Probability 2018
2

Giovanni Pistone
www.giannidiorestino.it

Version February 19, 2018

Simple functions

Let $(S, \mathcal{F})$ be a measurable space.

Definition

A measurable real function, $f \colon S \to \mathbb{R}$, $f^{-1} \colon \mathcal{B} \to \mathcal{F}$, is simple if it takes a finite number of values; equivalently, it is of the form $f = \sum_{k=1}^{m} a_k \mathbf{1}_{A_k}$, $a_k \in \mathbb{R}$, $A_k \in \mathcal{F}$, $k = 1, \dots, m$, $m \in \mathbb{N}$. The algebra with unity of all simple functions is denoted by $\mathcal{S}$; the cone of all non-negative simple function is denoted by $\mathcal{S}_+$.

- Both $\mathcal{S}$ and $\mathcal{S}_+$ are closed for $\vee$ and $\wedge$; $f = f^+ - f^-$, $f \in \mathcal{S}$.
- If f is measurable and non-negative, there exist an increasing sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{S}_+$ such that $\lim_{n\to\infty} f_n(s) = f(s)$, $s \in S$.
- If f is measurable and bounded, there exist an sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{S}$ such that $\lim_{n\to\infty} f_n = f$ uniformely.

Ch. 5 of D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991

Integral of a non-negative function

Let $(S, \mathcal{F}, \mu)$ be a measure space.

Definition

- If $f \in \mathcal{S}$, $f = \sum_{k=1}^{m} a_k \mathbf{1}_{A_k}$, we define its integral to be

$$\int f \; d\mu = \sum_{k=1}^{m} a_k \mu(A_k) \quad \text{where } 0 \cdot \infty = \infty \cdot 0 = 0$$

- If $f : S \to [0, +\infty]$ is measurable, namely $f \in \mathcal{L}_+$, we define its integral to be

$$\int f \; d\mu = \sup \left\{ \int h \; d\mu \middle| h \in \mathcal{S}_+, h \leq f \right\}$$

- The integral is linear and monotone on $\mathcal{S}^1 = \left\{ f \in \mathcal{S} \middle| \int f^+ \; d\mu, \int f^- \; d\mu \leq \infty \right\}$. The integral is convex and monotone on $\mathcal{L}_+$.
- If $f \in \mathcal{L}_+$ and $\int f \; d\mu = 0$, then $\mu \{ f > 0 \} = 0$.

Monotone-Convergence Theorem

Theorem (MON)

let $(f_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence in $\mathcal{L}_+$. Then the pointwise limit $f = \lim_{n\to\infty} f_n$ belongs to $\mathcal{L}_+$ and $\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu$.

- A sequence of simple functions converging to f is always available.
- If $\alpha, \beta \in \mathbb{R}_{>0}$, $f, g \in \mathcal{L}_+$, then :

$$\int (\alpha f + \beta g) \, d\mu = \int \alpha f \, d\mu + \int \beta g \, d\mu$$

- Exercise: If $\mu(S) = 1$, then $\int f \, d\mu = \int_0^\infty \mu\{f > u\} \, du$.

Proof of MON: Appendix A5 of D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991; Bertrand Lods *Lecture Notes*.

Fatou Lemmas

Theorem (FATOU)

Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}_+$.

1. $\int (\liminf_{n\to\infty} f_n)\ d\mu \leq \liminf_{n\to\infty} \int f_n\ d\mu$.
2. *If, moreover,* $f_n \leq g$, $n \in \mathbb{N}$, *and* $\int g\ d\mu < \infty$, *then* $\int (\limsup_{n\to\infty} f_n)\ d\mu \geq \limsup_{n\to\infty} \int f_n\ d\mu$.

- Exercise: Prove 1. by observing that lim sup is the limit of an increasing sequence.
- Exercise: If $(f_n)_{n\in\mathbb{N}}$ is a decreasing sequence in $\mathcal{L}_+$ and $\int f_1\ d\mu < \infty$, then $\int (\lim_{n\to\infty} f_n)\ d\mu = \lim_{n\to\infty} \int f_n\ d\mu$.
- Exercise: If $(f_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}_+$ and $f_n \leq g$, $n \in \mathbb{N}$, $\int g\ d\mu < \infty$, then $\int (\lim_{n\to\infty} f_n)\ d\mu = \lim_{n\to\infty} \int f_n\ d\mu$.

