Probability Theory

Teacher: Patrícia Gonçalves

The classes

Our classes have the following schedule:

- Wednesday from 9h to 11h at 4.35 (maths building).
- Thursday from 11h to 13h at P1 (maths building).

My office is in the maths building, 4th floor, the office number is 4.06.

My email is patricia.goncalves@math.tecnico.ulisboa.pt.

Office hours are on Wednesdays 11h-12h: please send me an email up the day before saying that you will show up.

The bibliography of the course

- Principal bibliography:
  Kai Lai Chung "A course in probability theory".
- Barry James: "Probabilidade: um curso em nível intermediário".
- Richard Durrett: "Probability: Theory and Examples".
- Patrick Billingsley: "Probability and measure".
- Sheldon Ross: "Probability".

The program of the course

- Set theory: framework and basic results;
  Probability measures and distribution functions;
- Random variables and random vectors;
  Expectation;
  Stochastic independence;
- Types of convergence: almost surely, in $\mathbb{L}^p$, in probability, in distribution;
  Weak and strong law of large numbers;
- Characteristic functions: framework and basic results;
  Central limit theorem: the convergence to the Gaussian/Poisson;
- Conditional expectation;
  Discrete time martingales: examples and main results;
  Applications.

Evaluation

The final mark is computed according to the rule

$$\frac{1}{3}T_{exam} + \frac{1}{3}T_{exercises} + \frac{1}{3}T_{presentation}, \tag{1}$$

where T_{exam} is the mark obtained in the exam, $T_{exercises}$ is the mark obtained in the exercise lists and $T_{presentation}$ is the mark obtained on the oral presentation (the date will be discussed).

- T_1 - 30th november at the class.

- The minimum mark on the exam has to be greater or equal to 8, otherwise the student can only approve by doing the final exam on the 3rd february from 11:30h to 14:30h.

- If the final mark obtained from (1) (or in the final exam) is greater or equal to 17, the student has to do an oral exam, otherwise the student gets the score 17.

1st Lecture: Set theory

Classes of sets

Let Ω be an abstract space and we shall denote its elements by ω.

Definition (Algebra)

An non-empty collection $\mathcal{F}$ of subsets of Ω is an algebra if and only if:

- $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$
- $E_1, E_2 \in \mathcal{F} \Rightarrow E_1 \cup E_2 \in \mathcal{F}$

Definition (Monotone class)

An non-empty collection $\mathcal{F}$ of subsets of Ω is a monotone class if and only if:

- $E_j \in \mathcal{F}\,, E_j \subset E_{j+1}\,, \forall j \Rightarrow \cup_{j\geq 1} E_j \in \mathcal{F}$
- $E_j \in \mathcal{F}\,, E_j \supset E_{j+1}\,, \forall j \Rightarrow \cap_{j\geq 1} E_j \in \mathcal{F}$

Definition (σ-algebra)

An non-empty collection $\mathcal{F}$ of subsets of Ω is a σ-algebra if and only if:

- $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$
- $E_j \in \mathcal{F}\,, \forall j \Rightarrow \cup_{j\geq 1} E_j \in \mathcal{F}$

Theorem

An algebra is a σ-algebra if and only if it is a monotone class.

Exercise: do the proof of the theorem.

Example

The collection $\mathcal{S}$ of all subsets of Ω is a σ-algebra and it is called the total σ-algebra. The collection $\{\varnothing, \Omega\}$ is a σ-algebra and is it called the trivial σ-algebra.

Remark

1. *If A is an index set and if for $\alpha \in A$, $\mathcal{F}_\alpha$ is a σ-algebra (or a monotone class), then $\cap_{\alpha \in A} \mathcal{F}_\alpha$ is a σ-algebra (or a monotone class).*
2. *Given a non empty collection of sets $\mathcal{C}$, there exists a minimal σ-algebra (or algebra or monotone class) containing $\mathcal{C}$, which consists in the intersection of all σ-algebras (or algebras or monotone classes) containing $\mathcal{C}$. There is at least one, namely the total σ-algebra $\mathcal{S}$. This σ-algebra (or algebra or monotone class) is called the σ-algebra generated by $\mathcal{C}$.*

Theorem

Let $\mathcal{F}_0$ be an algebra, $\mathcal{C}$ the minimal monotone class containing $\mathcal{F}_0$ and $\mathcal{F}$ the minimal σ-algebra containing $\mathcal{F}_0$. Then $\mathcal{F} = \mathcal{C}$.

Exercise: do the proof of the theorem.

Probability measures

Probability measure

Definition

Let Ω be an abstract space and $\mathcal{F}$ a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}(\cdot)$ in $\mathcal{F}$ is a function $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ which satisfies the following properties:

1. $\forall E \in \mathcal{F}, \mathbb{P}(E) \geq 0$.
2. If $\{E_j\}_{j\geq 1}$ is a countable collection of disjoint sets of $\mathcal{F}$, then

$$\mathbb{P}(\cup_{j\geq 1} E_j) = \sum_{j\geq 1} \mathbb{P}(E_j) \quad \text{(countable additivity).}$$

3. $\mathbb{P}(\Omega) = 1$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space, Ω is called the sample space and its elements ω are called the sample points.

Exercise: Prove that, as a consequence of the previous definition, we have

1. $\forall E \in \mathcal{F}, \mathbb{P}(E) \leq 1.$
2. $\mathbb{P}(\varnothing) = 0.$
3. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E).$
4. $\mathbb{P}(E \cup F) + \mathbb{P}(E \cap F) = \mathbb{P}(E) + \mathbb{P}(F).$
5. $E \subset F \Rightarrow \mathbb{P}(E) = \mathbb{P}(F) - \mathbb{P}(F \backslash E) \leq \mathbb{P}(F).$
6. Monotone property: If $E_j \uparrow E$ or $E_j \downarrow E$, then $\mathbb{P}(E_j) \to \mathbb{P}(E).$
7. Boole's inequality $\mathbb{P}(\cup_{j\geq 1} E_j) \leq \sum_{j\geq 1} \mathbb{P}(E_j).$

Recall that

1. If $\{E_j\}_{j\geq 1}$ is a countable collection of disjoint sets of $\mathcal{F}$, then
$$\mathbb{P}(\cup_{j\geq 1} E_j) = \sum_{j\geq 1} \mathbb{P}(E_j) \quad \text{(countable additivity).}$$
2. When above we have a finite collection we say it is the finite additivity property.
3. If $E_j \downarrow \varnothing$, then $\mathbb{P}(E_j) \to 0 \quad$ (continuity).

Theorem

The finite additivity and the continuity together are equivalent to countable additivity.

Exercise: do the proof of the theorem.

Trace

Let $\Lambda \in \Omega$. The trace of the σ-algebra $\mathcal{F}$ in Λ is the collection of all the sets of the form $\Lambda \cap F$, where $F \in \mathcal{F}$. It is easy to see that this is a σ-algebra that we denote by $\Lambda \cap \mathcal{F}$.
Suppose now that $\Lambda \in \mathcal{F}$ and $\mathbb{P}(\Lambda) > 0$. Then we can define $\mathbb{P}_\Lambda$ in $\Lambda \cap \mathcal{F}$ in the following way: for any $E \in \Lambda \cap \mathcal{F}$:

$$\mathbb{P}_\Lambda(E) = \frac{\mathbb{P}(E)}{\mathbb{P}(\Lambda)}.$$

$\mathbb{P}_\Lambda$ is a probability measure in $\Lambda \cap \mathcal{F}$.

The triple $(\Lambda, \Lambda \cap \mathcal{F}, P_\Lambda)$ is called the trace of $(\Omega, \mathcal{F}, \mathbb{P})$ in Λ.

Examples:

Example (1. Discrete sample space)

Let $\Omega = \{w_j, j \in \mathbb{N}\}$ and let $\mathcal{F}$ be the total σ-algebra in $\mathcal{F}$. Choose a sequence of numbers $\{p_j, j \in \mathbb{N}\}$ such that for all $j \in \mathbb{N}$, $p_j \geq 0$ and $\sum_{j \in \mathbb{N}} p_j = 1$ and let $\mathbb{P} : \mathcal{F} \to [0,1]$ defined on $E \in \mathcal{F}$ by

$$\mathbb{P}(E) = \sum_{w_j \in E} p_j.$$

Show that $\mathbb{P}$ is a probability measure and that all the probability measures on $(\Omega, \mathcal{F})$ are of the form above.

