi, X_{t_{n+2}} \in \Xi, \ldots X_{t_m} \in \Xi\right) \end{aligned}$$

This is going to get really awkward to write over and over, so let's introduce some simplifying notation. $\mathrm{Fin}(T)$ will denote the class of all finite sub-sets of our index set T, and likewise $\mathrm{Denum}(T)$ all denumerable sub-sets. We'll indicate such sub-sets, for the moment, by capital letters like J, K, etc., and extend the definition of coordinate maps (Definition 15) so that π_J maps from Ξ^T to Ξ^J in the obvious way, and π_J^K maps from Ξ^K to Ξ^J, if $J \subset K$. If μ is the measure for the whole process, then the finite-dimensional distributions are $\{\mu_J | J \in \mathrm{Fin}(T)\}$. Clearly, $\mu_J = \mu \circ {\pi_J}^{-1}$.

Definition 24 (Projective Family of Distributions) *A family of distributions μ_J, $J \in \mathrm{Denum}(T)$, is* projective *when for every $J, K \in \mathrm{Denum}(T)$, $J \subset K$ implies*

$$\mu_J = \mu_K \circ \left(\pi_J^K\right)^{-1} \quad (2.2)$$

Such a family is also said to be consistent *or* compatible *(with one another).*

Lemma 25 (FDDs Form Projective Families) *The finite-dimensional distributions of a stochastic process always form a projective family.*

PROOF: This is just the fact that we get marginal distributions by integrating out some variables from the joint distribution. But, to proceed formally: Letting J and K be finite sets of indices, $J \subset K$, we know that $\mu_K = \mu \circ {\pi_K}^{-1}$, that $\mu_J = \mu \circ {\pi_J}^{-1}$ and that $\pi_J = \pi_J^K \circ \pi_K$. Hence

$$\begin{aligned} \mu_J &= \mu \circ \left(\pi_J^K \circ \pi_K\right)^{-1} && (2.3) \\ &= \mu \circ \pi_K^{-1} \circ \left(\pi_J^K\right)^{-1} && (2.4) \\ &= \mu_K \circ \left(\pi_J^K\right)^{-1} && (2.5) \end{aligned}$$

as required. □

I claimed that the reason to care about finite-dimensional distributions is that if we specify them, we specify the distribution of the whole process. Lemma 25 says that a putative family of finite dimensional distributions must be consistent, if they are to let us specify a stochastic process. Theorem 23 says that there can't be more than one process distribution with all the same finite-dimensional marginals, but it doesn't guarantee that a given collection of consistent finite-dimensional distributions *can* be extended to a process distribution — it gives uniqueness but not existence. Proving the existence of an extension requires some extra assumptions. Either we need to impose topological conditions on Ξ, or we need to ensure that all the finite-dimensional distributions can be related through conditional probabilities. The first approach is due to Daniell and Kolmogorov, and will finish this lecture; the second is due to Ionescu-Tulcea, and will begin the next.

We'll start with Daniell's theorem on the existence of random sequences, i.e., where the index set is the natural numbers, which uses mathematical induction to extend the finite-dimensional family. To get *there*, we need a useful proposition about our ability to represent non-trivial random variables as functions of uniform random variables on the unit interval.

Proposition 26 (Randomization, transfer) *Let X and X' be identically-distributed random variables in a measurable space Ξ and Y a random variable in a Borel space Υ. Then there exists a measurable function $f : \Xi \times [0,1] \mapsto \Upsilon$ such that $\mathcal{L}(X', f(X', Z)) = \mathcal{L}(X, Y)$, when Z is uniformly distributed on the unit interval and independent of X'.*

Proof: See Kallenberg, Theorem 6.10 (p. 112–113). □

Basically what this says is that if we have two random variables with a certain joint distribution, we can always represent the pair by a copy of one of the variables (X), and a transformation of an independent random number. It is important that Υ be a Borel space here; the result, while very natural-sounding, does not hold for arbitrary measurable spaces, because the proof relies on having a regular conditional probability.

