Probability 2019
Measure Theory

Giovanni Pistone
www.giannidiorestino.it

DE CASTRO STATISTICS Collegio Carlo Alberto

Version March 25, 2019

Measurable space

Definition

- A family $\mathcal{B}$ of subsets of S is an field on S if it contains $\emptyset$ and S, and it is stable for the complements, finite unions, and finite intersection.
- A family $\mathcal{F}$ of subsets of S is a σ-field on S if it is an field on S and it is stable for denumerable unions and intersections.
- A measurable space is a couple $(S, \mathcal{F})$, where S is a set and $\mathcal{F}$ is a σ-field on S.

- Given the family $\mathcal{C}$ of subsets of S, the σ-field generated by $\mathcal{C}$ is $\sigma(\mathcal{C}) = \cap \{\mathcal{A} | \mathcal{C} \subset \mathcal{A} \text{ and } \mathcal{A} \text{ is a } \sigma\text{-field}\}$.
- Examples: the field generated by a finite partition; the Borel σ-field of $\mathbb{R}$ is generated by the open intervals, or by the closed intervals, or by the intervals, or by the open sets, or by semi-infinite intervals.

§1.1 of D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

Measure space

Definition

- A measure μ of the measurable space $(S, \mathcal{F})$ is a mapping $\mu\colon \mathcal{F} \to [0, +\infty]$ such that $\mu(\emptyset) = 0$ and for each sequence $(A_n)_{n\in\mathbb{N}}$ of disjoint elements of $\mathcal{F}$, $\mu(\cup_{n\in\mathbb{N}} A_n) = \sum_{i=1}^{\infty} \mu(A_n)$.
- A measure is finite if $\mu(S) < +\infty$; a measure is σ-finite if there is a sequence $(S_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ such that $\cup_{n\in\mathbb{N}} S_n = S$ and $\mu(S_n) < +\infty$ for all $n \in \mathbb{N}$.
- A probability measure is a finite measure such that $\mu(S) = 1$; a probability space is the triple $(S, \mathcal{F}, \mu)$, where μ is a probability measure.

- Examples: probability measure on a partition; probability measure on a denumerable set.
- Equivalently, a probability measure is finitely additive and sequentially continuous at $\emptyset$

§1.3-5 of Williams

Product system aka π-system

Definition

Let S be a set. A π-system on S is a family $\mathcal{I}$ of subsets of S which is stable under finite intersection.

- Examples: the family of all points of a finite set and the empty set; the family of open intervals of $\mathbb{R}$; the familily of closed intervals of $\mathbb{R}$; the family of cadlàg intervals of $\mathbb{R}$; the family of convex (resp. open convex, closed convex) subsets of $\mathbb{R}^2$; the family of open (resp. closed) set in a topological space.
- If $\mathcal{I}_i$ is a π-system of S_i, $i = 1, \dots, n$, then $\{\times_{i=1}^n I_i | I_i \in \mathcal{I}_i\}$ is a π-system of $\times_{i=1}^n S_i$.
- The family of all real functions of the form $\alpha_0 + \sum_{j=1}^n \alpha_j \mathbf{1}_{I_i}$, $n \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, $j = 0, \dots, n$ is a vector space and it is stable for multiplication.

§1.6 of Williams

Dynkin system aka d-system

Definition

Let S be a set. A *d-system* on S is a family $\mathcal{D}$ of subsets of S such that

1. $S \in \mathcal{D}$
2. If $A, B \in \mathcal{D}$ and $A \subset B$, then $B \setminus A \in \mathcal{D}$. (Notice that $S \setminus A = A^c$)
3. If $(A_n)_{n\in\mathbb{N}}$ is an increasing sequence in $\mathcal{D}$, then $\cup_{n\in\mathbb{N}} \in \mathcal{D}$

- Given probabilities μ_i and $i = 1, 2$ on the measurable space $(S, \mathcal{F})$, the family $\mathcal{D} = \{A \in \mathcal{F} | \mu_1(A) = \mu_2(A)\}$ in a d-system.
- Given measurable spaces $(S_i, \mathcal{F}_i)$, $i = 1, 2$, the product space $(S, \mathcal{F}) = (S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \{A_1 \times A_2 | A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$, and $x \in S_1$, the family $\mathcal{D} = \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 | A \cap \{x\} \times S_2 = \{x\} \times A_x, A_x \in \mathcal{F}_2\}$ is a d-system.