Examples:

Example (2. Continuous sample spaces)

Let $\mathcal{U} = (0,1]$ and let $\mathcal{C} := \{(a,b] : 0 < a < b \leq 1\}$, $\mathcal{B}$ the minimal σ-algebra containing $\mathcal{C}$, m the Lebesgue measure on $\mathcal{B}$. Then $(\mathcal{U},\mathcal{B},m)$ is a probability space.

Analogously, consider in $\mathbb{R}$ the collection $\mathcal{C}$ of the intervals of the form $(a,b]$, $-\infty < a < b < +\infty$. The algebra $\mathcal{B}_0$ generated by $\mathcal{C}$ consists of finite unions of disjoint sets of the form $(a,b]$, $(-\infty,a]$ or $(b,+\infty)$. The Borel σ-algebra is the σ-algebra, denoted hereafter by $\mathcal{B}$, generated by $\mathcal{B}_0$ or by $\mathcal{C}$. Not that the Lebesgue measure m in $\mathbb{R}$ is NOT a probability measure.

2nd Lecture: Distribution functions

Distribution function

Definition

A distribution function $F : \mathbb{R} \longrightarrow \mathbb{R}$ is increasing, right continuous and satisfies $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Example

$F_1(x) = \mathbf{1}_{[0,+\infty)}(x)$.

$F_2(x) = \frac{1}{2}\mathbf{1}_{[0,\frac{1}{2})}(x) + \mathbf{1}_{[\frac{1}{2},+\infty)}$.

$F_3(x) = x\mathbf{1}_{[0,1)}(x) + \mathbf{1}_{[1,+\infty)}$.

$F_4(x) = x\mathbf{1}_{[0,\frac{1}{2})}(x) + \mathbf{1}_{[\frac{1}{2},+\infty)}$.

Pay attention to the graph of the functions.

Distribution function and probability measure

Lemma

Each probability measure μ *in* $\mathcal{B}$ *defines a distribution function* F *through the following correspondence:* $\forall x \in \mathbb{R}, \mu((-\infty, x]) = F(x).(^*)$

Exercise: do the proof of the lemma.

Remark

As a consequence we have for $-\infty < a < b < \infty$ *that*

- $\mu((a,b]) = F(b) - F(a);$ $\mu([a,b)) = F(b^-) - F(a^-);$
- $\mu((a,b)) = F(b^-) - F(a);$ $\mu([a,b]) = F(b) - F(a^-);$

For a dense subset D of $\mathbb{R}$ the correspondence above in $(^*)$ is determined for $x \in D$ or if in the previous equalities we take $a, b \in D$.

Theorem

Each distribution function F determines a probability measure μ in $\mathcal{B}$ through one of the correspondences given above.

The question now is: Is this probability measure μ unique?

Theorem

Let μ and ν be two probability measures defined in the same σ-algebra $\mathcal{F}$ generated by the algebra $\mathcal{F}_0$. If $\mu(E)=\nu(E)$ for any $E\in\mathcal{F}_0$ then $\mu=\nu$.

Exercise: do the proof of the two theorems above.

Theorem

Given a probability measure μ in $\mathcal{B}$ there exists a unique distribution function F which satisfies $\mu((-\infty, x]) = F(x)\ \forall x \in \mathbb{R}$. Conversely, given a distribution function F, there exists a unique probability measure μ in $\mathcal{B}$ satisfying $\mu((-\infty, x]) = F(x)\ \forall x \in \mathbb{R}$.

We shall call μ the probability measure of F and F the distribution function of μ.

Example

Instead of $(\mathbb{R},\mathcal{B})$ we can consider a restriction to a fixed interval $[a,b]$. As example take $\mathcal{U}=[0,1]$. Let us see how to define F.

Let F be a distribution function such that $F(x)=0$, if $x<0$ and $F(x)=1$, if $x\geq 1$.

The probability measure μ will have support $[0,1]$ since $\mu(-\infty,0)=0=\mu(1,+\infty)$.

The trace of $(\mathbb{R},\mathcal{B},\mu)$ in $\mathcal{U}$ can be denoted by $(\mathcal{U},\mathcal{B}_{\mathcal{U}},m)$, where $\mathcal{B}_{\mathcal{U}}$ is the trace of $\mathcal{B}$ in $\mathcal{U}$ and any probability measure in $\mathcal{B}_{\mathcal{U}}$ can be seen as that trace.

As example, we have the uniform distribution given for $F_3(\cdot)$ above.

Atoms

Definition

An atom of a measure μ defined in $\mathcal{B}$ is a singleton $\{x\}$ such that $\mu(\{x\}) > 0$.

Definition

A measure is said to be atomic if and only if μ is zero on any set not containing any atom.

Prove that if F is the distribution function of μ then

$$\mu(\{x\}) = F(x) - F(x^-).$$

Prove that μ is atomless (that is μ does not have atoms) if and only if F is continuous.

Monotone functions

Let f be an increasing function defined on $\mathbb{R}$. This means that for all $x \leq y$ it holds $f(x) \leq f(y)$. Let us see some properties of these kind of functions.

1. Both lateral limits exist and are finite for any $x \in \mathbb{R}$:

$$\lim_{y \downarrow x} f(y) = f(x^+) \quad \text{and} \quad \lim_{y \uparrow x} f(y) = f(x^-).$$

2. When $x = \pm\infty$ the limits above exist but can be equal to $\pm\infty$.
3. The function is continuous (resp. right-continuous) at x if and only if the limits above are both (resp. $f(x^+)$ is) equal to $f(x)$.
4. We say that the function has a jump at x if the limits above exist but are different. The value $f(x)$ has to satisfy

$$f(x^-) \leq f(x) \leq f(x^+.)$$

5. When there is a jump at x, we say that x is a point of jump of f and $f(x^+) - f(x^-)$ is the size of the jump.

The set of jumps of f is countable (can be finite.)

To prove this, first associate to each point of jump x, the interval $I_x = (f(x^-), f(x^+))$. Then, if x' is another point of jump of f and $x < x'$, then there exists $\tilde{x}$ such that $x < \tilde{x} < x'$ and

$$f(x^+) \leq f(\tilde{x}) \leq f(x'^-).$$

As a consequence the intervals I_x and $I_{x'}$ are disjoint and can be consecutive if $f(x^+) = f(x'^-)$. Therefore we associate to the set of points of jump of f a collection of disjoint intervals in the range of f. Now, this collection is, at most, countable since each interval contains a rational number, so that the collection of intervals is in one-to-one correspondence with a certain subset of the rational numbers, being the latter countable.

Since the set of points of jump of f is in one-to-one correspondence with the set of intervals associated with it, then the proof ends.

Example

Let $\{a_n\}_{n\geq 1}$ be any given enumeration of the rational numbers and let $\{b_n\}_{n\geq 1}$ be a sequence of non-negative real numbers such that $\sum_{n\geq 1} b_n < +\infty$. Consider $f(x) = \sum_{n\geq 1} b_n \delta_{a_n}(x)$ where for each $n \geq 1$ we have $\delta_{a_n}(x) = \mathbf{1}_{[a_n,+\infty)}(x)$ - the Heaviside function at a_n.
Since $0 \leq \delta_{a_n}(x) \leq 1$, the series above is absolutely and uniformly convergent.
Since $\delta_{a_n}(x)$ is increasing, then if $x_1 \leq x_2$ we have that $f(x_2) - f(x_1) = \sum_{n\geq 1} b_n(\delta_{a_n}(x_2) - \delta_{a_n}(x_1)) \geq 0$, so that f is increasing.
Then $f(x^+) - f(x^-) = \sum_{n\geq 1} b_n(\delta_{a_n}(x^+) - \delta_{a_n}(x^-))$. But for each $n \geq 1$, $\delta_{a_n}(x^+) - \delta_{a_n}(x^-)$ is zero or one if $x \neq a_n$ or $x = a_n$.
From this we conclude that f is discontinuous (jumps) in the rational numbers and nowhere else.

The previous example shows that the set of points of jump of an increasing function may be dense.