Theorem 27 (Daniell Extension Theorem) *For each $n \in \mathbb{N}$, let Ξ_n be a Borel space, and μ_n be a probability measure on $\prod_{i=1}^{n} \Xi_i$. If the μ_n form a projective family, then there exist random variables $X_i : \Omega \mapsto \Xi_i$, $i \in \mathbb{N}$, such that $\mathcal{L}(X_1, X_2, \ldots X_n) = \mu_n$ for all n, and a measure μ on $\prod_{i=1}^{\infty} \Xi_i$ such that μ_n is equal to the projection of μ onto $\prod i = 1^n \Xi_i$.*

Proof: For any fixed n, $X_1, X_2, \ldots X_n$ is just a random vector with distribution μ_n, and we can always construct such an object. The delicate part here is showing that, when we go to $n+1$, we can use the *same* random elements for the first n coordinates. We'll do this by using the representation-by-randomization proposition just introduced, starting with an IID sequence of uniform random variables on the unit interval, and then transforming them to get a sequence of variables in the Ξ_i which have the right joint distribution. (This is like the quantile transform trick for generating random variates.) The proof will go inductively, so first we'll take care of the induction step, and then go back to reassure ourselves about the starting point.

Induction: Assume we already have $X_1, X_2, \ldots X_n$ such that $\mathcal{L}(X_1, X_2, \ldots X_n) = \mu_n$, and that we have a $Z_{n+1} \sim U(0,1)$ and independent of all the X_i to date. As remarked, we can always get $Y_1, Y_2, \ldots Y_{n+1}$ such that $\mathcal{L}(Y_1, Y_2, \ldots Y_{n+1}) = \mu_{n+1}$. Because the μ_n form a projective family, $\mathcal{L}(Y_1, Y_2, \ldots Y_n) = \mathcal{L}(X_1, X_2, \ldots X_n)$. Hence, by Proposition 26, there is a measurable f such that, if we set $X_{n+1} = f(X_1, X_2, \ldots X_n, Z_{n+1})$, then $\mathcal{L}(X_1, X_2, \ldots X_n, X_{n+1}) = \mu_{n+1}$.

First step: We need there to be an X_1 with distribution μ_1, and we need a (countably!) unlimited supply of IID variables $Z_2, Z_3, \ldots$ all $\sim U(0,1)$. But the existence of X_1 is just the existence of a random variable with a well-defined distribution, which is unproblematic, and the existence of an infinite sequence of IID uniform random variates is too. (See 36-752, or Lemma 3.21 in Kallenberg.)

Finally, to convince yourself of the existence of the measure μ on the product space, recall Theoerem 16. □

Remark: Kallenberg, Corollary 6.15, gives a somewhat more abstract version of this theorem.

Daniell's extension theorem works fine for one-sided random sequences, but we often want to work with larger and more interesting index sets. For this we need the full Kolmogorov extension theorem, where the index set T can be completely arbitrary. This in turn needs the Carathéodory Extension Theorem, which I re-state here for convenience.

Proposition 28 (Carathéodory Extension Theorem) *Let μ be a non-negative, finitely additive set function on a field $\mathcal{C}$ of subsets of some space Ω. If μ is also countably additive, then it extends to a measure on $\sigma(\mathcal{C})$, and, if $\mu(\Omega) < \infty$, the extension is unique.*

PROOF: See 36-752 lecture notes (Theorem 50, Exercise 51), or Kallenberg, Theorem 2.5, pp. 26–27. Note that "extension" here means extending from a mere field to a σ-field, not from finite to infinite index sets. □

Theorem 29 (Kolmogorov Extension Theorem) *Let Ξ_t, $t \in T$, be a collection of Borel spaces, with σ-fields $\mathcal{X}_i$, and let μ_J, $J \in \mathrm{Fin}(T)$, be a projective family of finite-dimensional distributions on those spaces. Then there exist Ξ_t-valued random variables X_t such that $\mathcal{L}(X_J) = \mu_J$ for all $J \in \mathrm{Fin}(T)$.*

PROOF: This will be easier to follow if we first consider the case there T is countable, which is basically Theorem 27 again, and then the general case, where we need Proposition 28.