Integrability

Definition

- Let $\mathcal{L}^1$ be the vector space of measurable real functions such that

$$\int |f| \ d\mu = \int f^+ \ d\mu + \int f^- \ d\mu < \infty$$

- Define the integral to be the linear mapping

$$\mathcal{L}^1 \ni f \mapsto \int f \ d\mu = \int f^+ \ d\mu - \int f^- \ d\mu \in \mathbb{R}$$

- Assignment: Revise L^1-convergence and Dominated Convergence Theorem in Ch 5 of D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991 or Bertrand Lods *Lecture Notes*.

Expectation

- Let $\mathbb{E}\colon \mathcal{L}^{\infty}(S,\mathcal{S}) \to \mathbb{R}$ be such that
  - $\mathbb{E}(\mathbf{1}) = 1$.
  - $\mathbb{E}$ is linear and positive (hence monotone).
  - $\mathbb{E}$ is continuous on non-increasing sequence converging to 0.
- Every such $\mathbb{E}$ defines a probability measure when restricted to indicators, $\mathrm{P}(A) = \mathbb{E}(\mathbf{1}_A)$ and $\mathbb{E}(f) = \int f \; d\,\mathrm{P}$
- A similar observation holds for a $\mathbb{E}\colon \mathcal{L}_+(S,\mathcal{S})$.
- If $f \in \mathcal{L}(S,\mathcal{L})$, as $f = f_+ - f_-$ and $|f| = f_+ + f_-$, if $\mathbb{E}(|f|) < \infty$ then $\mathbb{E}(f_+), \mathbb{E}(f_-) < \infty$. In such a case, we say that $f \in \mathcal{L}^1(S,\mathcal{S},\mathrm{P})$ and define $\mathbb{E}(f) = \mathbb{E}(f_+) - \mathbb{E}(f_-)$.
- $\mathbb{E}\colon \mathcal{L}^1(S,\mathcal{S},\mathrm{P}) \to \mathbb{R}$ is positive, linear, normalized, continuous for the bounded pointwise convergence.
- Assigment: carefully check everything!

Densities

Let be given a measure space $(S, \mathcal{F}, \mu)$ and a measurable non-negative mapping $p \colon S \to \mathbb{R}$ such that $\int p \; d\mu < \infty$.
The set function

$$p \cdot \mu \colon \mathcal{F} \to \mathbb{R}_{>0}, \quad A \mapsto \int \mathbf{1}_A p \; d\mu$$

is a bounded measure. In fact: $p \cdot \mu(\emptyset) = \int 0 \; d\mu = 0$; given a sequence $(A_n)_{n \in \mathbb{N}}$ of disjoint events, then MON implies

$$p \cdot \mu(\cup_{n \in \mathbb{N}} A_n) = \int \mathbf{1}_{\cup_{n \in \mathbb{N}} A_n} p \; d\mu = \int \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n} \; d\mu = \sum_{n \in \mathbb{N}} \int \mathbf{1}_{A_n} \; d\mu = \sum_{n \in \mathbb{N}} p \cdot \mu(A_n)$$

Exercise: If $f \colon S \to \mathbb{R}$ is measurable and $fp \in \mathcal{L}^1(S, \mathcal{F}, \mu)$, then $f \in \mathcal{L}^1(S, \mathcal{F}, p \cdot \mu)$ and $\int f \; d(p \cdot \mu) = \int fp \; d\mu$. [Hint: try first simple functions, then use MON]

Inequalities I

- Expectation is a positive operator, hence it preserves the order. Most common application is a family of inequalities whose simplest form is Markov inequality: If $x \geq 0$ and $a > 0$, then $\mathbf{1}_{[a,+\infty[} \leq a^{-1}x$. It follows that for each non-negative random variable X we have $\mathrm{P}(X \geq a) \leq a^{-1}\,\mathbb{E}(X)$.
- The previous inequality can be optimised to get, for example, the exponential Markov inequality. Observe that for all $t > 0$ it holds $\{X \geq a\} = \left\{\mathrm{e}^{tX} \geq \mathrm{e}^{ta}\right\}$. It follows that

$$\mathrm{P}(X \geq a) \leq \mathrm{e}^{-ta}\,\mathbb{E}\left(\mathrm{e}^{tX}\right) = \exp\left(-\left(ta - \log\mathbb{E}\left(\mathrm{e}^{tX}\right)\right)\right).$$

