MTP24, Lecture 2: Random Variables, Independence, Integration and Conditioning

Alexander Gnedin

Queen Mary, University of London

Measurable functions

Definition Given two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, a function $X : \Omega \to \Omega'$ is said to be *measurable* if it satisfies

$$X^{-1}(B') \in \mathcal{F} \quad \text{for all} \quad B' \in \mathcal{F}' \tag{1}$$

The σ-algebra $\sigma(X)$ *induced* (or generated) by X is comprised of the sets $X^{-1}(B')$.

• It is sufficient to require the condition (1) to hold for any system $\mathcal{G}'$ of generators (s.t. $\mathcal{F}' = \sigma(\mathcal{G}')$).

• $\sigma(X)$ is the smallest σ-algebra in Ω necessary to make X measurable.

• When the target space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we speak of a (real-valued) random variable. Then the measurability condition holds if $X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.

Product σ-algebras

Definition For a family $(X_t, t \in T)$ of measurable functions $X_t : \Omega \in \Omega'$, we denote $\sigma(X_t, t \in T)$ the σ-algebra generated by the sets $X_t^{-1}(B')$, where $B' \in \mathcal{F}', t \in T$.

Example Let $\Omega = \{0, 1\}^\infty$, $X_n(\omega) = \omega_n$. Then $\sigma(X_n, n \in \mathbb{N})$ coincides with the σ-algebra generated by the finite-dimensional cylinders $A(\epsilon_1, \dots, \epsilon_n) = \{\omega \in \Omega : \omega_1 = \epsilon_1, \dots, \omega_n = \epsilon_m\}$. This is also an example of product σ-algebra.

Definition For a family $((\Omega_t, \mathcal{F}_t), t \in T)$ of measurable spaces, let $\Omega = \prod_{t \in T} \Omega_t$ (Cartesian product), and for $\omega \in \Omega$ let $X_t(\omega)$ be the t-th coordinate. The *product* σ-algebra

$$\bigotimes_{t \in T} \mathcal{F}_t = \sigma(X_t, t \in T)$$

is the σ-algebra generated by the family $(X_t, t \in T)$.

Pushforward of measures

Definition For $(\Omega, \mathcal{F}, \mu)$ measure space and measurable $X : \Omega \to \Omega'$, where $(\Omega', \mathcal{F}')$ another measurable space, the measure on $(\Omega', \mathcal{F}')$

$$\mu_X(B') := \mu(X^{-1}(B')), \quad B' \in \mathcal{F}'$$

is called *pushforward* of μ (induced by X).

Example Distribution of a random variable X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is the induced probability measure $P(A) = \mathbb{P}[X \in A]$,$A \in \mathcal{B}(\mathbb{R})$.

Definition $(\Omega, \mathcal{F}, \mu), (\Omega', \mathcal{F}, \mu')$ are isomorphic $\bmod 0$ if there exist nullsets $A \in \mathcal{F}, A' \in \mathcal{F}'$ and a bijection $f : \Omega \setminus A \to \Omega' \setminus A'$ such that f, f^{-1} are both measurable and preserve the measure.

• Many probability spaces are isomorphic $\bmod 0$ to the 'standard probability space' $([0,1], \mathcal{B}([0,1]), \lambda)$.

Product of measures

Definition Let $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces. The *product measure space* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$ carries the *product measure* uniquely defined by extending the formula for 'rectangles'

$$\mu_1 \times \mu_2(B_1 \times B_2) = \mu_1(B_1)\mu_2(B_2), \quad B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2$$

to $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Definition Let $(P_t, t \in T)$ be probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. There exists a unique priobability measure P on the infinite product space $(\mathbb{R}^T, \mathcal{B}(R^T))$ with marginal measures that define the joint distribution of coordinates $(X_{t_1}, \ldots, X_{t_n})$ for $\{t_1, \ldots, t_n\} \subset T$ by the product formula

$$P_{t_1, \ldots, t_n}(B_1 \times \cdots \times B_n) = P_{t_1}(B_1) \cdots P_{t_n}(B_n),$$

where $B_i \in \mathcal{B}(\mathbb{R}), n \in \mathbb{N}$.

Independence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Definition

(i) Events $(A_t, t \in T)$ are independent if

$$\mathbb{P}(A_{t_1} \cap \cdots \cap A_{t_n}) = \mathbb{P}(A_{t_1}) \cdots \mathbb{P}(A_{t_n})$$

for any $\{t_1, \ldots, t_n\} \subset T$, $n \in \mathbb{N}$.

