17. Measures on Infinite Product Spaces

We will now move to the concept of infinite products of measurable spaces and measures on these product spaces. The constructions and results we will present have important applications in probability and ergodic theory.

17.1 Infinite Product Spaces.

We start with a remark on certain set-theoretical concepts used below. Given any collection of sets \( \{ X_\alpha : \alpha \in I \} \) indexed by a set \( I \), we write \( \prod_{\alpha \in I} X_\alpha \) to denote their Cartesian product. For finite \( I \), or even \( I \) having just two elements, the Cartesian product is defined using the Axioms of Pairing and Union in Zermelo-Frankel axiomatics. This will not work for \( I \) infinite, where we define \( \prod_{\alpha \in I} X_\alpha \) as the set of functions, \( \bigotimes_{\alpha \in I} X_\alpha \) as

\[
\bigotimes_{\alpha \in I} X_\alpha := \left\{ f : I \rightarrow \bigcup_{\alpha \in I} X_\alpha : \left( \forall \alpha \in I : f(\alpha) \in X_\alpha \right) \right\}.
\]

(17.1)

(The reader will recall that the concept of a function, which is a special case of a relation, is still based on the concept of a Cartesian product of two sets.) Unlike finite Cartesian products, there is no guarantee that the Cartesian product is non-empty, i.e., that

\[
\prod_{\alpha \in I} X_\alpha \neq \emptyset.
\]

(17.2)

Indeed, we need to invoke the Axiom of Choice for this to be true. (One case where this is not required when \( X_\alpha = X \) for all \( \alpha \in I \) where we write \( X^I := \prod_{\alpha \in I} X \).)

A starting point of our considerations is:

**Definition 17.1 (Infinite Product Space)** Let \( \{(X_\alpha, \mathcal{F}_\alpha) : \alpha \in I \} \) be an arbitrary family of measurable spaces. The product measurable space is then the pair \((\prod_{\alpha \in I} X_\alpha, \bigotimes_{\alpha \in I} \mathcal{F}_\alpha)\)

where

\[
\bigotimes_{\alpha \in I} \mathcal{F}_\alpha := \sigma(S)
\]

(17.3)

for

\[
S := \left\{ \bigotimes_{\alpha \in I} A_\alpha : \left( \forall \alpha \in I : A_\alpha \in \mathcal{F}_\alpha \right) \land \{ \alpha \in I : A_\alpha \neq X_\alpha \} \text{ finite} \right\}
\]

(17.4)

is the product \(\sigma\)-algebra on \( \prod_{\alpha \in I} X_\alpha \). If \( X_\alpha = X \) for all \( \alpha \in I \), we use \( X^\otimes := \bigotimes_{\alpha \in I} X \).

The restriction that \( \{ \alpha \in I : A_\alpha \neq X_\alpha \} \) be finite is imposed for several reasons. First off, it ensures that \( S \) is a semi-algebra, which is convenient for construction of measures. Second, the restriction is consistent with the construction of product topology, which is relevant when \( X_\alpha \) is a topological space and \( \mathcal{F}_\alpha \) are the Borel sets. However, the restriction also comes with less desirable consequences:

**Lemma 17.2 (Countability Curse)** Whenever \( I \) is infinite, the word “finite” in (17.4) can be replaced by “countable” with no change of \( \sigma(S) \). Moreover, setting

\[
S_J := \left\{ \bigotimes_{\alpha \in I} A_\alpha : \left( \forall \alpha \in I : A_\alpha \in \mathcal{F}_\alpha \right) \land \{ \alpha \in I : A_\alpha \neq X_\alpha \} \subseteq J \right\}
\]

(17.5)
we then have
\[ \bigotimes_{a \in I} \mathcal{F}_a = \bigcup_{\text{countable} \sigma(\mathcal{G})}. \]  
(17.6)

In particular, every \( A \in \bigotimes_{a \in I} \mathcal{F}_a \) is non-trivial in at most a countable number of coordinates.

We leave the proof to a homework exercise. To explain why we label this using the phrase “countability curse,” let us consider the example of \( \mathbb{R} \), viewed as the space of all real-valued functions of real variables. We endow \( \mathbb{R} \) with the product Borel \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R})^{<\mathbb{R}} \) which corresponds to the Borel sets in the topology of pointwise convergence. Then, while
\[ \{ f \in \mathbb{R}^\mathbb{R} : f \text{ is uniformly continuous on } \mathbb{Q} \} \in \mathcal{B}(\mathbb{R})^{<\mathbb{R}} \]  
(17.7)
because the set involves “looking” at \( f \) at only a countably many coordinates, we have
\[ \{ f \in \mathbb{R}^\mathbb{R} : f \text{ is continuous everywhere} \} \notin \mathcal{B}(\mathbb{R})^{<\mathbb{R}} \]  
(17.8)
because continuity everywhere cannot be checked from a dense subset only. The moral is that, for uncountable products, the \( \sigma \)-algebra is often too small to allow asking natural regularity questions about the objects in the product space. This is of relevance in the study of random functions — called stochastic processes — in probability.