§A1.2 of Williams

Dynkin's lemma

Theorem

1. *A family of subsets of S is a σ-field if, and only if, it is both a d-system and a π-system.*
2. *If $\mathcal{I}$ is a π-system, then $d(\mathcal{I}) = \sigma(\mathcal{I})$.*
3. *Any d-system that contains a π-system contains the σ-field generated by the π-system.*

Theorem

If two probability measures on the same measurable space agree on a π-system $\mathcal{I}$ they are equal on $\sigma(\mathcal{I})$.

§A1.3 of Williams

Probability space

Definition

A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ of a sample space Ω (set of possible worlds), a σ-field $\mathcal{F}$ on Ω, a probability measure $\mathbb{P}\colon \mathcal{F} \to [0,1]$. An element $\omega \in \Omega$ is a sample point (world); an element $A \in \mathcal{F}$ is an event; the value $\mathbb{P}(A)$ is the probability of the event A.

- Examples: a finite set, all its subsets, a probability function $p\colon \Omega \to \mathbb{R}_{>0}$ such that $\sum_{\omega\in\Omega} p(\omega) = 1$; $\mathbb{Z}_{\geq}$ with all its subsets, and a probability function $p\colon \mathbb{Z}_{\geq} \to \mathbb{R}_{>0}$ such that $\sum_{k=0}^{\infty} p(k) = 1$; the restriction of a probability space to a sub-σ-field; the product of two probability spaces.
- Bernoulli trials. Let $\Omega = \{0,1\}^{\mathbb{N}}$ and let $\mathcal{F}_n = \left\{A \times \{0,1\} \times \{0,1\} \times \cdots \middle| A \subset \{0,1\}^n\right\}$, $\mathcal{F} = \sigma(\mathcal{F}_n\colon n \in \mathbb{N})$. Given $\theta \in [0,1]$, the function $p_n(x_1x_2\cdots x_n\cdots) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$ uniquely defines probability spaces $(\Omega, \mathcal{F}_n, \mathbb{P}_n)$, $n \in \mathbb{N}$, such that $\mathbb{P}_{n+1}|_{\mathcal{F}_n} = \mathbb{P}_n$, hence a probability measure $\mathbb{P}$ on $\mathcal{F}$.

$\limsup$ and $\liminf$

Definition

- Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers.

$$\limsup_{n\to\infty} a_n = \wedge_{m\in\mathbb{N}} \vee_{n\geq m} a_n \quad \text{(maximum limit)}$$

$$\liminf_{n\to\infty} a_n = \vee_{m\in\mathbb{N}} \wedge_{n\geq m} a_n \quad \text{(minimum limit)}$$

- Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of events in the measurable space $(\Omega, \mathcal{F})$.

$$\limsup_{n\to\infty} E_n = \cap_{m\in\mathbb{N}} \cup_{n\geq m} E_n \quad (E_n \text{ infinitely often})$$

$$\liminf_{n\to\infty} E_n = \cup_{m\in\mathbb{N}} \cap_{n\geq m} E_n \quad (E_n \text{ eventually})$$

A similar definition applies to sequences of functions. If $(f_n)_n$ is a sequence of non-negative functions, then the set of $x \in S$ such that $\lim_n f_n(x) = 0$ is equal to the set $\{\limsup_n f_n = 0\}$.