3rd Lecture: Random variables

Random variable

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

Definition

A function X with domain $\Lambda \in \mathcal{F}$ taking values in $\mathbb{R}^* := [-\infty, \infty]$ is a random variable if: $\forall B \in \mathcal{B}^*$ we have that

$$X^{-1}(B) \in \Lambda \cap \mathcal{F}, \tag{2}$$

where $\Lambda \cap \mathcal{F}$ is the trace of $\mathcal{F}$ in Λ, $X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in \mathcal{B}\}$ and $\mathcal{B}^*$ is the extended Borel σ-algebra, that is, its elements are sets in $\mathcal{B}$ with one or both $+\infty$, $-\infty$.

Remark

A random variable that takes values in the complex numbers is a function from $\Lambda \in \mathcal{F}$ to the complex plane whose real and imaginary parts are random variables taking finite values.

Random variable

From now on we assume that $\Lambda = \Omega$ and that X is real and takes finite values with probability one. The general case can be reduced to this one, considering the trace of $(\Omega, \mathcal{F}, \mathbb{P})$ in the set

$$\Lambda_0 := \{\omega \in \Omega : |X(\omega)| < \infty\}$$

and taking the real and imaginary parts of X.

Consider now the inverse application $X^{-1} : \mathbb{R} \longrightarrow \Omega$ defined on $A \subset \mathbb{R}$, by $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$. The condition (2) tells us that X^{-1} takes elements of $\mathcal{B}$ into elements of $\mathcal{F}$: $X^{-1}(\mathcal{B}) \in \mathcal{F}$. A function which satisfies this property is said to be measurable wrt $\mathcal{F}$. Therefore a random variable is a measurable function from Ω to $\mathbb{R}$ (or $\mathbb{R}^*$).

Theorem

For each function $X:\Omega\longrightarrow\mathbb{R}$ *(or* $\mathbb{R}^*$*), the inverse application* X^{-1} *satisfies the following properties:*

- $X^{-1}(A^c)=(X^{-1}(A))^c$,
- $X^{-1}(\cup_\alpha A_\alpha)=\cup_\alpha X^{-1}(A_\alpha)$,
- $X^{-1}(\cap_\alpha A_\alpha)=\cap_\alpha X^{-1}(A_\alpha)$

where α *belongs to an index set not necessarily countable.*

Theorem

X *is a random variable if and only if* $\forall x\in\mathbb{R}$ *(or* x *in a dense subset of* $\mathbb{R}$*) we have* $\{\omega\in\Omega : X(\omega)\leq x\}\in\mathcal{F}$.

In this case since $\mathbb{P}$ is defined in $\mathcal{F}$ we denote the probability wrt $\mathbb{P}$ of the set $\{\omega\in\Omega : X(\omega)\in B\}$ simply by $\mathbb{P}(X\in B)$, for $B\in\mathcal{B}$.

Theorem

Each random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ induces a probability space $(\mathbb{R}, \mathcal{B}, \mu)$ through the following correspondence $\forall B \in \mathcal{B}$ $\mu(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$.

Remark

1. The collection of sets $\{X^{-1}(S); S \in \mathbb{R}\}$ is a σ-algebra for any function X.

2. In case X is a random variable, the collection $\{X^{-1}(B); B \in \mathcal{B}\}$ is the σ-algebra generated by X, which consists in the smallest sub σ-algebra of $\mathcal{F}$ which contains all the sets of the form $\{\omega \in \Omega : X(\omega) \leq x\}$ with $x \in \mathbb{R}$.

3. The measure μ is going to be denoted by $\mu := \mathbb{P} o X^{-1}$ and it is called the probability distribution measure of X and its associated F is the distribution function of X: $F(x) = \mu((-\infty, x]) = \mathbb{P}(X \leq x)$.

Identically distributed random variables

Note that X determines μ and μ determines F, the converse is false. Two random variables which have the same distribution are said to be identically distributed.

Example

Consider the probability space $(\mathcal{U}, \mathcal{B}, m)$, $\mathcal{U} = [0,1]$, $\mathcal{B}$ is the Borel σ-algebra in $\mathcal{U}$ and m is the Lebesgue measure; and the random variables $X_i : \mathcal{U} \longrightarrow \mathcal{U}$ given by $X_1(\omega) = \omega$ and $X_2(\omega) = 1 - \omega$.
We observe that $X_1 \neq X_2$ but they are identically distributed:
$m(\omega \in \mathcal{U} : X_1(\omega) \leq x) = m(\omega \in \mathcal{U} : \omega \leq x) = m([0,x]) = x$ and
$m(\omega \in \mathcal{U} : X_2(\omega) \leq x) = m(\omega \in \mathcal{U} : 1 - \omega \leq x) = m(\omega \in \mathcal{U} : 1 - x \leq \omega) =$
$1 - m(\omega < 1 - x) = 1 - m([0, 1 - x]) = 1 - (1 - x) = x$.

Theorem (Constructing random variables)

If X is a random variable and $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a Borel measurable function (that is $f^{-1}(\mathcal{B}) \in \mathcal{B}$), then $f(X)$ is a random variable.

4th Lecture: Distribution functions

Distribution functions

Recall the definition of a distribution function. Let $\{a_j\}_{j\geq 1}$ be the countable set of points of jump of F and let b_j be the size of the jump at a_j: $F(a_j^+) - F(a_j^-) = F(a_j) - F(a_j^-) = b_j$. Let

$$F_d(x) = \sum_{j\geq 1} b_j \delta_{a_j}(x),$$

where $\delta_{a_j}(x)$ is the Heaviside function at a_j. The function F_d represents all the jumps of F in $(-\infty, x]$. Note that F_d is increasing, right-continuous, $F_d(-\infty) = 0$ and $F_d(+\infty) = \sum_{j\geq 1} b_j \leq 1$. The function F_d is the jumping part of F.

Theorem

The function $F_c(x) = F(x) - F_d(x)$ is positive, increasing, continuous.

Exercise: do the proof of the theorem.

Theorem

Let F be a distribution function. Suppose that there exists a continuous function G_c and a function G_d of the form $G_d(x) = \sum_{j\geq 1} b'_j \delta_{a'_j}(x)$, where $\{a'_j\}_{j\geq 1}$ is a countable set of real numbers and $\sum_{j\geq 1} b'_j < \infty$, such that $F = G_c + G_d$. Then $G_c = F_c$ and $G_d = F_d$ where F_c and F_d were defined above.

Exercise: do the proof of the theorem.

Definition

A distribution function that can be represented in the form $F = \sum_{j\geq 1} b_j \delta_{a_j}$, where $\{a_j\}_{j\geq 1}$ is a countable (or finite) set of real numbers $b_j > 0$ for every j and $\sum_{j\geq 1} b_j = 1$ is called a *discrete distribution function.* A distribution function that is continuous everywhere is called a *continuous distribution function.*

Suppose that for a distribution function F we have that $F_c \neq 0$ and $F_d \neq 0$. Let $\alpha = F_d(+\infty)$ such that $0 < \alpha < 1$ and let

$$F_1 = \frac{1}{\alpha} F_d \quad \text{and} \quad F_2 = \frac{1}{1-\alpha} F_c.$$

Then

$$F = F_d + F_c = \alpha F_1 + (1-\alpha) F_2 (^*)$$

and F_1 is a discrete distribution function and F_2 is a continuous distribution function and F is a convex combination of F_1 and F_2.

Remark

If $F_c = 0$ then F is discrete and we take $\alpha = 1$, so that $F_1 = F$ and $F_2 = 0$; and if $F_d = 0$, then F is continuous and we take $\alpha = 0$ and $F_1 = 0$ and $F_2 = F$ and in both cases $(^*)$ holds.

Theorem (Convex combination of distribution functions)

Every distribution function can be written as the convex combination of a discrete and a continuous distribution function. Such decomposition is unique.

Continuous distributions: absolutely continuous and singular distributions

Definition

A function F is said to be absolutely continuous (in $\mathbb{R}$ wrt the Lebesgue measure) iff there exists a function $f \in \mathbb{L}^1$ such that $\forall x < x'$ we have that

$$F(x') - F(x) = \int_x^{x'} f(y)\, dy.$$

***(A function f is in $\mathbb{L}^1$ iff $\int_{\mathbb{R}} |f(y)|\, dy < \infty$.)

There is a result in measure theory that says that such a function F has a derivative equal to f almost everywhere (on a set of full Lebesgue measure). In particular if F is a distribution function then

$$f \geq 0 \quad \text{a.e. and} \quad \int_{\mathbb{R}} f(y)\, dy = 1(^{**}).$$

Conversely, given any $f \in \mathbb{L}^1$ satisfying the previous conditions in (**), the function F defined for all $x \in \mathbb{R}$ as

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

is a distribution function that is absolutely continuous.