In order to overcome the “countability curse” one needs to assume some regularity of the underlying functions from the outset or introduce so-called modifications. For instance, supposing we can somehow prove that, under a measure \( \mu \) on \( (\mathbb{R}^\mathbb{R}, \mathcal{B}(\mathbb{R})^{<\mathbb{R}}) \), we have
\[ \mu\left( \{ f \in \mathbb{R}^\mathbb{R} : f \text{ is uniformly continuous on } \mathbb{Q} \}^c \right) = 0. \]  
(17.9)
Then (except on this null set where we set \( \bar{f} := 0 \)) we can define
\[ \bar{f}(x) := \lim_{q \in \mathbb{Q} : q \uparrow x} f(x) \]  
(17.10)
to get a continuous modification of \( f \) that agrees with \( f \) at all rational points. In order to make \( \bar{f} \) agree with \( f \) at most other points, it then suffices to assume that \( x \mapsto f(x) \) is continuous in measure. This strategy underlies the proof of:

**Theorem 17.3 (Kolmogorov-Čenstov)** Assume \( \mu \) is a probability measure on \( (\mathbb{R}^\mathbb{R}, \mathcal{B}(\mathbb{R})^{<\mathbb{R}}) \) such that, for some \( a, \beta > 0 \) and \( c \in (0, \infty) \),
\[ \forall s < t : \int |f(t) - f(s)|^a \mu(\mathbf{d}f) \leq c |t - s|^{1+\beta}. \]  
(17.11)
For any \( \gamma \in (0, \beta/a) \) and writing \( \mathcal{D} := \{ k2^{-n} : k \in \mathbb{Z} \wedge n \in \mathbb{N} \} \) for the dyadic rationals in \( \mathbb{R} \), we then have
\[ \forall N \geq 1 : \mu\left( \left\{ f \in \mathbb{R}^\mathbb{R} : \sup_{s,t \in \mathcal{D}, -N \leq s < t \leq N} \frac{|f(t) - f(s)|}{|t - s|^{\gamma}} < \infty \right\} \right) = 1. \]  
(17.12)
Denoting the set in (17.12) by $A$,

$$
\bar{f}(x) := \begin{cases} 
\lim_{q \to D, q \to x} f(x), & \text{on } A, \\
0, & \text{on } A^c,
\end{cases}
$$  \hspace{1cm} (17.13)

is well defined and locally $\gamma$-H"older continuous. Moreover, $f \mapsto \bar{f}(x)$ is Borel measurable for each $x \in \mathbb{R}$ and

$$
\forall x \in \mathbb{R}: \quad \mu(\{ \bar{f}(x) = f(x) \}) = 1. \hspace{1cm} (17.14)
$$

We leave the proof of Kolmogorov-Čenstov’s Theorem to (an annotated) homework exercise. In probability, this theorem (or its natural extensions) is a great tool for control of regularity of an uncountable family of random variables using a suitably chosen countable sub-collection thereof. (Other, somewhat more refined methods exist as well, e.g., generic chaining, but they are all based on similar ideas.)

A natural question to address next is the construction of measures on the infinite product space. We will start with product measures as these are those most easily described by their value of the measure “rectangles” — i.e., the sets in $S$ above. Unfortunately, since rectangles now have infinitely many “sides” we have to restrict attention to probability measures from the outset.

**Theorem 17.4** (Infinite product measure)  Let $\{(X_\alpha, \mathcal{F}_\alpha, \mu_\alpha): \alpha \in I \}$ be measure spaces with

$$
\forall \alpha \in I: \quad \mu_\alpha(X_\alpha) = 1. \hspace{1cm} (17.15)
$$

Then there exists a unique probability measure $\bar{\mu}$ on $\bigotimes_{\alpha \in I} X_\alpha$ such that

$$
\forall A = \bigotimes_{\alpha \in I} A_\alpha \in S: \quad \bar{\mu}(A) = \prod_{\alpha \in I} \mu_\alpha(A_\alpha) \hspace{1cm} (17.16)
$$

**Proof.** The class of sets $S$ is readily checked to be a semi-algebra, so defining $\bar{\mu}$ on $S$ by (17.16) we just need to check that $\bar{\mu}$ is finitely additive and countably subadditive on $S$. Lemma 11.8 shows that

$$
A := \left\{ \bigcap_{j=1}^n E_j: n \geq 1 \land E_1, \ldots, E_n \in S \text{ disjoint} \right\} \hspace{1cm} (17.17)
$$

is an algebra and Proposition 11.9 then shows that $\bar{\mu}$ is finitely (and, by Theorem 16.3, even countably) additive on $A$, and thus also on $S$. As in the case of finite product spaces, we will aim to prove directly countable additivity of $\bar{\mu}$ on $S$. We will build on the argument (16.12–16.13) based on integration; however, attention needs to be paid to the fact that, in the absence of an extension of $\bar{\mu}$ to the full product space, we can integrate at most a finite number of variables at a time.

Let $\{A_n\}_{n \geq 1} \subseteq S$ be disjoint with $A := \bigcup_{n \geq 1} A_n \in S$. Since

$$
A = A_1 \cup \cdots \cup A_{n-1} \cup \bigcup_{i \geq n} A_i \hspace{1cm} (17.18)
$$
the disjointness shows
\[ B_n := \bigcup_{i \geq n} A_i = A \cap A_1^c \cap \ldots \cap A_{n-1}^c \]  
\[(17.19)\]
Hence \( B_n \in \mathcal{A} \) and \( \tilde{\mu}(B_n) \) is defined. Moreover, finite additivity implies
\[ 0 \leq \tilde{\mu}(A) - \sum_{n \geq 1} \tilde{\mu}(A_i) = \lim_{n \to \infty} \tilde{\mu}(B_n). \]  
\[(17.20)\]
We thus need:

**Lemma 17.5** Let \( \tilde{\mu} \) be a finitely additive set function on \( \mathcal{A} \) defined from \( \mathcal{S} \) via (17.17). Then for any \( \{B_n\}_{n \geq 1} \subseteq \mathcal{A} \) satisfying
\[ \left( \forall n \geq 1: B_{n+1} \subseteq B_n \right) \land \bigcap_{n \geq 1} B_n = \emptyset \]  
\[(17.21)\]
we have
\[ \lim_{n \to \infty} \tilde{\mu}(B_n) = 0. \]  
\[(17.22)\]

