Measure Theory with Ergodic Horizons
Lecture 22

Theorem (Fubini-Tonelli for $\mathcal{I} \otimes \mathcal{J}$). Let $(X, \mathcal{I}, \mu)$ and $(Y, \mathcal{J}, \nu)$ be σ-finite measure spaces. Let $f : X \times Y \to \overline{\mathbb{R}} := [-\infty, \infty]$ be a $\mathcal{I} \otimes \mathcal{J}$-measurable function. Then:

(a) $f_x : Y \to \overline{\mathbb{R}}$ and $f^y : X \to \overline{\mathbb{R}}$ are $\mathcal{J}$ and $\mathcal{I}$-measurable for all $x \in X$ and $y \in Y$.

(b) Tonelli. If $f \geq 0$, then:

(i) $x \mapsto \int_Y f_x \, d\nu$ and $y \mapsto \int_X f^y \, d\mu$ are $\mathcal{I}$ and $\mathcal{J}$-measurable.

(ii) $$\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(y).$$

(c) Fubini. If f is $\mu \times \nu$-integrable then:

(i) $x \mapsto \int_Y f_x \, d\nu$ and $y \mapsto \int_X f^y \, d\mu$ are $\mathcal{I}$ and $\mathcal{J}$-measurable and integrable.

(ii) $$\int_X \int_Y f_x(y) \, d\nu(y) \, d\mu(x) = \int_{X \times Y} f \, d(\mu \times \nu) = \int_Y \int_X f^y(x) \, d\mu(x) \, d\nu(y).$$

Proof. We have already proven (a).

For (b), we have shown it for indicator functions, so the linearity of the integral implies (b) for simple. Finally, any $\mathcal{I} \otimes \mathcal{J}$-measurable function $f \geq 0$ is an increasing limit of simple functions, so (b) follows by MCT.

For (c), write $f = f^+ - f^-$ and note that both f^+, f^- are intergrable, and apply part (b) to f^+ and f^-. The only thing to note is that the functions $x \mapsto \int_Y f_x^{\pm}\, d\nu$ and $y \mapsto \int_X (f^{\pm})^y\, d\mu$ are finite a.e. by (b) because $\int_X\int_Y f_x^{\pm}\, d\nu\, d\mu = \int_Y \int_X (f^{\pm})^y\, d\mu\, d\nu(y) = \int_{X\times Y} f^{\pm}\, d\mu\times\nu < \infty$. $\square$

Infinite product spaces.

Let I be an index set, possibly unctbl and let $(X_i, \mathcal{B}_i, \mu_i)_{i\in I}$ be a sequence of measure spaces. We would like to define a measure μ on the product

$$X := \prod_{i\in I} X_i.$$

Firstly, we let $\mathcal{B}$ be the σ-algebra generated by $[i \mapsto B_i] := B_i \times \prod_{j\in I\setminus\{i\}} X_j$ where $i \in I$, $B_i \in \mathcal{B}_i$. We would like μ to extend the finite products $\mu_{i_1} \times \mu_{i_2} \times \dots \times \mu_{i_n}$ for $i_1, \dots, i_n \in I$ distinct, i.e.

$$\mu\left([i_1 \mapsto B_{i_1}, \dots, i_n \mapsto B_{i_n}]\right) = \mu_{i_1}(B_{i_1}) \cdot \mu_{i_2}(B_{i_2}) \cdot \ldots \cdot \mu_{i_n}(B_{i_n}) \cdot \prod_{j\in I\setminus\{i_1,\dots,i_n\}} \mu_j(X_j) \qquad (*)$$

where

$$[i_1 \mapsto B_{i_1}, \dots, i_n \mapsto B_{i_n}] := [i_1 \mapsto B_{i_1}] \cap \dots \cap [i_n \mapsto B_{i_n}].$$

To ensure that cylinders can have finite nonzero measure, we need that all but finitely many μ_i are finite and in fact bounded above and below. Since we handled finite products, we can restrict our attention to the case that each μ_i is a probability measure.

To prove the existence of such a measure μ satisfying $(*)$, let $\mathcal{A}$ be the algebra generated by the cylinders and note, as usual, that each $A \in \mathcal{A}$ is a finite disjoint union of cylinders. Thus, in the usual fashion we may define a finitely additive measure μ on $\mathcal{A}$ satisfying $(*)$. This is done by the fact that and finite union of cylinders has all finitely many coordinates equal to X_i, so the statements of well-definedness and finite additivity reduces to that for finite products, which we already proved.

Finite additivity implies ctbl super-additivity:

$$\mu\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) \geq \sum_{n \in \mathbb{N}} \mu(A_n)$$

whenever $\bigsqcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$, for $A_n \in \mathcal{A}$. Thus, we only need to show that μ is also ctbly subadditive.

<u>Theorem</u> (Kakutani 1943). μ is ctbly subadditive, hence by Carathéodory's thm, the is a unique probability measure μ on $\mathcal{B}$ satisfying $(*)$, for arbitrary probability spaces $(X_i, \mathcal{B}_i, \mu_i)$.

Remark. This theorem was proved by Kolmogorov in 1933 for $I := \mathbb{N}$ and $(X_i, \mathcal{B}_i, \mu_i) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and is known as the Kolmogorov consistency theorem. It was later extended by Doob in 1938 to include other Borel probability measures on $[0,1]$. All these proofs relied on the regularity and tightness of the involved measures, which are inherently topological notions, but Kakutani found a topology-free proof.