Definition

A function F is called singular if and only if it is not identically zero and F' exists and is equal to zero a.e.

Theorem

Let F be bounded increasing with $F(-\infty) = 0$ and let F' denote its derivative whenever it exists. Then:

1. *If S is the set of points x for which $F'(x)$ exists with $0 \leq F'(x) < +\infty$, then $m(S^c) = 0$.*
2. *The F' belongs to L^1 and we have for every $x < x'$ that*
$$\int_x^{x'} F'(y)\,dy \leq F(x') - F(x).$$
3. *If for all $x \in \mathbb{R}$*
$$F_{ac}(x) = \int_{-\infty}^x F'(y)\,dy \quad \text{and} \quad F_s(x) = F(x) - F_{ac}(x),$$
then $F'_{ac} = F'$ a.e., so that $F'_s = F' - F'_{ac} = 0$ a.e. and consequently F_s is singular if it is not identically zero.

Definition

Any positive function f that is equal to F' a.e. is called a density of F. F_{ac} is the absolutely continuous part of F and F_s is its singular part.

Remark

Note that:
1. the discrete part F_d defined above is part of the singular part F_s defined above;
2. F_{ac} is increasing and that $F_{ac} \leq F$. (Check it!)
Moreover, if $x < x'$ then $F_s(x') - F_s(x) = F(x') - F(x) - \int_x^{x'} f(y)\,dy \geq 0$, (from (2) of the previous theorem) therefore F_s is also increasing and $F_s \leq F$. (Check it!)

Theorem

Every distribution function F can be written as the convex combination of a discrete, a singular and an absolutely continuous distributions function and such decomposition is unique.

Continuous distribution function: the singular case

Let us construct the Cantor set. From the closed interval $[0,1]$ remove the central interval $(\frac{1}{3},\frac{2}{3})$. Then in the two remaining intervals remove the central intervals $(\frac{1}{9},\frac{2}{9})$ and $(\frac{7}{9},\frac{8}{9})$. After the 1st step we remain with two intervals of size $\frac{1}{3}$. In the 2nd step we remain with four intervals of size $\frac{1}{3^2}$ and so on. After n steps we have removed $1+2+4+8+\cdots+2^{n-1}=2^n-1$ disjoint intervals and remain 2^n closed intervals of size $\frac{1}{3^n}$. Let us order these intervals, by order from left to right and denote them by $J_{n,k}$, where $1\leq k\leq 2^n-1$ and denote their union by U_n. Note that

$$m(U_n)=1-\Big(\frac{2}{3}\Big)^n.$$

As n increases the set U_n increases to an open set U and let $\mathcal{C}:=U^c$ (the complementary wrt $[0,1]$) be the Cantor set. Then $m(\mathcal{C})=1-m(U)=1-\lim_{n\to\infty}m(U_n)=1-1=0.$

The Cantor distribution function

Now we define the Cantor distribution function. For each n,k, with $n \geq 1$ and $k = 1, \cdots, 2^n - 1$ let $c_{n,k} = \frac{k}{2^n}$ and let us define F in U in the following way:

if $x \in J_{n,k}$ then $F(x) = c_{n,k}$.

In each $J_{n,k}$ the function F is constant and it is strictly greater on any $J_{n,k'}$ at the right of $J_{n,k}$. Therefore, F is increasing and $F(0^+) = 0$ and $F(1^-) = 1$.
Now we complete the definition by setting $F(x) = 0$ for $x \leq 0$ and $F(x) = 1$ for $x \geq 1$.

$\Rightarrow$ Up to here the function F is defined on the domain $\mathcal{D} = (-\infty, 0] \cup U \cup [1, +\infty)$ and is increasing.

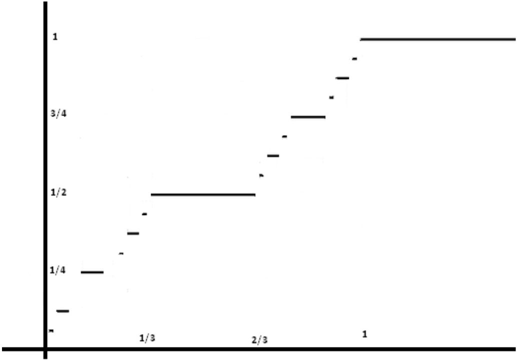
Now, since each $J_{n,k}$ is at a distance which is greater or equal than $1/3^n$ from any other $J_{n,k'}$ and since the total variation of F over each of the 2^n disjoint intervals that remain after removing $J_{n,k}$ is $\frac{1}{2^n}$, it follows that

$$0 \leq x' - x \leq \frac{1}{3^n} \Rightarrow 0 \leq F(x') - F(x) \leq \frac{1}{2^n}$$

Then, the function F is uniformly continuous on $\mathcal{D}$. ($\mathcal{D}$ is dense in $\mathbb{R}$). By a result (*) there exists a continuous and increasing function $\tilde{F}$ defined on $\mathbb{R}$ that coincides with F on $\mathcal{D}$.
This function $\tilde{F}$ is a continuous distribution function that is constant on each $J_{n,k}$ so that $\tilde{F}' = 0$ on U and also on $\mathbb{R} \backslash \mathcal{C}$, which means that $\tilde{F}$ is singular.

(*) Let f be increasing on a dense subset $\mathcal{D}$ of $\mathbb{R}$. If for any $x \in \mathbb{R}$ $\tilde{f}(x) = \inf_{x<t\in\mathcal{D}} f(t)$, then $\tilde{f}$ is increasing and right continuous everywhere. If f uniformly continuous, then $\tilde{f}$ is uniformly continuous.

A singular distribution function: the Cantor distribution function



Random variables

Definition

A random variable X is said to be discrete if it takes values in a finite or countable set, that is, if there exists a finite or countable set $B \in \mathbb{R}$ such that $\mathbb{P}(X \in B) = 1$.

Definition

A random variable X whose distribution function F has a density f is said to be absolutely continuous.

Note that,

1. If X is discrete, then $\mathbb{P}(X \in B) = \sum_{i:x_i \in B} \mathbb{P}(X = x_i)$;
2. If X is absolutely continuous with density f, then $\mathbb{P}(X \in A) = \int_A f(y)\,dy$, for any $A \in \mathcal{B}$.

Example

1. Let X be uniformly distributed in $(0,1)$. Find the distribution of $Y = X^n$, for $n \in \mathbb{N}$.
2. Let X be a continuous r.v. with density f_X. Find the distribution of $Y = X^2$ and of $Y = |X|$.
3. Let X be a r.v. with density given by

$$f(x) = \frac{1}{(1+x)^2}\mathbf{1}_{(0,+\infty)}(x).$$

Let $Y = max(X,c)$, where c is a strictly positive constant.

a) Find the distribution of X and Y and do the graphical representation.

b) Decompose the distribution function of Y in its discrete, absolutely continuous and singular parts.

5th Lecture: Random vectors

Random vectors

A random vector is just a vector whose components are random variables. We focus on the case $d = 2$.

• The Borel σ-algebra in $\mathbb{R}^2$ is the σ-algebra generated by rectangles of the form

$$\{(x,y) : a < x \leq b; c < y \leq d\}$$

and it is also generated by products sets of the form

$$B_1 \times B_2 = \{(x,y) : x \in B_1; y \in B_2\},$$

where $B_1, B_2 \in \mathcal{B}$.

A function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is Borel measurable iff $f^{-1}(\mathcal{B}) \in \mathcal{B}^2$.

Definition

Let X and Y be two random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The random vector (X,Y) induces a probability measure $\nu \in \mathcal{B}^2$ such that for $A \in \mathcal{B}^2$

$$\nu(A) = \mathbb{P}((X,Y) \in A) = \mathbb{P}(\omega \in \Omega : (X(\omega), Y(\omega)) \in A)$$

The measure ν is called the distribution measure of (X,Y).

We also define the inverse application $(X,Y)^{-1}$ in the following way:

$$\forall A \in \mathcal{B}^2 : (X,Y)^{-1}(A) = \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\}.$$

We note that the results that we have seen above for X^{-1} are also true for $(X,Y)^{-1}$.