**Proof.** By the fact that each set in \( \mathcal{S} \) depends on a finite number of indices, for each \( n \geq 1 \) there is a finite \( I_n \subseteq I \) such that \( B_n \) is non-trivial only for \( a \in I_n \). We may assume that \( I_n \subseteq I_{n+1} \) for all \( n \geq 1 \) and denote \( I_\infty := \bigcup_{n \geq 1} I_n \) and \( I_0 := \emptyset \). For \( m < n \), abbreviate
\[ X_{m,n} := \bigotimes_{a \in I_m \setminus I_n} X_a \quad \text{and} \quad F_{m,n} := \bigotimes_{a \in I_m \setminus I_n} F_a \]  
\[(17.23)\]
and let
\[ \mu_{m,n} := \bigotimes_{a \in I_m \setminus I_n} \mu_a \]  
\[(17.24)\]
be the product measure on \( (X_{m,n}, F_{m,n}) \), which exists by Theorem 16.3.

The indicator \( 1_{B_n} \) depends on the variables of \( \{x_a\}_{a \in I_n} \). Define \( h_{m,n} : X_{0,m} \to [0,1] \) by
\[ h_{m,n}((x_a)_{a \in I_n}) := \begin{cases} 1_{B_n}((x_a)_{a \in I_n}), & \text{if } n < m, \\ \int_{X_{0,n}} 1_{B_n}((x_a)_{a \in I_n}) \, d\mu_{m,n} & \text{if } m \leq n, \end{cases} \]  
\[(17.25)\]
where the integral is over the variables indexed by \( I_n \setminus I_m \). We now observe three facts. First, in light of the product structure of \( \mu_{m,n} \), Tonelli’s Theorem implies
\[ \forall m, n \geq 1: \quad h_{m,n} = \int_{X_{0,m+1}} h_{m+1,n} \, d\mu_{m,m+1}, \]  
\[(17.26)\]
Second, the monotonicity of \( \{B_n\}_{n \geq 1} \) implies \( h_{m,n+1} \leq h_{m,n} \) and so
\[ h_m := \lim_{n \to \infty} h_{m,n} \]  
\[(17.27)\]
exists for all \( m \geq 1 \). Third, by the defining property of \( \tilde{\mu} \),
\[ \tilde{\mu}(B_n) = \int_{X_{0,n}} h_{m,n} \, d\mu_{0,m} \]  
\[(17.28)\]
for all \( m, n \geq 1 \).
We now use these properties to prove the above claim. First, as \(0 \leq h_{n,m} \leq 1\) and \(\mu_{m,n}\) is a probability measure, the Bounded Convergence Theorem implies

\[
h_m = \int_{X_{m,m+1}} h_{m+1} \, d\mu_{m,m+1}
\]

and

\[
\lim_{n \to \infty} \bar{\mu}(B_n) = \int_{X_{0,m}} h_m \, d\mu_{0,m}
\]

for all \(m \geq 1\). If \(\delta := \lim_{n \to \infty} \bar{\mu}(B_n) > 0\), then (17.30) shows the existence of \(\bar{x}_1 := \{x_a\}_{a \in I_1}\) such that \(h_1(\bar{x}_1) \geq \delta\). By (17.29), for each \(m \geq 1\) this \(\bar{x}_1\) can be extended to \(\bar{x}_m := \{x_a\}_{a \in I_m}\) such that \(h_m(\bar{x}_m) \geq \delta\). This defines \(\bar{x}_\infty := \{x_a\}_{a \in I_\infty}\) such that \(h_m(\bar{x}_\infty) \geq \delta\) for all \(m \geq 1\) and thus also \(h_{m,n}(\bar{x}_\infty) \geq h_m(\bar{x}_\infty) \geq \delta\) for all \(m, n \geq 1\). Taking \(m > n\) implies that \(\bar{x}_\infty \in \bigcap_{n \geq 1} B_n\) and so the intersection is non-empty. \(\Box\)

Returning to the proof of Theorem 17.4, Lemma 17.5 implies by \(\bar{\mu}\) is countably additive on \(\mathcal{S}\) and so, by Corollary 11.10, it extends to a unique (probability) measure on \(\sigma(\mathcal{S}) = \bigotimes_{a \in I} \mathcal{F}_a\). \(\Box\)

While we used quite strongly that \(\bar{\mu}\) can, with the help of Theorem 16.3, be assumed finitely and even countably additive on \(\mathcal{A}\), this alone is not sufficient for an extension to exist. Indeed, the above argument uses the underlying product structure in a number of places and, without this structure, the extension generally requires additional assumptions on the underlying spaces. This is the content of the Kolmogorov Extension Theorem to be discussed later.

On the other hand, if \(I\) is countable and \(X_a = \mathbb{R}\), which is often the case for applications in probability, the above abstract reasoning can be by-passed completely:

**Lemma 17.6** For any sequence of Radon measures \(\{\mu_n\}_{n \geq 1}\) on \((\mathbb{R}, B(\mathbb{R}))\), there exists an explicit measurable map \(f : [0, 1] \to \mathbb{R}^N\) such that

\[
\bigotimes_{n \geq 1} \mu_n = \lambda \circ f^{-1}
\]

where \(\lambda\) is the Lebesgue measure on \(([0, 1], B([0, 1]))\).

Although \((\mathbb{R}^N, B(\mathbb{R}^N))\) is a standard Borel space, and is thus Borel isomorphic to \(([0, 1], B([0, 1]))\), the main point of this lemma is to have a map that pushes the Lebesgue measure into the prescribed product measure on \((\mathbb{R}^N, B(\mathbb{R}^N))\). A small technicality is also the proof of \(B(\mathbb{R}^N) = B(\mathbb{R})^{\otimes N}\). We leave this lemma to a homework exercise.