(ii) σ-algebras $(\mathcal{F}_t, t \in T), \mathcal{F}_t \subset \mathcal{F}$, are independent if any finite collection of events $A_{t_1} \in \mathcal{F}_1, \ldots, A_{t_n} \in \mathcal{F}_n, n \in \mathbb{N}$, is independent.

(iii) Random variables $(X_t, t \in T)$ are independent if the generated σ-algebras $\sigma(X_t)$ are independent.

Zero-one laws and tail events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For a series of events $A_n \in \mathcal{F}, n \in \mathbb{N}$,

$$\{A_n \ \text{i.o.}\} := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Borel-Cantelli Lemma

($\rightarrow$) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$.

($\leftarrow$) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Example (the problem of records) Let $X_1, X_2, \ldots$ be i.i.d. with continuous c.d.f., $A_n = \{X_n = \max(X_1, \ldots, X_n)\}$ the event 'X_n is a record'.
By symmetry (exchangeability of X_n's) $\mathbb{P}(A_n) = 1/n$ and the events A_n are independent. Hence there are infinitely many records almost surely.

Definition Let $\mathcal{F}_n$ be σ-algebras $\mathcal{F}_n \subset \mathcal{F}$. The *tail σ-algebra* is defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{k=n}^{\infty} \mathcal{F}_k\right).$$

Each $A \in \mathcal{T}$ is a *tail event*.

Example Let $X_1, X_2, \ldots$ be random variables, $\mathcal{F}_n = \sigma(X_n)$. Tail events:

(i) $\{X_n > 2024 \text{ i.o. }\}$,

(ii) $\left\{\frac{X_1+\cdots+X_n}{n} \to 0 \text{ as } n \to \infty\right\}$,

(iii) $\{\sum_{n=1}^{\infty} X_n \text{ converges}\}$,

Not tail events:

(a) $\{X_n > X_1 \text{ i.o. }\}$,

(b) $\{\sum_{n=1}^{\infty} X_n = 27\}$.

Kolmogorov's 0-1 **Law** Suppose $\mathcal{F}_n, n \in \mathbb{N}$ are independent σ-algebras. Then their tail σ-algebra $\mathcal{T}$ is trivial, meaning that $\mathbb{P}(A) = 0$ or 1 for every $A \in \mathcal{T}$.

Example Suppose $X_n \sim \mathcal{N}(m_n, \sigma_n^2), n \in \mathbb{N}$, are independent normal r.v. Does the series $\sum_{n=1}^{\infty} X_n$ converge?

Kolmogorov Three-Series Theorem Let $X_1, X_2, \ldots$ be independent random variables. The series $\sum_n X_n$ converges a.s. (almost surely) if and only if for some $c > 0$

(i) $\sum_n \mathbb{P}[|X_n| > c] < \infty$,

(ii) $\sum_n \mathbb{E}[X_n \mathbf{1}(|X_n| \leq c)] < \infty$,

(iii) $\sum_n \mathrm{Var}[X_n \mathbf{1}(|X_n| \leq c)] < \infty$,

• If (i), (ii), (iii) hold for some $c > 0$ then also for all $c > 0$.

• For normal r.v.'s the convergence holds iff both series $\sum_n m_n$ and $\sum_n \sigma_n^2$ converge.

Lebesgue integral and expectation

$(\Omega, \mathcal{F}, \mu)$ measure space, $X = X(\omega)$ measurable function with values in $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, $\quad \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Suppose first $X \geq 0$, and let

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \, \mathbf{1}\left[\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right] + n\,\mathbf{1}[X \geq n],$$

so $X_n \uparrow X$ a.s. Then set

$$\int_\Omega X_n(\omega)\mu(\mathrm{d}\omega) := \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \, \mu\left\{\omega : \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n}\right\} + n\,\mu\{\omega : X(\omega) \geq n\},$$

and define the Lebesgue intergal as the (monotone) limit

$$\int_\Omega X(\omega)\mu(\mathrm{d}\omega) := \lim_{n\to\infty} \int_\Omega X_n(\omega)\mu(\mathrm{d}\omega).$$

For the general X, split $X = X_+ - X_-$ with $X_\pm := \max(\pm X, 0)$, and define

$$\int_\Omega X(\omega)\mu(\mathrm{d}\omega) := \int_\Omega X_+(\omega)\mu(\mathrm{d}\omega) - \int_\Omega X_-(\omega)\mu(\mathrm{d}\omega)$$

provided at least one of the integrals in the r.h.s. is finite.