Fatou lemma

Theorem

$$\mathbb{P}\left(\liminf_{n\to\infty} E_n\right) \le \liminf_{n\to\infty} \mathbb{P}\left(E_n\right) \le \limsup_{n\to\infty} \mathbb{P}\left(E_n\right) \le \mathbb{P}\left(\limsup_{n\to\infty} E_n\right)$$

- $(\limsup_n E_n)^c = \liminf_n E_n^c$; $\limsup_n \mathbf{1}_{E_n} = \mathbf{1}_{\limsup_n E_n}$.
- Proof of FL. Write $\cup_m \cap_{n\ge m} E_n = \cup_m G_m$ so that $G_m \uparrow G = \liminf_n E_n$. We have $\mathbb{P}\left(G_m\right) \le \wedge_{n\ge m} \mathbb{P}\left(E_n\right)$; monotone continuity (increasing) implies $\mathbb{P}\left(G_m\right) \uparrow \mathbb{P}\left(G\right)$ hence, $\vee_m \mathbb{P}\left(G_m\right) = \mathbb{P}\left(G\right)$. The middle inequality is a property of $\liminf$ and $\limsup$. The least inequality follows from a similar proof using continuity on decreasing sequences or, by taking the complements.
- BC1. Assume $\sum_{n=1}^{\infty} \mathbb{P}\left(E_n\right) < +\infty$. We have for all $m \in \mathbb{N}$ that

$$\mathbb{P}\left(\limsup_n E_n\right) \le \mathbb{P}\left(\cup_{n\ge m} E_n\right) \le \sum_{n=m}^{\infty} \mathbb{P}\left(E_n\right) \to 0 \quad \text{if } m \to \infty$$

hence $\mathbb{P}\left(\limsup_n E_n\right) = 0$.

Measurable function

Definition

Given measurable spaces $(S_i, \mathcal{S}_i)$, $i = 1, 2$, we say that the function $h \colon S_1 \to S_2$ is measurable, or is a random variable, if for all $B \in \mathcal{S}_2$ the set $h^{-1}(B) = \{s \in S_1 | h(s) \in B\}$ belongs into $\mathcal{S}_1$.

Theorem

- *Let $\mathcal{C} \subset \mathcal{S}_2$ and $\sigma(C) = \mathcal{S}_2$. If $h^{-1} \colon \mathcal{C} \to \mathcal{S}_1$, then h is measurable.*
- *Given measurable spaces $(S_i, \mathcal{S}_i)$, $i = 1, 2, 3$, if both $h \colon S_1 \to S_2$, $g : S_2 \to S_3$ are measurable functions, then $g \circ f : S_1 \to S_3$ is a measurable function.*
- *Given measurable spaces $(S_i, \mathcal{S}_i)$, $i = 0, 1, 2$ and $h_i \colon S_0 \to S_j$, $j = 1, 2$, consider $h = (h_1, h_2) \colon S_0 \to S_1 \times S_2$. with product space $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$, Then both h_1 and h_2 are measurable if, and only if, h is measurable.*

- Ch. 3 of Ch. 2 of Williams

Image measure

Definition

Given measurable spaces $(S_i, \mathcal{S}_i)$, $i = 1, 2$, a measurable function $h\colon S_1 \to S_2$, and a measure μ_1 on $(S_1, \mathcal{S}_1)$, then $\mu_2 = \mu_1 \circ h^{-1}$ is a measure on $(S_2, \mathcal{S}_2)$. We write $h_\# \mu_1 = \mu_2 \circ h^{-1}$ and call it image measure. If μ_1 is a probability measure, we say that $h_\# \mu_1$ is the distribution of the random variable h.

- Bernoulli scheme Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the Bernoulli scheme, and define $X_t\colon \Omega \to \{0, 1\}$ to be the t-projection, $X_t(x_1 x_2 \cdots) = x_t$. It is a random variable with Bernoulli distribution $\mathrm{B}(\theta)$. The random variable $Y_n = X_1 + \cdots + X_n$ has distribution $\mathrm{Bin}(\theta, n)$. The random variable $T = \inf \{k \in \mathbb{N} | X_k = 1\}$ has distribution $\mathrm{Geo}(\theta)$.