Theorem

If X and Y are random variables and if $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is Borel measurable, then $f(X,Y)$ is a random variable.

Example

1. If X is a random variable and if $f:\mathbb{R}\longrightarrow\mathbb{R}$ is continuous, then $f(X)$ is a random variable. Therefore:
   - X^r; $|X|^r$ for positive real r; $e^{-\lambda X}$, for real λ, e^{itX}, for real t are random variables;
2. If X and Y are random variables then all these are random variables:
   - $X \pm Y;\ X.Y;\ X/Y;\ X\wedge Y := min(X,Y); X\vee Y := max(X,Y);$

Theorem

If $\{X_j\}_{j\geq 1}$ *is a sequence of random variables, then*

$$\inf_j X_j;\ \sup_j X_j;\ \liminf_j X_j;\ \limsup_j X_j;$$

are random variables not necessarily finite but a.e. defined and $\lim_{j\to+\infty} X_j$ *is a random variable on the set where there is convergence or divergence to* $\pm\infty$*.*

The distribution function of a random vector

Definition

The distribution function of a random vector (X,Y) is defined on $(x,y) \in \mathbb{R}^2$ by

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

F is also called the joint distribution function of the r.v. X and Y.

The distribution function just defined satisfies the following properties:

1. F is increasing in each variable.
2. F is right-continuous in each variable.
3. $\lim_{x\to-\infty} F(x,y) = 0$ and $\lim_{y\to-\infty} F(x,y) = 0$.
4. $\lim_{x\to+\infty, y\to+\infty} F(x,y) = 1$

Note that the distribution function of X (resp. Y) is obtained from $\lim_{y\to\infty} F(x,y) = F(x)$ (resp. $\lim_{x\to\infty} F(x,y) = F(y)$).

The distribution function of a random vector

The properties above are not sufficient to guarantee that a function $F:\mathbb{R}^2\to\mathbb{R}^2$ is the distribution function of a random vector. Let us see an example.

Example (Let $F(x,y)=\mathbf{1}_{\{x\geq 0,y\geq 0,x+y\geq 1\}}$.)

It is easy to see that F satisfies the properties above, nevertheless it is not the distribution function of a random vector. Suppose it is. Then we would have, for example, that: $\mathbb{P}(X\in(0,1],Y\in(0,1])=-1$, which cannot happen since $\mathbb{P}$ is a probability.

We need to introduce some extra condition, in order to avoid the previous example, which is the following:
- For any $a_1<b_1$ and $a_2<b_2$ we have $\mathbb{P}(X\in(a_1,b_1],Y\in(a_2,b_2])\geq 0$.

A function F satisfying the properties above is the distribution function of a random vector.

Exercises

1. Show that the function $F:\mathbb{R}^2 \to \mathbb{R}$ given by

$$F(x,y) = (1-e^{-x})(1-e^{-y})\mathbf{1}_{\{x\geq 0, y\geq 0\}}$$

is the distribution function of some random vector (X,Y).

2. Is the function $F:\mathbb{R}^2 \to \mathbb{R}$ given by

$$F(x,y) = (1-e^{-x-y})\mathbf{1}_{\{x\geq 0, y\geq 0\}}$$

the distribution function of some random vector (X,Y)?

Definition

A random vector (X,Y) is discrete iff it takes a finite or countable number of values.

Definition

Let (X,Y) be a random vector and let F be its distribution function. If there exists a function $f:\mathbb{R}^2\to\mathbb{R}$ such that $f(x,y)\geq 0$ and $\iint_{R^2} f(x,y)\,dx\,dy=1$ and if for any $(x,y)\in\mathbb{R}^2$

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,du\,dv,$$

then f is called the density function of the random vector (X,Y) or the joint density of the r.v. X and Y. In this case, we say that the random vector is absolutely continuous.

6th Lecture: Stochastic Independence

Independence

Definition

The collection of random variables $\{X_j\}_{j=1,\cdots,n}$ are said to be independent iff for any $\{B_j\}_{j=1,\cdots,n}$ with $B_j \in \mathcal{B}$, for any $j = 1, \cdots, n$, we have that

$$\mathbb{P}\Big(\cap_{j=1}^{n} (X_j \in B_j)\Big) = \Pi_{j=1}^{n} \mathbb{P}(X_j \in B_j). \tag{3}$$

Remark

1. The r.v. of an infinite family are said to be independent iff the r.v. in any finite subfamily are independent.
2. The r.v. are said to be pairwise independent iff every two of them are independent.
3. Note that (3) *implies that any of its subfamilies is independent, since*
$\mathbb{P}\Big(\cap_{j=1}^{k}(X_j \in B_j)\Big)=\mathbb{P}\Big(\cap_{j=1}^{n}(X_j \in B_j)\Big)=\Pi_{j=1}^{n}\mathbb{P}(X_j \in B_j)=\Pi_{j=1}^{k}\mathbb{P}(X_j \in B_j)$

Remark

We note that (3) is equivalent to $\mathbb{P}\Big(\cap_{j=1}^{n}(X_j \leq x_j)\Big) = \Pi_{j=1}^{n}\mathbb{P}(X_j \leq x_j)$, *for every set of real numbers* $\{x_j\}_{j=1}^{n}$. *(Show this!)*

We can rewrite (3) in terms of the probability measure $\mu_{(X_1,\cdots,X_n)}$ induced by the random vector $(X_1,\cdots,X_n)$ on $(\mathbb{R}^n,\mathcal{B}^n)$ as

$$\mu_{(X_1,\cdots,X_n)}(B_1\times\cdots\times B_n) = \Pi_{j=1}^{n}\mu_j(B_j) = \mu_1(B_1)\times\cdots\times\mu_n(B_n),$$

where $\mu_j := \mu_{X_j}$ is the probability measure induced by each random variable X_j in $(\mathbb{R},\mathcal{B})$. The induced measure is the product measure!

Remark

We can define the n-dimensional distribution function $F_{(X_1,\cdots,X_n)}$ *as*

$$\begin{aligned} F_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) =& \mathbb{P}(X_1 \leq x_1,\cdots,X_n \leq x_n) \\ =& \mu_{(X_1,\cdots,X_n)}((-\infty,x_1]\times\cdots\times(-\infty,x_n]). \end{aligned}$$

and the condition above is rewritten as $F(x_1,\cdots,x_n) = \Pi_{j=1}^{n}F_j(x_j)$.

Example

Let X_1 and X_2 be independent r.v. given by

$$X_1 = \begin{cases} 1, & 1/2 \\ -1, & 1/2 \end{cases} \quad \text{and} \quad X_2 = \begin{cases} 1, & 1/2 \\ -1, & 1/2. \end{cases}$$

Then, the three r.v. $\{X_1, X_2, X_1X_2\}$ are pairwise independent but they are not totally independent.

Independent events

Whenever a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed, the sets in $\mathcal{F}$ will be called events. We have seen above the notion of independent r.v. but what about independent events?

Definition

We say that the events $\{E_j\}_{j=1}^n$ are independent iff their indicators are independent, that is, for any subset $\{j_1, \cdots, j_\ell\}$ of $\{1, \cdots, n\}$ we have that $\mathbb{P}\Big(\cap_{k=1}^{\ell} E_{j_k}\Big) = \Pi_{k=1}^{\ell}\mathbb{P}(E_{j_k})$.

Theorem

If $\{X_j\}_{j=1}^n$ are independent r.v. and $\{f_j\}_{j=1}^n$ are Borel measurable functions, then $\{f_j(X_j)\}_{j=1}^n$ are independent r.v.

Do the proof of the theorem.

We have seen above that if $X_1, \cdots, X_n$ are independent r.v. then $F_{(X_1,\cdots,X_n)}(x_1, \cdots, x_n) = \Pi_{j=1}^n F_{X_j}(x_j)$. Now let us see the reciprocal.

Proposition

If there exist functions $F_1, \cdots, F_n$ such that

$$\lim_{x_j \to \infty} F_j(x_j) = 1$$

for all $j = 1, \cdots, n$ and if for all $(x_1, \cdots, x_n) \in \mathbb{R}^n$

$$F_{(X_1,\cdots,X_n)}(x_1, \cdots, x_n) = \Pi_{j=1}^n F_j(x_j),$$

then $\{X_j\}_{j=1}^n$ are independent and $F_j := F_{X_j}$ for all $j = 1, \cdots, n$.