  If $I(a) = \sup_{t>0}\left(ta - \log\mathbb{E}\left(\mathrm{e}^{tX}\right)\right)$, then $\log\mathrm{P}(X \geq a) \leq -I(a)$.
- Jensen inequality: Let $\Phi\colon \mathbb{R} \to \overline{\mathbb{R}}$ be convex with proper domain D. Assume X is an integrable random variable such that $\Phi(X)$ is integrable. Then $\Phi(\mathbb{E}(X)) \leq \mathbb{E}(\Phi \circ X)$. In fact, for each $x_o \in D$ there is an affine function such that $\Phi(x_0) + b(x_o)(x - x_0) \leq \Phi(x)$, $x \in \mathbb{R}$. It follows that $\Phi(x_0) + b(x_o)(\mathbb{E}(X) - x_0) \leq \mathbb{E}(\Phi \circ X)$. In particular, Jensen inequality follows if $x_0 = \mathbb{E}(X)$.

Inequalities II

- Fenchel's inequality: Given the convex function Φ, there exists a convex function Ψ such that $\Psi(y) = \sup_x (xy - \Phi(x))$. In particular, the inequality $xy \leq \Phi(x) + \Psi(y)$ holds for all x, y. It follows that $\mathbb{E}(XY) \leq \mathbb{E}(\Phi \circ X) + \mathbb{E}(\Psi \circ Y)$ if all terms are well defined.

- An important example of Fenchel inequality follows from $xy \leq \frac{1}{\alpha}|x|^\alpha + \frac{1}{\beta}|y|^\beta$, where $\alpha, \beta > 1$ and $\alpha^{-1} + \beta^{-1} = 1$. The integral inequality is $\mathbb{E}(XY) \leq \frac{1}{\alpha}\mathbb{E}(|X|^\alpha) + \frac{1}{\beta}\mathbb{E}\left(|Y|^\beta\right)$.

- For $x \in \mathbb{R}$, $q > 0$ and have $xq \leq \mathrm{e}^x - 1 + q\log q$. If f is a random variable and q is a probability density w.r.t. μ, then $\int fq\ d\mu \leq \int \mathrm{e}^f\ d\mu - 1 + \int q \log q\ d\mu$.

- Lebesgue space: For each $\alpha \geq 1$, define

$$\mathcal{L}^\alpha = \{X \in \mathcal{L} | \mathbb{E}(|X|^\alpha) < \infty\} \ .$$

Define

$$X \mapsto (\mathbb{E}(|X|^\alpha))^{1/\alpha} = \|X\|_\alpha$$

.

Inequalities III

- Hölder inequality. Apply Fenchel inequality to $f = X / \|X\|_\alpha$ and $g = Y / \|Y\|_\beta$. It follows

$$\mathbb{E}(fg) = \mathbb{E}\left(\frac{X}{\|X\|_\alpha}\frac{Y}{\|Y\|_\beta}\right) \leq \frac{1}{\alpha} + \frac{1}{\beta} = 1.$$

  It follows that $\mathbb{E}(XY) \leq \|X\|_\alpha \|Y\|_\beta$.

- Minkowski inequality. Apply Hölder inequality to

$$\mathbb{E}(|X+Y|^\alpha) = \mathbb{E}\left(|X+Y|\,|X+Y|^{\alpha-1}\right) \leq$$
$$\mathbb{E}\left(|X|\,|X+Y|^{\alpha-1}\right) + \mathbb{E}\left(|Y|\,|X+Y|^{\alpha-1}\right)$$

  to get $\|X+Y\|_\alpha \leq \|X\|_\alpha + \|Y\|_\alpha$.

Change of variable formula

Let be given a measure space $(S, \mathcal{F}, \mu)$, a measurable space $(\mathbb{X}, \mathcal{G})$ and a measurable mapping $\phi\colon S \to \mathbb{X}$, $p^{-1}\colon \mathcal{G} \to \mathcal{F}$. Let $\phi_{\#}\mu = \mu \circ \phi^{-1}$ be the push-forward measure.