### 17.2 Zero-one laws.

Having constructed the product measures, we move to note that these measures tend to be trivial on certain natural \(\sigma\)-algebras. Such statements go under the banner zero-one laws. The best known example of these is due to Kolmogorov. Consider the product space \((\bigotimes_{a \in I} X_a, \bigotimes_{a \in I} \mathcal{F}_a)\) and let \(S_j\) as in (17.6). For any \(j \leq I\) (not necessarily countable),
denote by
\[ \mathcal{F}_I := \sigma\left( \bigcup_{K \subseteq I \text{ finite}} \mathcal{S}_K \right) \] (17.32)
the \( \sigma \)-algebra (and sub-\( \sigma \)-algebra of \( \mathcal{F}_I = \bigotimes_{a \in I} \mathcal{F}_a \)) of sets that depend only on the coordinates in \( I \). We then put forward:

**Definition 17.7** (Tail sets and \( \sigma \)-algebra) A set \( A \in \bigotimes_{a \in I} \mathcal{F}_a \) is a tail set if it belongs to the tail \( \sigma \)-algebra
\[ \mathcal{T} := \bigcap_{J \subseteq I \text{ finite}} \mathcal{F}_J \setminus I \] (17.33)
on \( \bigotimes_{a \in I} X_a, \bigotimes_{a \in I} \mathcal{F}_a \).

The tail \( \sigma \)-algebra contains those sets that, in a matter of speech, “do not depend” on any finite number of coordinates. Although this may seem to disqualify all but trivial sets, there are in fact many sets in \( \mathcal{T} \) whenever \( I \) is infinite. For instance, for \( I := \mathbb{N} \) and \( X_a := \{0,1\} \) for all \( a \in \mathbb{N} \),
\[ \left\{ \sigma \in \{0,1\}^\mathbb{N} : \sum_{n \geq 1} \sigma_n = \infty \right\} \in \mathcal{T}. \] (17.34)
This is true because changing any finite number of the coordinates \( \sigma \) will not affect convergence or divergence of \( \sum_{n \geq 1} \sigma_n \). Other examples of tail sets are
\[ \{\sigma_n = 1 \text{ i.o.}\} \quad \text{or} \quad \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i \text{ exists} \right\}. \] (17.35)

For the tail \( \sigma \)-algebra under product measures we then get:

**Theorem 17.8** (Kolmogorov’s zero-one law) Every product probability measure \( \mu \) on the product space \( \bigotimes_{a \in I} X_a, \bigotimes_{a \in I} \mathcal{F}_a \) is trivial on \( \mathcal{T} \) in the sense
\[ \forall E \in \mathcal{T} : \quad \mu(E) \in \{0,1\}. \] (17.36)

**Proof.** We will show that, in fact,
\[ \forall A \in \mathcal{F}_I, \forall B \in \mathcal{T} : \quad \mu(A \cap B) = \mu(A) \mu(B) \] (17.37)
which in the language of probability means that \( \mathcal{T} \) is independent of the \( \sigma \)-algebra \( \mathcal{F}_I \).
This does imply the claim because \( \mathcal{T} \subseteq \mathcal{F}_I \) and so taking \( A = B := E \in \mathcal{T} \) we get \( \mu(E) = \mu(E)^2 \), proving (17.36). (The converse, namely (17.36) \( \Rightarrow \) (17.37), is also true.)

For any \( B \in \mathcal{F}_I \), define
\[ \mathcal{L}_B := \{ A \in \mathcal{F}_I : \mu(A \cap B) = \mu(A) \mu(B) \}. \] (17.38)
Using the \( \sigma \)-additivity and finiteness of \( \mu \) we readily check that \( \mathcal{L}_B \) is a \( \lambda \)-system. Moreover, the symmetry of the condition defining \( \mathcal{L}_B \) shows
\[ \forall A, B \in \mathcal{F}_I : \quad A \in \mathcal{L}_B \quad \Leftrightarrow \quad B \in \mathcal{L}_A. \] (17.39)
Next observe that, thanks to the product structure of $\mu$, for all $J, K \subseteq I$ finite
\[ J \cap K = \emptyset \implies \forall A \in S_J: \ S_K \subseteq \mathcal{L}_A \quad (17.40) \]
Since the union of $S_K$ over all finite $K \subseteq I \setminus J$ is a $\pi$-system, Dynkin’s $\pi/\lambda$-Theorem implies
\[ \forall J \subseteq I \text{ finite } \forall A \in S_J: \ F_{I \setminus J} \subseteq \mathcal{L}_A. \quad (17.41) \]
Since $\mathcal{T} \subseteq F_{I \setminus J}$ for all finite $J \subseteq I$, this implies
\[ \forall A \in \bigcup_{J \subseteq I \text{ finite}} S_J: \ \mathcal{T} \subseteq \mathcal{L}_A. \quad (17.42) \]
Then (17.39) turns this into
\[ \forall B \in \mathcal{T}: \ \bigcup_{J \subseteq I \text{ finite}} S_J \subseteq \mathcal{L}_B \quad (17.43) \]
and since the giant union is a $\pi$-system, Dynkin’s $\pi/\lambda$-Theorem shows that $F_I \subseteq \mathcal{L}_B$ for each $B \in \mathcal{T}$. This gives (17.37) as desired. \hfill \Box

Note that the above proof does not make a distinction between the case when $I$ is finite or infinite, even though for $I$ finite we always have $\mathcal{T} = \{\emptyset, X_\alpha \times_{\alpha \in I} X_\beta\}$ and so the measure is trivially trivial. Note also that in the definition of the tail $\sigma$-algebra we cannot replace “finite” by “countable.”