Proof of a special case of Kakutani's theorem: the Kolmogorov consistency theorem.

We only prove in the case that $I := \mathbb{N}$ and each $(X_i, \mathcal{B}_i, \mu_i)$ is standard, hence by the Borel isomorphism theorem, we may assume that $X_i = 2^{\mathbb{N}}$ and μ_i is strongly regular and tight. To show the ctbl subadditivity of μ on the algebra $\mathcal{A}$, it is enough to show that if A and the A_n are cylinders with $A = \bigsqcup_{n \in \mathbb{N}} A_n$ then $\mu(A) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

We do the usual trick of replacing A with a compact and the A_n with open sets. Let $A = B_0 \times B_1 \times \dots \times B_\ell \times \prod_{i > \ell} X_i$ and $A_n = B_{n0} \times B_{n1} \times \dots \times B_{n\ell_n} \times \prod_{i > \ell_n} X_i$. By tightness, there are compact sets $K_j \subseteq B_j$ with $\mu_j(K_j) \approx_{\varepsilon/2\ell} \mu_j(B_j)$, so $\mu(A') \approx_{\varepsilon/2} \mu(A)$, where $A' := K_0 \times K_1 \times \dots \times K_\ell \times \prod_{i > \ell} X_i$. By strong regularity, there are open sets $U_{nj} \supseteq B_{nj}$ with $\mu_j(U_{nj}) \approx$ $\approx_{2^{-(n+1)} \cdot \varepsilon/2\ell_n} \mu_j(B_{nj})$, so $\mu(\tilde{A}_n) \approx_{2^{-(n+1)} \cdot \varepsilon/2} \mu(A_n)$, where $\tilde{A}_n := U_{n0} \times U_{n1} \times \dots \times U_{n\ell_n} \times \prod_{i > \ell_n} X_i$. By the definition of product topology (look up), the $\tilde{A}_n$ are open, and by Tychonoff's theorem, A' is compact. Since $A' \subseteq \bigcup_{n \in \mathbb{N}} \tilde{A}_n$, there is a finite $N \in \mathbb{N}$ such that $A' \subseteq \bigcup_{n < N} \tilde{A}_n$.

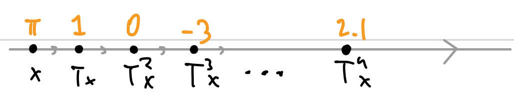
Thus, $\mu(A) \approx_{\varepsilon/2} \mu(A') \le \mu\left(\bigcup_{n \le N} \tilde{A}_n\right) \le \sum_{n \le N} \mu(\tilde{A}_n) \le \sum_{n \in \mathbb{N}} \mu(\tilde{A}_n) \approx_{\varepsilon/2} \sum_{n \in \mathbb{N}} \mu(A_n)$. $\square$

Birkhoff ergodic theorem.

Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ be a $(\mathcal{B}, \mathcal{B})$-measurable μ-preserving transformation. Recall that a set $S \subseteq X$ is called T-invariant if it is a union of T-orbits; equivalently $T^{-1}(S) = S$. Finally, recall that T is said to be ergodic if every T-invariant set in $\mathcal{B}$ is null or conull. In particular, ergodicity is a global property: we need to check something for every T-invariant set in $\mathcal{B}$. The following famous theorem translates this global property into a local finitary property at a.e. point of X. The first version of it is due to Birkhoff (1931), inspired by a weaker result of von Neumann. This theorem is considered the birth of ergodic theory.

Classical pointwise ergodic theorem (Birkhoff 1931). Let $(X, \mathcal{B}, \mu)$ be a probability space and $T : X \to X$ be a $(\mathcal{B}, \mathcal{B})$-measurable transformation that preserves μ, i.e. $\mu(T^{-1}(S)) = \mu(S)$ for all $S \in \mathcal{B}$. Then T is ergodic iff for each $f \in L^1(X, \mu)$,

$$(*) \quad \lim_{n \to \infty} A_n f(x) = \int f \, d\mu \quad \text{a.e. } x \in X,$$



where $A_n f(x) :=$ the average of f over $\{x, Tx, \dots, T^n x\} = \frac{1}{n+1} \sum_{i=0}^{n} f(T^i x)$.

The $\Leftarrow$ direction of this theorem is rather trivial:

Proof of $\Leftarrow$. Let $S \in \mathcal{B}$ be a T-invariant set. Then $f := \mathbb{1}_S$ is constant on every orbit of T, with values either 0 or 1. Hence $\lim\limits_{n\in\mathbb{N}} A_n \mathbb{1}_S$ is either 0 or 1. By $(*)$, $\lim\limits_{n\to\infty} A_n \mathbb{1}_S = \int \mathbb{1}_S \, d\mu = \mu(S)$, so $\mu(S)$ is 0 or 1. $\square$

The nontrivial direction $\Rightarrow$ says that for an ergodic T,

time averages $A_n f \xrightarrow{\text{converge to}}$ spatial average $\int f \, d\mu$.

In other words, the global statistic $\int f \, d\mu$ can be seen by examining ctbly many points $\{x, Tx, \ldots, T^n x, T^{n+1} x, \ldots\} =:$ the forward T-orbit of x.

Next time we will give applications of this powerful theorem and give a short and modern proof of the nontrivial direction $\Rightarrow$.