Countable T: We can, by definition, put the elements of T in $1-1$ correspondence with the elements of $\mathbb{N}$. This in turn establishes a bijection between the product space $\bigotimes_{t \in T} \Xi_t = \Xi_T$ and the sequence space $\bigotimes_{i=1}^{\infty} \Xi_t$. This bijection also induces a projective family of distributions on finite sequences. The Daniell Extension Theorem (27) gives us a measure on the sequence space, which the bijection takes back to a measure on Ξ_T. To see that this μ does not depend on the order in which we arranged T, notice that any two arrangements must give identical results for any finite set J, and then use Theorem 23.

Uncountable T: For each countable $K \subset T$, the argument of the preceding paragraph gives us a measure μ_K on Ξ_K. And, clearly, these μ_K themselves form a projective family. Now let's define a set function μ on the countable cylinder sets, i.e., on the class $\mathcal{D}$ of sets of the form $A \times \Xi_{T \setminus K}$, for some $K \in \mathrm{Denum}(T)$ and some $A \in \mathcal{X}_K$. Specifically, $\mu : \mathcal{D} \mapsto [0,1]$, and $\mu(A \times \Xi_{T \setminus K}) = \mu_K(A)$. We would like to use Carathéodory's theorem to extend this set function to a measure on the product σ-algebra $\mathcal{X}_T$. First, let's check that the countable cylinder sets form a field: (i) $\Xi_T \in \mathcal{D}$, clearly. (ii) The complement, in Ξ_T, of a countable cylinder $A \times \Xi_{T\setminus K}$ is another countable cylinder, $A^c \times \Xi_{T\setminus K}$. (iii) The union of two countable cylinders $B_1 = A_1 \times \Xi_{T\setminus K_1}$ and $B_2 = A_2 \times \Xi_{T\setminus K_2}$ is another countable cylinder, since we can always write it as $A \times \Xi_{T\setminus K}$ for some $A \in \mathcal{X}_K$, where $K = K_1 \cup K_2$. Clearly, $\mu(\emptyset) = 0$, so we just need to check that μ is countably additive. So consider any sequence of disjoint cylinder sets $B_1, B_2, \ldots$. Because they're cylinder sets, each i, $B_i = A_i \times \Xi_{T\setminus K_i}$, for some $K_i \in \mathrm{Denum}(T)$, and some $A_i \in \mathcal{X}_{K_i}$. Now set $K = \bigcup_i K_i$; this is a countable union of countable sets, and so itself countable. Furthermore, say $C_i = A_i \times \Xi_{K\setminus K_i}$, so we can say that $\bigcup_i B_i = (\bigcup_i C_i) \times \Xi_{T\setminus K}$. With this notation in place,

$$\mu \bigcup_i B_i = \mu_K \bigcup_i C_i \tag{2.6}$$

$$= \sum_i \mu_K C_i \tag{2.7}$$

$$= \sum_i \mu_{K_i} A_i \tag{2.8}$$

$$= \sum_i \mu B_i \tag{2.9}$$

where in the second line we've used the fact that μ_K is a probability measure on Ξ_K, and so countably additive on sets like the C_i. This proves that μ is countably additive, so by Proposition 28 it extends to a measure on $\sigma(\mathcal{D})$, the σ-field generated by the countable cylinder sets. But we know from Definition 12 that this σ-field *is* the product σ-field. Since $\mu(\Xi_T) = 1$, Proposition 28 further tells us that the extension is unique. □

Borel spaces are good enough for most of the situations we find ourselves modeling, so the Daniell-Kolmogorov Extension Theorem (as it's often known) see a lot of work. Still, some people dislike having to make topological assumptions to solve probabilistic problems; it seems inelegant. The Ionescu-Tulcea Extension Theorem provides a purely probabilistic solution, available if we can write down the FDDs recursively, in terms of regular conditional probability distributions, even if the spaces where the process has its coordinates are not nice and Borel. Doing this properly will involve our revisiting and extending some ideas about conditional probability, which you will have seen in 36-752, so it will be deferred to the next lecture.