Real random variable

Definition

Let $(S, \mathcal{S})$ be a measurable space. A real random variable is a real function $h \colon S \to \mathbb{R}$ with is measurable into $(\mathbb{R}, \mathcal{B})$.

Theorem

- *$h \colon S \to \mathbb{R}$ is a real random variable if, and only if, for all $c \in \mathbb{R}$ the level set $\{s \in S\}\, h(s) \leq c$ is measurable. The same property holds with $\leq$ replaced by $<$ or $\geq$ or $>$. The condition can be taken as a defintion of extended random variable i.e.*
  $h \colon S \to \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.
- *If $g, h \colon S \to \mathbb{R}$ are real random variables and $\Phi \colon \mathbb{R}^2 \to \mathbb{R}$ is continuous, then $\Phi \circ (g, h)$ is a real random variable.*
- *Let $(h_n)_{n \in \mathbb{N}}$ be a sequence of real random variables on $(S, \mathcal{S})$. Then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ are real random variable.*

A monotone-class theorems

Theorem

Let $\mathcal{H}$ be a vector space of bounded real functions of a set S and assume $\mathbf{1} \in \mathcal{H}$. Assume

1. *$\mathcal{H}$ is a monotone class i.e., if for each bounded increasing sequence $(f_n)_n \in \mathbb{N}$ in $\mathcal{H}$ the function $\vee_n f_n$ belong to $\mathcal{H}$.*
2. *$\mathcal{H}$ contains the indicator functions of a π-system $\mathcal{I}$.*

Then, $\mathcal{H}$ contains all bounded measurable functions of $(S, \sigma(I))$.

- Application. Consider measurable spaces $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$. Define $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{I} = \{A_1 \times A_2 | A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$. Then $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{I})$. Let $\mathcal{H}$ be the set of all bounded real funtions $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ such that for each fixed $x \in \Omega_1$ the mapping $\Omega_2 \ni y \mapsto f(x, y)$ is $\mathcal{F}_2$-measurable and for each fixed $y \in \Omega_2$ the mapping $\Omega_1 \ni x \mapsto f(x, y)$ is $\mathcal{F}_1$-measurable.

- §3.14 and §A3.1 of Williams

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 Williams

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 $f \colon S \to [0, +\infty]$ 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 Williams

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

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

- Exercise. Revise L^1-convergence and Dominated Convergence Theorem in Ch 5 of Williams

Expectation

- Let $\mathrm{E}\colon \mathcal{L}^\infty(S, \mathcal{S}) \to \mathbb{R}$ be such that
  - $\mathbb{E}(\mathbf{1}) = 1$.
  - E is linear and positive (hence monotone).
  - E is continuous on non-increasing sequence converging to 0.
- Every such E defines a probability measure when restricted to indicators, $\mathbb{P}(A) = \mathbb{E}(\mathbf{1}_A)$ and $\mathbb{E}(f) = \int f \ d\,\mathbb{P}$
- A similar observation holds for a $\mathrm{E}\colon \mathcal{L}_+(S, \mathcal{S})$.
- 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 $f \in \mathcal{L}^1(S, \mathcal{S}, \mathbb{P})$ and define $\mathbb{E}(f) = \mathbb{E}(f_+) - \mathbb{E}(f_-)$.
- $\mathrm{E}\colon \mathcal{L}^1(S, \mathcal{S}, \mathbb{P}) \to \mathbb{R}$ is positive, linear, normalized, continuous for the bounded pointwise convergence.
- Exercise: 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 $\mathbb{P}(X \geq a) \leq a^{-1}\,\mathbb{E}(X)$.
- 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

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

  If $I(a) = \sup_{t>0}\left(ta - \log \mathbb{E}\left(\mathrm{e}^{tX}\right)\right)$, then $\log \mathbb{P}(X \geq a) \leq -I(a)$.
- 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

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

  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 $f: S_1 \times S_2 \to \mathbb{R}$ 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. $\square$

Proof: Integrable integrand.

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

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 $A_i = X_i^{-1}(B_i)$, $i = 1, \ldots, n$,

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