- If $h \in \mathcal{S}(\mathbb{X}, \mathcal{G})$ i.e. $h = \sum_{k=1}^{n} b_k \mathbf{1}_{B_k}$, $b_k \in \mathbb{R}$, $B_k \in \mathcal{G}$, $k = 1, \dots, n$. Then

$$\int h \, d\phi_{\#}\mu = \sum_{k=1}^{n} b_k \phi_{\#}\mu(B_k) = \sum_{k=1}^{n} b_k \mu[\phi^{-1}(B_k)] =$$
$$\sum_{k=1}^{n} b_k \int \mathbf{1}_{B_k} \circ \phi \, d\mu = \int h \circ \phi \, d\mu$$

- If $f \in \mathcal{L}_+$, then MON implies $\int f \, d\phi_{\#}\mu = \int f \circ \phi \, d\mu$
- If $f : \mathbb{X} \to \mathbb{R}$ is measurable and $f \circ \phi \in \mathcal{L}^1(S, \mathcal{F}, \mu)$ then $f \in \mathcal{L}^1(\mathbb{X}, \mathcal{G}, \phi_{\#}\mu)$ and $\int f \, d\phi_{\#}\mu = \int f \circ \phi \, d\mu$.

Product measure I

Definition

Given measure spaces $(S_i, \mathcal{F}_i, \mu_i)$, $i = 1, \ldots, n$, $n = 2, 3, \ldots$, the product measure space is

$$(S, \mathcal{F}, \mu) = \otimes_{i=1}^{n}(S_i, \mathcal{F}_i, \mu_i) = (\times_{i=1}^{n} S_i, \otimes_{i=1}^{n} \mathcal{F}_i, \otimes_{i=1}^{n} \mu_i),$$

where

$$\mathcal{F} = \otimes_{i=1}^{n} \mathcal{F}_i = \sigma\left\{\times_{i=1}^{n} A_i | A_i \in \mathcal{F}_i, i = 1, \ldots, n\right\}$$

and $\mu = \otimes_{i=1}^{n} \mu_i$ is the unique measure on $(\times_{i=1}^{n} S_i, \otimes_{i=1}^{n} \mathcal{F}_i)$ such that

$$\mu(\times_{i=1}^{n} A_i) = \prod_{i=1}^{n} \mu_i(A_i), \quad A_i \in \mathcal{F}_i, i = 1, \ldots, n.$$

- Let $X_i \colon S \mapsto S_i$, $i = 1, \ldots, n$, be the projections. Then $\otimes_{i=1}^{n} \mathcal{F}_i = \sigma\left\{X_i | i = 1, \ldots, n\right\}$.
- Examples: Counting measure on $\mathbb{N}^2$, Lebesgue measure on $\mathbb{R}^2$, the finite Bernoulli scheme.
- Product measure of probability measures is a probability measure.

Product measure II

Recall all measures are σ-finite. Assume $n = 2$.

Sections

If $C \in= \mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, then for each $x_1 \in S_1$ the set $\{x_2 \in S_2 | (x_1, x_2) \in C\}$ belongs to $\mathcal{F}_2$.

Proof.

Let $\mathcal{O}$ be the family of all subsets of S for which the proposition is true. $\mathcal{O}$ is a σ-algebra that contains all the measurable rectangles, hence $\mathcal{F} \subset \mathcal{O}$. □

Partial integration

The mapping $S_1 \colon x_1 \mapsto \mu_2 \{x_2 \in S_2 | (x_1, x_2) \in C\}$ is non-negative and $\mathcal{F}_1$-measurable.

Proof.

If $C \in \mathcal{F}$ then the function is well defined. The set of all $C \in \mathcal{F}$ such that the function is measurable contains measurable rectangles, is a π-system, and is a d-system. □

Product measure III

Product measure: existence

The set function $\mu\colon \mathcal{F} \ni C \mapsto \int \mu_2 \{x_2 \in S_2 | (x_1, x_2) \in C\} \ \mu_1(dx - 1)$ is a measure such that $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ on measurable rectangles. Hence, $\mu = \mu_1 \otimes \mu_2$.

Proof.

The integral exists because the integrand is non-negative. $\mu(\emptyset) = 0$; if $(C_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{F}$ of disjoint events, then for all $x_1 \in S_1$ we have

$$\mu_2 \{x_2 \in S_2 | (x_1, x_2) \in \cup_{n\in\mathbb{N}} C_n\} = \mu_2 \left(\cup_{n\in\mathbb{N}} \{x_2 \in S_2 | (x_1, x_2) \in C_n\}\right) = \sum_{n\in\mathbb{N}} \mu_2 \{x_2 \in S_2 | (x_1, x_2) \in C_n\}$$