There are other zero-one laws that apply to product measures, albeit only under some uniformity restrictions on the underlying space. We will focus on the concept of exchangeability. We again start with:

**Definition 17.9** (Finite permutations) Given a non-empty set $I$, a map $\varphi: I \to I$ is a finite permutation if $\varphi$ is bijective and
\[ \{\alpha \in I: \varphi(\alpha) \neq \alpha\} \text{ is finite.} \quad (17.44) \]

**Definition 17.10** (Exchangeable sets) Given non-empty sets $X$ and $I$, a set $A \subseteq X^I$ is said to be exchangeable if for all finite permutations $\varphi$, with $h_\varphi: X^I \to X^I$ defined by
\[ h_\varphi(\{x_\alpha\}_{\alpha \in I}) := \{x_{\varphi(\alpha)}\}_{\alpha \in I}, \quad (17.45) \]
we have $h_\varphi^{-1}(A) = A$.

Exchangeable sets are thus those that are invariant under any “shuffle” of any finite number of coordinates. We observe:

**Lemma 17.11** Consider the product spaces $(X^I, \mathcal{F}^I)$. For any finite permutation $\varphi$ on $I$, the map $h_\varphi$ is measurable. In particular,
\[ \mathcal{E} := \{A \in \mathcal{F}^I: \text{exchangeable}\} \quad (17.46) \]
is a $\sigma$-algebra with $\mathcal{T} \subseteq \mathcal{E}$.

**Proof.** For measurability we note that $h_\varphi^{-1}(S) = S$; the rest follows by the fact that the preimage map commutes around countable unions and complements. This also shows that $\mathcal{E}_\varphi := \{A \in \mathcal{F}^I: h_\varphi^{-1}(A) = A\}$ is a $\sigma$-algebra. As $\mathcal{E}$ is the intersection of $\mathcal{E}_\varphi$ over all
finite permutations \( \varphi \), we get \( \mathcal{E} \) is a \( \sigma \)-algebra. Since tail sets do not depend on any finite set of coordinates, they are invariant under finite permutations.

Next we introduce:

**Definition 17.12 (Exchangeable measures)** A measure \( \mu \) on \((X^I, \mathcal{F}^I)\) is said to be exchangeable if \( \mu = \mu \circ h_{\varphi}^{-1} \) for any finite permutation \( \varphi \) of \( I \).

We note:

**Lemma 17.13** Let \( \mu \) be a probability measure on \((X, \mathcal{F})\). Then the product measure \( \mu \otimes I \) on \((X^I, \mathcal{F}^I)\) is exchangeable.

**Proof.** Let \( S \) be as in (17.4) and, given a finite permutation, let

\[
\mathcal{L} := \{ A \in \mathcal{F}^I : \mu \otimes I (A) = \mu \otimes I (h_{\varphi}^{-1}(A)) \}.
\]

(17.47)

Then \( \mathcal{L} \) is again readily checked to be a \( \lambda \)-system. Since the defining property (17.16) of product measures ensures \( S \subseteq \mathcal{L} \), and \( S \) is a \( \pi \)-system, Dynkin’s \( \pi / \lambda \)-Theorem gives \( \mathcal{F}^I = \mathcal{L} \). Hence, \( \mu = \mu \circ h_{\varphi}^{-1} \) for all finite permutations \( \varphi \), as claimed.

We now give:

**Theorem 17.14 (Hewitt-Savage zero-one law)** Let \( \mu \) be a probability measure on \((X, \mathcal{F})\) and let \( \mu \otimes I \) be the associated product measure on \((X^I, \mathcal{F}^I)\). If \( I \) is infinite, then \( \mu \otimes I \) is trivial on \( \mathcal{E} \), meaning

\[
\forall E \in \mathcal{E} : \mu \otimes I (A) \in \{0, 1\}. \]

(17.48)

**Proof.** Since \( \mu \otimes I \) is a probability measure, the claim will follow once we show

\[
\forall E \in \mathcal{E} : \mu \otimes I (E) \leq \mu \otimes I (E)^2. \]

(17.49)

As \( \mathcal{E} \subseteq \mathcal{F}^I \) and \( \mathcal{F}^I \) is generated by the procedure underlying Proposition 11.9 as intersections of countable unions of sets in the semi-algebra (17.4), for each \( n \geq 1 \) there is a set \( E_n \) depending only on coordinates in a finite set \( I_n \subseteq I \) such that \( \mu \otimes I (E_n \Delta E) < 1/n \). Since \( I \) is infinite, there exists \( J_n \subseteq I \setminus I_n \) such that \( J_n \) has the same cardinality of \( I_n \). Let \( \varphi : I \to I \) be a bijection such that \( \varphi(\alpha) = \alpha \) for \( \alpha \notin I_n \cup J_n \) and such that \( \varphi \) maps \( I_n \) bijectively onto \( J_n \). Define

\[
F_n := h_{\varphi}^{-1}(E_n). \]

(17.50)

Then \( E_n \) depends on different coordinates of \( F_n \) and so, by the product structure of \( \mu \otimes I \),

\[
\mu \otimes I (E_n \cap F_n) = \mu \otimes I (E_n) \mu \otimes I (F_n). \]

(17.51)

But \( h_{\varphi}^{-1}(E) = E \) and so \( \mu \otimes I (E \cap F_n) = \mu \otimes I (E \cap E_n) < 1/n. \) A computation now gives

\[
[\mu \otimes I (E) + 1/n] \geq \mu \otimes I (E_n) \mu \otimes I (F_n) = \mu \otimes I (E_n \cap F_n) \geq \mu \otimes I (E) - 2/n. \]

(17.52)

Taking \( n \to \infty \) we then get (17.49) as desired.