Do the proof of the proposition.

Criterion for independence

Proposition

- *If* $\{X_j\}_{j=1}^n$ *are independent r.v. with densities* $f_{X_1}, \cdots, f_{X_n}$*, then the function*
$$f(x_1, \cdots, x_n) = \Pi_{j=1}^n f_{X_j}(x_j)$$
is the joint density of $\{X_j\}_{j=1}^n$ *or the density of the random vector* $(X_1, \cdots, X_n)$.
- *On the other hand, if* $X_1, \cdots, X_n$ *has a joint density* f *which satisfies*
$$f(x_1, \cdots, x_n) = \Pi_{j=1}^n f_j(x_j)$$
for all $(x_1, \cdots, x_n) \in \mathbb{R}^n$ *with* $f_j(x) \geq 0$ *and* $\int_{\mathbb{R}} f_j(x)dx = 1$*, then* $X_1, \cdots, X_n$ *are independent and* f_j *is the density of* X_j.

Do the proof of the theorem.

Constructing two independent random variables I

Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ where ($\mathcal{F}_j$ is the total σ-algebra) be discrete probability spaces.
We define the product space $\Omega^2 := \Omega_1 \times \Omega_2$ as the space of points $\omega = (\omega_1, \omega_2)$ with $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. The product σ-algebra $\mathcal{F}^2$ is the collection of all the subsets of Ω^2. We know from the beginning of the course that the probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ are determined by their values in ω_1, ω_2 respectively. Since Ω^2 is also countable we can define a probability measure $\mathbb{P}^2$ in $\mathcal{F}^2$ as

$$\mathbb{P}^2(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\})$$

which is the product measure of $\mathbb{P}_1$ and $\mathbb{P}_2$. Check that it is a probability measure. It has the property that if $S_1 \in \mathcal{F}_1$ and $S_2 \in \mathcal{F}_2$, then

$$\mathbb{P}^2(S_1 \times S_2) = \mathbb{P}_1(S_1)\mathbb{P}_2(S_2).$$

Now, let X_1 be a r.v. on Ω_1 and X_2 a r.v. on Ω_2; B_1 and B_2 Borel sets and $S_1 = X_1^{-1}(B_1) := \{\omega_1 \in \Omega_1 : X_1 \in B_1\}$ and $S_2 = X_2^{-1}(B_2)$. Note that $S_1 \in \mathcal{F}_1$ and $S_2 \in \mathcal{F}_2$. Then

$$\begin{aligned}\mathbb{P}^2(&X_1 \in B_1 \times X_2 \in B_2)\\ &=\mathbb{P}^2(S_1 \times S_2) = \mathbb{P}_1(S_1)\mathbb{P}_2(S_2) = \mathbb{P}_1(X_1 \in B_1)\mathbb{P}_2(X_2 \in B_2).\end{aligned}$$

To X_1 on Ω_1 and X_2 in Ω_2, we associate the function $\tilde{X}_1$ and $\tilde{X}_2$ defined on $\omega \in \Omega^2$ as $\tilde{X}_1(\omega) = X_1(\omega_1)$ and $\tilde{X}_2(\omega) = X_2(\omega_2)$. Now we have

$$\begin{aligned}\cap_{j=1}^2&\{\omega \in \Omega^2 : \tilde{X}_j(\omega) \in B_j\}\\ &= \Omega_1 \times \{\omega_2 \in \Omega_2 : X_2(\omega_2) \in B_2\} \cap \{\omega_1 \in \Omega_1 : X_1(\omega_1) \in B_1\} \times \Omega_2\\ &= \{\omega_1 \in \Omega_1 : X_1(\omega_1) \in B_1\} \times \{\omega_2 \in \Omega_2 : X_2(\omega_2) \in B_2\}.\end{aligned}$$

From where we conclude that

$$\mathbb{P}^2(\cap_{j=1}^2\{\tilde{X}_j \in B_j\}) = \mathbb{P}^2(\tilde{X}_1 \in B_1)\mathbb{P}^2(\tilde{X}_2 \in B_2),$$

so that the random variables $\tilde{X}_1$ and $\tilde{X}_2$ are independent!

Constructing n independent random variables I

Let $n \geq 2$ and $(\Omega_j, \mathcal{F}_j, \mathbb{P}_j)$ ($\mathcal{F}_j$ is the total σ-algebra) be n discrete probability spaces. We define the product space $\Omega^n := \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ as the space of points $\omega = (\omega_1, \cdots, \omega_n)$ with $\omega_j \in \Omega_j$. The product σ-algebra $\mathcal{F}^n$ is the collection of all the subsets of Ω^n. We know from the beginning of the course that for each j, the probability measure $\mathbb{P}_j$ is determined by its value in ω_j. Since Ω^n is also countable we can define a probability measure $\mathbb{P}^n$ in $\mathcal{F}^n$ as

$$\mathbb{P}^n(\{(\omega_1, \cdots, \omega_n)\}) = \Pi_{j=1}^n \mathbb{P}_j(\{\omega_j\})$$

which is the product measure of the $\{\mathbb{P}_j\}_{j=1}^n$. Check that it is a probability measure.
It has the property that if $S_j \in \mathcal{F}_j$, then

$$\mathbb{P}^n(S_1 \times \cdots \times S_n) = \Pi_{j=1}^n \mathbb{P}_j(S_j).$$

Now, let X_j be a r.v. on Ω_j, B_j a Borel set and $S_j = X_j^{-1}(B_j) := \{\omega_j \in \Omega_j : X_j \in B_j\}$. Note that $S_j \in \mathcal{F}_j$. Then

$$
\begin{aligned}
&\mathbb{P}^n(X_1 \in B_1 \times \cdots \times X_n \in B_n) \\
&\qquad = \mathbb{P}^n(S_1 \times \cdots \times S_n) = \Pi_{j=1}^n \mathbb{P}_j(S_j) = \Pi_{j=1}^n \mathbb{P}_j(X_j \in B_j).
\end{aligned}
$$

To each function X_j on Ω_j we associate the function $\tilde{X}_j$ on Ω^n defined on $\omega \in \Omega$ as $\tilde{X}_j(\omega) = X_j(\omega_j)$. Now we have

$$
\begin{aligned}
&\cap_{j=1}^n \{\omega \in \Omega^n : \tilde{X}_j(\omega) \in B_j\} \\
&\qquad = \cap_{j=1}^n \Omega_1 \times \cdots \times \Omega_{j-1} \times \{\omega_j \in \Omega_j : X_j(\omega_j) \in B_j\} \times \Omega_{j+1} \times \cdots \times \Omega_n \\
&\qquad = \Pi_{j=1}^n \{\omega_j \in \Omega_j : X_j(\omega_j) \in B_j\}
\end{aligned}
$$

From where we conclude that

$$\mathbb{P}^n(\cap_{j=1}^n \{\tilde{X}_j \in B_j\}) = \Pi_{j=1}^n \mathbb{P}^n(\tilde{X}_j \in B_j),$$

so that the random variables $\{\tilde{X}_j\}_{j=1}^n$ are independent!

Constructing independent random variables II

Let $\mathcal{U}^n = \{(x_1, \cdots, x_n) : 0 \leq x_j \leq 1, 1 \leq j \leq n\}$. The trace on $\mathcal{U}^n$ of $(\mathbb{R}^n, \mathcal{B}^n, m^n)$ is a probability space. For $j = 1, \cdots, n$, let $f_j : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function and let $X_j(x_1, \cdots, x_n) = f_j(x_j)$.
Then, the r.v. $\{X_j\}_{j=1}^n$ are independent.
If $f_j(x_j) = x_j$ then we get the n-coordinate variables in the cube.

Theorem (Existence of product measures)

Let $\{\mu_j\}_j$ be a finite or infinite sequence of probability measures on $(\mathbb{R}, \mathcal{B})$ or equivalently, let their distribution functions be given. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of independent r.v. $\{X_j\}_j$ defined on it such that for each j, the measure μ_j is the probability measure of X_j.

Example

Let $X_1, \cdots, X_n$ be independent r.v. with Rayleight density with parameter θ given by

$$f(x) = \begin{cases} x^2/\theta e^{-\frac{x^2}{2\theta^2}}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

(a) Find the joint density of $Y_1, \cdots, Y_n$ where $Y_i = X_i^2$.
(b) Find the distribution of $U = \min_{1 \leq i \leq n}\{X_i\}$.