MON implies

$$\mu\left(\cup_{n\in\mathbb{N}} C_n\right) = \int \sum_{n\in\mathbb{N}} \mu_2 \{x_2 \in S_2 | (x_1, x_2) \in C_n\} \ \mu_1(dx_1) = \sum_{n\in\mathbb{N}} \int \mu_2 \{x_2 \in S_2 | (x_1, x_2) \in C_n\} \ \mu_1(dx_1) = \sum_{n\in\mathbb{N}} \mu(C_n)$$

Product measure IV

- Consider $n = 3$. The product measure space

$$\otimes_{i=1}^{3}(S_i, \mathcal{F}_i, \mu_i) = (\times_{i=1}^{n} S_i, \otimes_{i=1}^{n} \mathcal{F}_i, \otimes_{i=1}^{n} \mu_i)$$

  is identified with

$$(S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2) \otimes (S_3, \mathcal{F}_3, (\mu_1 \otimes \mu_2) \otimes \mu_3)$$

  One has to check that

$$(\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3 = \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \mathcal{F}_3$$

- The $n = \infty$ case requires Charateodory. See the Bernoulli scheme example.

Fubini theorem I

Section

Let $f : S_1 \times S_2 \to$ *reals* be $\mathcal{F}_1 \otimes \mathcal{F}_2$ measurable. For all $x_1 \in S_1$ the function $f_{x_1} : x_2 \mapsto f(x_1, x_2)$ is $\mathcal{F}_2$-measurable.

Proof.

For each $y \in \mathbb{R}$, consider the level set $C = \{(x_1, x_2) | f(x_1, x_2) \leq y\} \in \mathcal{F}_1 \otimes \mathcal{F}_2$. The set $\{(x_2) | f x_1(x_2) \leq y\}$ is the x_1-section of C. □

Theorem (Non-negative integrand)

Let $f : S_1 \times S_2 \to \mathbb{R}$ be $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable and non-negative. Then the mapping $S_1 \ni x_1 \mapsto \int f(x_1, x_2)\ \mu_2(dx_2)$ is $\mathcal{F}_1$-measurable and

$$\int f\ d\mu_1 \otimes \mu_2 = \int \left(\int f(x_1, x_2)\ \mu_2(dx_2) \right)\ \mu_1(dx_1)$$

Fubini theorem II

Theorem (Integrable integrand)

Let $f : S_1 \times S_2 \to \mathbb{R}$ be $\mu_1 \otimes \mu_2$-integrable. Then the mapping $S_1 \ni x_1 \mapsto \int f(x_1, x_2)\ \mu_2(dx_2)$ is μ_1-integrable and

$$\int f\ d\mu_1 \otimes \mu_2 = \int \left(\int f(x_1, x_2)\ \mu_2(dx_2) \right)\ \mu_1(dx_1)$$

Proof: Non-negative integrand.

Choose an increasing sequence of simple non-negative functions converging to f and use MON. □

Proof: Integrable integrand.

Decompose $f = f^+ - f^-$ and use the previous form of the theorem. □

Independence

Definition

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

1. The sub-σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if $A_i \in \mathcal{F}_i$, $i = 1, \ldots, n$, implies $\mu(A_1 \cap \cdots \cap A_n) = \mu(A_1) \cdots \mu(A_n)$.
2. The random variables $X_i \colon \Omega \to S_i$, $X_i^{-1} \colon \mathcal{G}_i \to \mathcal{F}$, $i = 1, \ldots, n$, are independent, if

$$(X_1, \ldots, X_n)_\# \mu = (X_1)_\# \mu \otimes \cdots \otimes (X_n)_\# \mu$$

If $\mathcal{F}_i = \sigma(X_i)$, the 1. and 2. are equivalent. If $A_i = X_i^{-1}(B_i)$, $i = 1, \ldots, n$,

$$\mu(A_1 \cap \cdots \cap A_n) = \mu(X_1^{-1}(B_1) \cap \cdots \cap X_n^{-1}(B_n)) =$$
$$\mu((X_1, \ldots, X_n)^{-1}(B_1 \times \cdots \times B_n)) = (X_1, \ldots, X_n)_\# \mu(B_1 \times \cdots \times B_n) =$$
$$(X_1)_\# \mu \otimes \cdots \otimes (X_n)_\# \mu(B_1 \times \cdots \times B_n) = (X_1)_\# \mu(B_1) \cdots (X_n)_\# \mu(B_n) =$$
$$\mu(X_1^{-1}(B_1)) \cdots \mu(X_n^{-1}(B_n)) = \mu(A_1) \cdots \mu(A_n)$$