To see the full strength of the Hewitt-Savage zero one law (which on homogeneous product spaces subsumes Komogorov’s zero one law because \( \mathcal{T} \subseteq \mathcal{E} \)), we prove:
Theorem 17.15 (Strong Law of Large Numbers) Let $(X, \mathcal{F}, \mu)$ be a probability space and $f \in L^1(X, \mathcal{F}, \mu)$ a real-valued function. For each $n \in \mathbb{N}$, let $f_n: X^N \to \mathbb{R}$ be defined by $f_n(x_i) := f(x_n)$. Then
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f_i = \int f \, d\mu, \quad \mu^\otimes N\text{-a.e.} \tag{17.53}
\]
The convergence takes place in $L^1(X^N, \mathcal{F}^\otimes N, \mu^\otimes N)$ as well.

Proof. The main point is a proof of $\mu^\otimes N$-a.e. convergence; the limit then has to be constant by the Hewitt-Savage zero-one law. Let $\mathcal{E}_n$ be the $\sigma$-algebra of sets in $\mathcal{F}^\otimes N$ that are invariant under any permutation that does not move indices larger than $n$. Consider the conditional expectations $(f_j)_{\mathcal{E}_n}$. Examining the permutations that just swap 1 with index $j \in \{2, \ldots, n\}$ while leaving the other indices unchanged, from $f_1 \circ h_j = f_j$ we get
\[
\forall j = 1, \ldots, n: \quad (f_1)_{\mathcal{E}_n} = (f_j)_{\mathcal{E}_n}. \tag{17.54}
\]
Averaging $(f_j)_{\mathcal{E}_n}$ over $j = 1, \ldots, n$ and noting that $\sum_{j=1}^{n} f_j$ is measurable with respect to $\mathcal{E}_n$ then gives
\[
(f_1)_{\mathcal{E}_n} = \frac{1}{n} \sum_{i=1}^{n} f_i, \quad \mu^\otimes N\text{-a.e.} \tag{17.55}
\]
As $\{\mathcal{E}_n\}_{n \geq 1}$ is non-increasing with $\bigcap_{n \geq 1} \mathcal{E}_n = \mathcal{E}$, Levy's Backward Theorem (see Theorem 10.14) shows
\[
(f_1)_{\mathcal{E}_n} \xrightarrow{n \to \infty} (f_1)_{\mathcal{E}}, \quad \mu^\otimes N\text{-a.e. and in } L^1(X^N, \mathcal{F}^\otimes N, \mu^\otimes N). \tag{17.56}
\]
But $\mu^\otimes N$ is trivial on $\mathcal{E}$ by Theorem 17.14 and so $(f_1)_{\mathcal{E}}$ is constant $\mu^\otimes N$-a.e., namely (since $\mu^\otimes N$ is a probability),
\[
(f_1)_{\mathcal{E}} = \int (f_1)_{\mathcal{E}} \, d\mu^\otimes N = \int f_1 \, d\mu^\otimes N = \int f \, d\mu, \quad \mu^\otimes N\text{-a.e.} \tag{17.57}
\]
Jointly with (17.55–17.56), this now gives the claim. \hfill \Box

We remark that the whole class of exchangeable measures on $(X^I, \mathcal{F}^\otimes I)$ for $I$ infinite can be fully characterized as convex combinations of (homogeneous) product measures. This the content of celebrated de Finetti's Theorem. This theorem also shows that the only exchangeable measures that are trivial on $\mathcal{E}$ are (homogeneous) product measures.

Yet another example of a zero-one law arises when $I$ is an additive group. A shift in the group induces a transformation on $(X^I, \mathcal{F}^\otimes I)$ that preserves (homogeneous) product measures. Moreover, whenever the orbits of the shift are infinite, these product measures are trivial on the $\sigma$-algebra of invariant sets. Theorem 17.15 can then be proved by an appeal to Birkhoff's Ergodic Theorem (see Theorem 9.9).

17.3 Kolmogorov Extension Theorem.

A majority of this section has been devoted to product measures but infinite product spaces support other measures as well. The question we wish to address now is under what conditions the information about finite-dimensional marginals ensures existence,
and determines, the measure on a full product space. This is the content of the celebrated Kolmogorov Extension Theorem.

Recall the notation $\mathcal{F}_J$ from (17.32) for the $\sigma$-algebra for the set of all measurable sets that depend only on the coordinates in $J \subseteq I$. We now have:

**Theorem 17.16 (Kolmogorov’s Extension Theorem)** Let $\{(X_a, \mathcal{F}_a)\}_{a \in I}$ be a family of measurable spaces and, for each finite $J \subseteq I$, let $\mu_J$ be a probability measure on $(\times_{a \in J} X_a, \mathcal{F}_J)$. Suppose that these measures obey the consistency conditions:

$$\forall K, J \subseteq I \text{ finite } \forall A \in \mathcal{F}_{K \cap J}: \quad \mu_J(A) = \mu_K(A). \quad (17.58)$$

If in addition each $(X_a, \mathcal{F}_a)$ is a standard Borel space, then there exists a unique probability measure $\bar{\mu}$ on $(\times_{a \in I} X_a, \mathcal{F}_I)$ such that

$$\forall J \subseteq I \text{ finite } \forall A \in \mathcal{F}_J: \quad \bar{\mu}(A) = \mu_J(A). \quad (17.59)$$

We start by a proposition that highlights the reasons why we need to restrict to standard Borel spaces in Theorem 17.16:

**Proposition 17.17 (Inner regularity)** Let $\mu$ be a finite measure on a standard Borel space $(X, \mathcal{B}(X))$. Then $\mu$ is inner regular, meaning

$$\forall A \in \mathcal{B}(X): \quad \mu(A) = \sup \{\mu(K): K \subseteq A \text{ compact}\}. \quad (17.60)$$

In particular, finite Borel measures on standard Borel spaces are automatically Radon.