Example

Let $X_1, \cdots, X_n$ be independent r.v. with exponential distribution of parameters $\alpha_1, \cdots, \alpha_n$. (to simplify take $n = 3$.)
(a) Find the joint density of $Y = \min_{1 \leq i \leq n}\{X_i\}$ by showing that it is exponential. Find the parameter.
(b) Prove that for $k = 1, \cdots, n$ it holds that $\mathbb{P}(X_k = Y) = \frac{\alpha_k}{\alpha_1 + \cdots + \alpha_n}$.

7th Lecture: Mathematical Expectation

Mathematical expectation is integration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to the probability measure $\mathbb{P}$. To avoid complications we assume that the r.v. are finite everywhere.

Definition

A countable partition of Ω is a countable family of disjoint sets A_j with $A_j \in \mathcal{F}$ and such that $\Omega = \cup_{j\geq} A_j$. In this case we have that $1 = \mathbf{1}_\Omega = \sum_j \mathbf{1}_{A_j}$.

Definition

A r.v. X is said to belong to the weighted partition $\{A_j, bj\}$ is for all $\omega \in \Omega$ we have that $X(\omega) = \sum_j b_j \mathbf{1}_{A_j}(\omega)$. Note that X is a discrete r.v.

Remark

Every discrete r.v. belongs to a weighted partition: take $\{b_j\}_j$ as the countable set of the possible values of X and $A_j = \{\omega \in \Omega; X(\omega) = b_j\}$. If j ranges over a finite set the r.v. is said to be simple.

Mathematical expectation: positive r.v.

• If X is a positive discrete r.v. belonging to the weighted partition $\{A_j, bj\}$, then its expectation is defined as

$$\mathbb{E}[X] = \sum_j b_j \mathbb{P}(A_j).$$

Note that $\mathbb{E}[X]$ is a number, in this case, since $b_j \geq 0$, positive or $+\infty$.

• Suppose now that X is a positive random variable and for each positive integers m, n, let

$$A_{mn} = \Big\{\omega : \frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}\Big\} = X^{-1}\Big(\Big[\frac{n}{2^m}, \frac{n+1}{2^m}\Big]\Big),$$

so that $A_{mn} \in \mathcal{F}$. For each m, let X_m be the random variable that takes the value $\frac{n}{2^m}$ in A_{mn}, that is

$$X_m(\omega) = \frac{n}{2^m} \quad \text{iff} \quad \frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}.$$

It is easy to see that for each m we have that for all $\omega \in \Omega$, $X_m(\omega) \leq X_{m+1}(\omega)$.
Now let $\omega \in \Omega$ and note that if $\frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m}$, then $X_m(\omega) = \frac{n}{2^m}$, so that

$$0 \leq X(\omega) - X_m(\omega) < \frac{1}{2^m},$$

from where we get that $\lim_{m\to\infty} X_m(\omega) = X(\omega)$. So the sequence of r.v. $\{X_m\}_m$ is increasing and converges pointwisely to X.
Note that

$$\mathbb{E}[X_m] = \sum_{n=0}^{\infty} \frac{n}{2^m} \mathbb{P}\Big(\frac{n}{2^m} \leq X < \frac{n+1}{2^m}\Big).$$

If $\mathbb{E}[X_m] = +\infty$ then we define $\mathbb{E}[X] = +\infty$, otherwise, we define $\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_m]$.
Note that the limit can be infinite.

For a general X we take $X = X^+ - X^-$, where $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$. Both X^+, X^- are positive, so their expectation is defined and unless both expectations are $+\infty$ we define $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$. We sat that X has finite or infinite expectation according to $\mathbb{E}[X]$ is finite or infinite.

When the expectation of X exists we use the notation

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

For $\Lambda \in \mathcal{F}$ we have

$$\mathbb{E}[X\mathbf{1}_\Lambda] = \int_\Lambda X(\omega)\mathbb{P}(d\omega) = \int_\Omega \mathbf{1}_\Lambda(\omega)X(\omega)d\mathbb{P}$$

and it is called the integral of X wrt $\mathbb{P}$ over the set Λ.

• When the integral above exists and is finite we say that X is integrable in Λ wrt $\mathbb{P}$.

Example (Lebesgue-Stietjes integral)

For $(\mathbb{R}, \mathcal{B}, \mu)$ and $X = f$ and $\omega = x$ we have

$$\int_\Lambda X(\omega)\mathbb{P}(d\omega) = \int_\Lambda f(x)\mu(dx) = \int_\Lambda f(x)d\mu.$$

When F is the distribution function of μ we also write (for $\Lambda = (a,b]$)

$$\int_{(a,b]} f(x)dF(x).$$

To distinguish the intervals $(a,b], [a,b], (a,b)$ and $[a,b)$ we use the notation

$$\int_{a+0}^{b+0}, \int_{a-0}^{b+0}, \int_{a+0}^{b-0}, \int_{a-0}^{b-0}.$$

For $(\mathcal{U}, \mathcal{B}, m)$ the integral is $\int_a^b f(x)m(dx) = \int_a^b f(x)dx$.
Since μ is atomless we do not need to distinguish the intervals.

Properties of the mathematical expectation

In what follows X and Y are r.v. and $a,b\in\mathbb{R}$ and $\Lambda\in\mathcal{F}$.

(1) Absolute integrability: $\int_\Lambda X d\mathbb{P}$ is finite iff $\int_\Lambda |X| d\mathbb{P}$ is finite.

(2) Linearity: $\int_\Lambda (aX+bY)d\mathbb{P} = a\int_\Lambda X d\mathbb{P} + b\int_\Lambda Y d\mathbb{P}$.

(3) Set additivity: If $\{\Lambda_n\}_{n\geq 1}$ are disjoint, then

$$\int_{\cup_{n\geq 1}\Lambda_n} X d\mathbb{P} = \sum_{n\geq 1}\int_{\Lambda_n} X d\mathbb{P}.$$

(4) Positivity: If $X\geq 0$ a.e. on Λ (this means there is a subset of Λ with weight one wrt $\mathbb{P}$ where X is positive), then

$$\int_\Lambda X d\mathbb{P}\geq 0.$$

Properties of the mathematical expectation

(5) Monotonicity: If $X_1 \leq X \leq X_2$ a.e. in Λ, then

$$\int_\Lambda X_1 d\mathbb{P} \leq \int_\Lambda X d\mathbb{P} \leq \int_\Lambda X_2 d\mathbb{P}.$$

(6) Mean value Theorem: If $a \leq X \leq b$ a.e. in Λ, then

$$a\mathbb{P}(\Lambda) \leq \int_\Lambda X d\mathbb{P} \leq b\mathbb{P}(\Lambda).$$

(7) Modulus inequality: $\left|\int_\Lambda X d\mathbb{P}\right| \leq \int_\Lambda |X| d\mathbb{P}$.

(8) Dominated convergence Theorem: If $\lim_{n\to\infty} X_n = X$ a.e. on Λ and if for $n \geq 1$ $|X_n| \leq Y$ a.e. on Λ and $\int_\Lambda Y d\mathbb{P} < \infty$, then

$$\lim_{n\to\infty} \int_\Lambda X_n d\mathbb{P} = \int_\Lambda X d\mathbb{P} = \int_\Lambda \lim_{n\to\infty} X_n d\mathbb{P}.$$

Properties of the mathematical expectation

(9) Bounded convergence Theorem: If $\lim_{n\to\infty} X_n = X$ a.e. on Λ and there exists a constant M such that $n \geq 1$ $|X_n| \leq M$ a.e. on Λ, then the previous equality is true.

(10) Monotone convergence Theorem: If $X_n \geq 0$ and $X_n \uparrow X$ a.e. on Λ, then the previous equality is true if we allow $+\infty$ as a value.