Example The Dirichlet function $f(x) = \mathbf{1}_{\mathbb{R}\setminus\mathbb{Q}}(x)$ has Lebesgue integral 0 (w.r.t. λ), but is not Riemann-integrable.

• For r.v. X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ the *expectation* is denoted

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(\mathrm{d}\omega).$$

This can be computed as the *Lebesgue-Stiltjes* integral

$$\mathbb{E}[X] := \int_{\mathbb{R}} x \, \mathrm{d}F_X(x),$$

where $F_X(x) := \mathbb{P}[X \le x]$ is the c.d.f. of X.

Constructing a non-product measure

Definition Given a product space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, call a function $K(\omega_1, B_2) : (\Omega_1, \mathcal{F}_2) \to ([0,1], \mathcal{B}([0,1]))$ a *transition function* (aka *Markov kernel*) if

(i) $K(\omega_1, \cdot)$ is a probability measure on $(\Omega_2, \mathcal{F}_2)$ for each $\omega_1 \in \Omega_1$,

(ii) $K(\cdot\,, B_2)$ is a measurable function for each $B_2 \in \mathcal{F}_2$.

Theorem For probability measure P_1 on $(\Omega_1, \mathcal{F}_1)$, the formula

$$P(B_1 \times B_2) = \int_{B_1} K(\omega_1, B_2) P_1(\mathrm{d}\omega_1)$$

defines a unique probability measure on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.

Exchanging $\mathbb{E}$ and lim

Let $X_1, X_2, \ldots, X, Y$ be (real) random variables on some $(\Omega, \mathcal{F}, \mathbb{P})$.

Monotone Convergence If $X_n \uparrow X$ a.s. and $X_n \geq Y$ for some Y with $\mathbb{E}[Y] > -\infty$ the $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$ as $n \to \infty$.

Fatou Lemma If $X_n \geq Y, \mathbb{E}[Y] > -\infty$ then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n\right].$$

Dominated Convergence If $|X_n| \leq Y$, where $\mathbb{E}[Y] < \infty$ then $X_n \to X$ a.s. implies

(i) $\mathbb{E}\left[X\right] < \infty$,

(ii) $\mathbb{E}\left[X_n\right] \to \mathbb{E}\left[X\right]$ as $n \to \infty$,

(iii) $\mathbb{E}|X_n - X| \to 0$ as $n \to \infty$.

Absolute continuity

Definition Given measures μ, ν on $(\Omega, \mathcal{F})$, we say that μ *dominates* ν (written $\mu \gg \nu$) if for $A \in \mathcal{F}$

$$\mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Then it is also said that ν is *absolutely continuous* w.r.t. μ. The measure are *equivalent*, written $\mu \sim \nu$, if $\mu \gg \nu$ and $\nu \gg \mu$.

The discrete case $\mu = \sum_k p_k \delta_{\omega_k}, \nu = \sum_k q_k \delta_{\omega_k}$, where $p_k > 0, q_k \geq 0$.

$$\int_\Omega f(\omega)\nu(\mathrm{d}\omega) =$$

$$\sum_k f(\omega_k) q_k = \sum_k f(\omega_k) p_k \left(\frac{q_k}{p_k}\right) =$$

$$\int_\Omega f(\omega)\xi(\omega)\mu(\mathrm{d}\omega),$$

where $\xi(\omega) = \sum_k \left(\frac{q_k}{p_k}\right) \mathbf{1}[\omega = \omega_k]$.

Radon-Nykodým Theorem If $\mu \gg \nu$ and μ is σ-finite, then there exists measurable $\xi : \Omega \to \mathbb{R}_+$ such that

$$\int_\Omega f(\omega)\nu(\mathrm{d}\omega) = \int_\Omega f(\omega)\xi(\omega)\mu(\mathrm{d}\omega)$$

for all measurable $f : \Omega \to \mathbb{R}$. Such function ξ is unique up to a set of μ-measure zero.

• Such ξ is called the Radon-Nykodým derivative, denoted $\xi = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

• For probability distributions on $\mathbb{R}$ the absolute continuity means existence of the density function f such that the c.d.f. satisfies

$$F(x) = \int_{-\infty}^{x} f(y)\mathrm{d}y.$$

So f is the RN derivative w.r.t. λ.