For the proof we first recall:

**Lemma 17.18** Let $X$ be a topological space (not necessarily Polish) and let $\mu$ be a finite measure on $(X, \mathcal{B}(X))$. Then for all $A \in \mathcal{B}(X)$:

$$\mu(A) = \inf \{\mu(O): A \subseteq O \text{ open}\} = \sup \{\mu(C): C \subseteq A \text{ closed}\}. \quad (17.61)$$

**Proof.** Consider the class of sets

$$\mathcal{F} := \{A \in \mathcal{B}(X): \text{(17.61) holds}\}. \quad (17.62)$$

It suffices to show that $\mathcal{F}$ is a $\sigma$-algebra because, since $\mathcal{F}$ trivially contains the class of all open and closed sets, it then contains $\mathcal{B}(X)$.

Note that $\mathcal{F}$ contains $\emptyset$ and $X$ and is closed under complements, by additivity of $\mu$ and the fact that open sets are complements of closed sets. It thus suffices to show that $\mathcal{F}$ is closed under finite intersections and increasing limits. Let $A_1, A_2 \in \mathcal{F}$ and let $\epsilon > 0$. Then there are $C_1 \subseteq A_1$ and $C_2 \subseteq A_2$ closed such that $\mu(A_i \setminus C_i) < \epsilon$ for $i = 1, 2$. Set $C := C_1 \cap C_2$. Then

$$A \setminus C = (A_1 \cup A_2) \setminus (C_1 \cup C_2) \subseteq \bigcup_{i=1,2} (A_i \setminus C_2) \quad (17.63)$$

and so

$$\mu(A \setminus C) < \sum_{i=1,2} \mu(A_i \setminus C) < 2\epsilon. \quad (17.64)$$

Since $C$ is closed, we have $A \in \mathcal{F}$.
Next assume that \( \{A_n\}_{n \geq 1} \subseteq \mathcal{F} \) obey \( A_n \uparrow A \) and, given \( \epsilon > 0 \), pick \( \{C_n\} \) closed such that \( C_n \subseteq A_n \) and \( \mu(A_n \setminus C_n) < \epsilon 2^{-n} \) for all \( n \geq 1 \). Then

\[
A \setminus C = \left( \bigcup_{n \geq 1} A_n \right) \setminus \left( \bigcup_{n \geq 1} C_n \right) \subseteq \bigcup_{n \geq 1} (A_n \setminus C_n)
\]

and so \( \mu(A \setminus C) \leq \sum_{n \geq 1} \epsilon 2^{-n} = \epsilon \). But then also

\[
\mu\left( A \setminus \bigcup_{k=1}^{n} C_k \right) < \epsilon
\]

because as \( n \) increases the left hand side decreases, by the Downward Monotone Convergence Theorem for sets, to \( \mu(A \setminus C) \). Since \( \bigcup_{k=1}^{n} C_k \) is closed, \( A \in \mathcal{F} \).

**Proof of Proposition 17.17.** We first observe that, by Lemma 17.18, the statement holds for the special case of the Hilbert cube \([0, 1]^\mathbb{N}\) because that space is compact in the product Euclidean topology and so its closed subsets are automatically compact.

In order to extend the claim to a general Polish space \( X \), we recall that, by Proposition 15.13, \( X \) is homeomorphic to a \( G_\delta \)-subset of \([0, 1]^\mathbb{N}\). Let \( f : X \hookrightarrow [0, 1]^\mathbb{N} \) be this homeomorphism. We claim

\[
\forall K \subseteq [0, 1]^\mathbb{N} \text{ compact: } f^{-1}(K) \text{ is compact.} \tag{17.67}
\]

This follows from the fact that a continuous image of a compact set is compact, but let us write a proof of this in detail: In metric spaces it suffices to prove sequential compactness, so let \( \{x_n\}_{n \geq 1} \subseteq f^{-1}(K) \). Then \( \{f(x_n)\}_{n \geq 1} \subseteq K \) and so there exists \( \{n_k\}_{k \geq 1} \subseteq \mathbb{N} \) increasing such that \( f(x_{n_k}) \to y \in K \). But \( f^{-1} \) is continuous and so, since \( X \) is complete, \( x_{n_k} = f^{-1}(f(x_{n_k})) \) converges to some \( x \in X \). But \( f \) is continuous too and so \( y = f(x) \), meaning that \( x \in f^{-1}(K) \). Hence \( f^{-1}(K) \) is compact.

Let now \( \mu \) be a finite measure on \((X, \mathcal{B}(X))\). Then (since \( f^{-1} \) is Borel measurable), \( \nu := \mu \circ f \) is a finite Borel measure on \([0, 1]^\mathbb{N}\). Given \( A \in \mathcal{B}(X) \) and \( \epsilon > 0 \), let \( K \subseteq f(A) \) be compact with \( \nu(f(A) \setminus K) < \epsilon \). But then

\[
\mu\left( A \setminus f^{-1}(K) \right) = \nu\left( f(A) \setminus K \right) < \epsilon.
\]

As \( f^{-1}(K) \) is compact by (17.67) with \( f^{-1}(K) \subseteq A \), we get (17.60). \( \square \)

We note the following concept:

**Definition 17.19 (Radon space)** A topological space for which every finite Borel measure is inner regular (which is what (17.60) adds to the general outer regularity statement in (17.61)) is called a Radon space.

Standard Borel spaces are thus Radon. Unfortunately, Radon spaces are not generally closed under products, which is what we need in the proof below. Hence we have to restrict to standard Borel spaces.

**Proof of Theorem 17.16.** We start with some a remark on notation. For \( J \) finite, the measurable space \((\times_{\alpha \in I} X_\alpha, \mathcal{F}_J)\) is naturally identified with the finite-product space \((\otimes_{\alpha \in J} X_\alpha, \otimes_{\alpha \in J} \mathcal{F}_\alpha)\) by the canonical projection \( \pi_J(\{x_\alpha\}_{\alpha \in I}) := \{x_\alpha\}_{\alpha \in J} \). Using \( \pi_J^{-1} \), any...
measure on \((\bigotimes_{\alpha \in \mathcal{A}} X_\alpha, \bigotimes_{\alpha \in \mathcal{A}} \mathcal{F}_\alpha)\) then induces a measure on \((\bigotimes_{\alpha \in I} X_\alpha, \mathcal{F}_I)\). This permits us to “lift” properties of former measures to the latter ones.

Consider the algebra \(\mathcal{A}\) defined in (17.17). Every \(A \in \mathcal{A}\) depends only a finite number of coordinates, say those in \(I \subseteq I\), and so \(A \in \mathcal{F}_I\). We may thus set

\[
\forall A \in \mathcal{F}_I: \quad \bar{\mu}(A) := \mu_I(A). \tag{17.69}
\]

By (17.58), this is independent of \(I\) such that \(A \in \mathcal{F}_I\) and so \(\bar{\mu}\) is well defined on \(\mathcal{A}\). The argument (17.18–17.20) can then be followed literally and so, in order to show that \(\bar{\mu}\) is countably additive on \(\mathcal{A}\), we have to prove that (17.21) implies (17.22).

We will again aim at proving the contrapositive. Let \(\{B_n\}_{n \geq 1} \subseteq \mathcal{A}\) be non-decreasing with \(\delta := \lim_{n \to \infty} \bar{\mu}(B_n) > 0\). By repeating terms in the sequence we may assume the existence of a sequence \(\{a_n\}_{n \geq 1}\) of distinct indices from \(I\) such that \(B_n \in \mathcal{F}_{a_n}\) for \(I_n := \{a_1, \ldots, a_n\}\). Under our assumption, each \(X_{a_n}\) is a Polish space and \(\mathcal{F}_{a_n} = B(X_{a_n})\).

Hence also \((\bigotimes_{i=1}^n X_{a_i}, \bigotimes_{i=1}^n \mathcal{F}_{a_i})\) is a standard Borel space. By Lemma 17.17, for each \(n \geq 1\) there is a compact set \(C_n \subseteq \bigotimes_{i=1}^n X_{a_i}\) such that

\[
C_n \times \bigotimes_{a \notin I_n} X_a \subseteq B_n \land \mu_I\left(C_n \times \bigotimes_{a \notin I_n} X_a\right) \geq \mu_I(B_n) - \delta 2^{-n-1}. \tag{17.70}
\]

Writing the objects using (17.69), the union bound shows

\[
\forall n \geq 1: \quad \bar{\mu}\left(\bigcap_{a \notin I_n} (C_i \times \bigotimes_{a \notin I_n} X_a)\right) \geq \bar{\mu}(B_n) - \frac{1}{2} \delta. \tag{17.71}
\]

But \(\bar{\mu}(B_n) \geq \delta\) and so we get

\[
\forall n \geq 1: \quad \bigcap_{a \notin I_n} (C_i \times \bigotimes_{a \notin I_n} X_a) \neq \emptyset. \tag{17.72}
\]

Denote \(I_x := \bigcup_{n \geq 1} I_n \neq I\) and, for each \(n \geq 1\), pick

\[
x(n) = \{x^{(n)}_{a_i}\}_{a_i \in I_x} \in \bigcap_{a_i \in I_x \setminus I_n} (C_i \times \bigotimes_{a_i \in I_x \setminus I_n} X_a) \tag{17.73}
\]

For each \(1 \leq i \leq n\), let \(P_n\) denote the intersection of the projections of \(\{C_i\}_{i=1}^n\) on \(X_{a_i}\). Since \(\{P_n\}_{n \geq 1}\) are compact and \(I_x\) is countable, the Cantor diagonal argument permits us to choose \(\{n_k\}_{k \geq 1}\) such that \(x^{(n_k)}_{a_i} \to x_{a_i} \in \bigcap_{n \geq 1} P_n\) for each \(a_i \in I_x\). Picking \(x_{a_i} \in X_{a_i}\) arbitrarily for \(a \in I \setminus I_x\) (this uses \(\times_{a \in I \setminus I_x} X_a \neq \emptyset\)) we obtain \(x = \{x_{a_i}\}_{a_i \in I}\) that belongs to all the sets in (17.72). Hence

\[
\bigcap_{n \geq 1} B_n \supseteq \bigcap_{n \geq 1} (C_n \times \bigotimes_{a \notin I_n} X_a) \neq \emptyset. \tag{17.74}
\]

This shows that (17.21) implies (17.22) and thus proves that \(\bar{\mu}\) is countably additive on \(\mathcal{A}\). By Theorem 11.6, \(\bar{\mu}\) extends to a unique measure on \(\sigma(\mathcal{A}) = \bigotimes_{\alpha \in I} \mathcal{F}_\alpha\). \(\square\)

Theorem 17.16 goes back to A.N. Kolmogorov’s “Grundbegriffe der Wahrscheinlichkeitsrechnung” (Ergebnisse Mathematik, Berlin, Springer, 1933); sometimes credit is given to PJ. Daniell as well for his development of infinite-dimensional integration theory in the context of what is now called Daniell’s integra. As shown by E.S. Andersen

Theorem 17.16 finds many uses in probability. Among the most elegant one is the construction of Gaussian measures on arbitrary product spaces. This relies heavily on the fact that a Gaussian measure on $\mathbb{R}^n$ is determined by the first and second moments of its one-dimensional projections — the so called mean and the covariances. This makes checking the consistency conditions (17.58) an elementary exercise and so, as a result, Gaussian measures can be constructed in remarkable generality.