(11) Integration term by term: If $\sum_{n\geq 1} \int_\Lambda |X_n| d\mathbb{P} < \infty$, then $\sum_{n\geq 1} |X_n| < \infty$ a.e. on Λ, so that $\sum_{n\geq 1} X_n$ converges a.e. on Λ and

$$\int_\Lambda \sum_{n\geq 1} X_n d\mathbb{P} = \sum_{n\geq 1} \int_\Lambda X_n d\mathbb{P}.$$

(12) Fatou's Lemma: If $X_n \geq 0$ a.e. on Λ, then

$$\int_\Lambda (\liminf_{n\to\infty} X_n) d\mathbb{P} \leq \liminf_{n\to\infty} \int_\Lambda X_n d\mathbb{P}.$$

The Lebesgue-Stieltjes integral

Let f be a continuous function defined on $[a,b]$ and let F be a distribution function. The Riemann-Stieltjes integral of f on $[a,b]$ wrt F is defined as the limit of the Riemann sums of the form

$$\sum_{i=1}^{n} f(\tilde{x}_i)(F(x_{i+1}) - F(x_i)),$$

where $x_1 = a, x_n = b$, $x_i < x_{i+1}$ and $\tilde{x}_i$ is an arbitrary point in $[x_i, x_{i+1}]$. The limit is taken by making the norm of the partition $\{x_i\}_i$ tending to 0, that is $\max_{i=1,\cdots,n}(x_{i+1} - x_i) \to 0$. The limit exists, when f is continuous, and it is denoted by $\int_a^b f(x)dF(x)$. Note that

$$\int_{\mathbb{R}} f(x)dF(x) = \lim_{a\to-\infty, b\to+\infty} \int_a^b f(x)dF(x).$$

Example

Compute $\int_{\mathbb{R}} F_0(x)dF_0(x)$, for $F_0(x) = \delta_0(x)$, the Heaviside function at 0.

The Lebesgue-Stieltjes integral

To extend the definition to discontinuous functions we do it like this. Let f be a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$. We want to define $\int_{\mathbb{R}} f(x)dF(x)$ for a distribution function F. First we define it for $f(x) = \mathbf{1}_{[a,b]}(x)$ as $\int_{\mathbb{R}} f(x)dF(x) = F(b) - F(a)$. Then we extend the definition as we did before.

Remark

• *When F is the distribution function of a discrete random variable X taking values $\{x_i\}_{i\geq 1}$ then $\int_{\mathbb{R}} f(x)dF(x) = \sum_i f(x_i)\mathbb{P}(X = x_i)$ and $\int_{(a,b]} f(x)dF(x) = \sum_{i:a<x_i\leq b} f(x_i)\mathbb{P}(X = x_i)$.*

• *When F is the distribution function of an absolutely continuous random variable X with density f_X, then $\int_{\mathbb{R}} f(x)dF(x) = \int_{\mathbb{R}} f(x)f_X(x)dx$ and $\int_a^b f(x)dF(x) = \int_a^b f(x)f_X(x)dx$.*

• *When $F = \alpha F_d + \beta F_{ac} + \gamma F_s$, then $\int_{\mathbb{R}} f(x)dF(x) = \alpha \int_{\mathbb{R}} f(x)dF_d + \beta \int_{\mathbb{R}} f(x)dF_{ac} + \gamma \int_{\mathbb{R}} f(x)dF_s$.*

8th Lecture: Integrability criterion and Classical Inequalities

Proposition

For a r.v. X with distribution function F we have that

$$\mathbb{E}[X] = \int_0^{+\infty} (1 - F(x))dx - \int_{-\infty}^{0} F(x)dx.$$

Corollary

For a non-negative r.v. X, we have that $\mathbb{E}[X] = \int_0^{+\infty} \mathbb{P}(X > x)dx$.

Theorem

Let $k \in \mathbb{N}$, then

$$\mathbb{E}[X^k] = k\Big[\int_0^{+\infty} (1 - F_X(x))x^{k-1}dx - \int_{-\infty}^{0} F_X(x)x^{k-1}dx\Big].$$

Do the proof of the previous results.

Integrability

Theorem (Integrability criterion)

For a r.v. X we have that

$$\sum_{n\geq 1}\mathbb{P}(|X|\geq n)\leq \mathbb{E}[|X|]\leq 1+\sum_{n\geq 1}\mathbb{P}(|X|\geq n),$$

so that $\mathbb{E}[|X|]<\infty$ iff the series above converges.

Lemma

If X is a non-negative r.v. then $\mathbb{E}[X]=\sum_{n\geq 1}\mathbb{P}(X\geq n)$.

Do the proof of the previous results.

Relation between integrals:

There is a basic relation between an integral wrt $\mathbb{P}$ over sets of $\mathcal{F}$ and the Lebesgue-Stieltjes integral wrt to μ over sets of $\mathcal{B}$.

Theorem

Let X be a r.v. defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which induces the probability space $(\mathbb{R}, \mathcal{B}, \mu)$ and let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. Then, $\int_\Omega f(X(\omega))\mathbb{P}(d\omega) = \int_\mathbb{R} f(x)\mu(dx)$.

Do the proof of the previous result.

In higher dimensions the result is the same. We state it for $d=2$ as

Theorem

Let (X,Y) be a random vector defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ which induces the probability space $(\mathbb{R}^2,\mathcal{B}^2,\nu)$ and let $f:\mathbb{R}^2\to\mathbb{R}$ be a Borel measurable function. Then,
$\int_\Omega f(X(\omega),Y(\omega))\mathbb{P}(d\omega)=\iint_{\mathbb{R}^2} f(x,y)\nu(dx,dy)$.

From the previous theorem, for a r.v. X with distribution function F_X and distribution measure μ_X, it holds that

$$\mathbb{E}[X]=\int_{\mathbb{R}} x\mu_X(dx)=\int_{\mathbb{R}} xdF_X(x)$$

and more generally $\mathbb{E}[f(X)]=\int_{\mathbb{R}} f(x)\mu_X(dx)=\int_{\mathbb{R}} f(x)dF_X(x)$.

Remark

An important consequence of the previous theorem is that for $f(x,y)=x+y$ we obtain that (linearity of the integral)

$$\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]. \quad \textit{Show this!}$$

Moments of a r.v.

Definition

Let $a \in \mathbb{R}$ and $r \geq 0$. The absolute moment of a r.v. X of order r about a is defined as $\mathbb{E}[|X-a|^r]$.

Remark

If μ_X and F_X are the distribution measure and the distribution function of X, then

$$\mathbb{E}[|X-a|^r] = \int_{\mathbb{R}} |x-a|^r \mu(dx) = \int_{\mathbb{R}} |x-a|^r dF_X(x),$$

$\mathbb{E}[(X-a)^r] = \int_{\mathbb{R}} (x-a)^r \mu(dx) = \int_{\mathbb{R}} (x-a)^r dF_X(x)$. *When $r=1$ and $a=0$, the previous moment is $\mathbb{E}[X]$. The moments about $a=\mathbb{E}[X]$ are called central moments and the one of order 2 is called the variance:*

$$Var(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Definition (The space $\mathbb{L}^p = \mathbb{L}^p(\Omega, \mathcal{F}, \mathbb{P})$)

For a positive number p, we say that $X \in \mathbb{L}^p$ iff $\mathbb{E}[|X|^p] < \infty$.

Theorem

Let X and Y be random variables and p, q such that $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$ (p and q are said to be conjugate). Then

1. *(Holder's Inequality)*

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} \mathbb{E}[|Y|^q])^{1/q} \quad (1)$$

2. *(Minkowski's inequality)*

$$(\mathbb{E}[|X+Y|^p])^{1/p} \leq (\mathbb{E}[|X|^p])^{1/p} + (\mathbb{E}[|Y|^p])^{1/p}$$

Remark

When $p = 2$, (1) is called the Cauchy-Schwarz's inequality.

9th Lecture: Classical Inequalities

Jensen's inequality

Theorem

If $\varphi : \mathbb{R} \to \mathbb{R}$ *is a convex function and* X *and* $\varphi(X)$ *are integrable r.v. then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Do the proof of the previous result.

Example

1. $\varphi(x) = |x|$;
2. $\varphi(x) = x^2$;
3. $\varphi(x) = |x|^p$, $p \geq 1$.

Tchebychev inequalities:

Theorem (Basic inequality)

Let X be a non-negative r.v. For any $\lambda > 0$, $\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}[X]}{\lambda}$.

Theorem (Classic inequality)

Let X be a r.v. with finite variance. For any $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \lambda) \leq \frac{Var(X)}{\lambda^2}.$$

Theorem (Markov's inequality)

Let X be a r.v. with $\mathbb{E}[|X|]^t < \infty$. For any $\lambda > 0$,

$$\mathbb{P}(|X| \geq \lambda) \leq \frac{\mathbb{E}[|X|^t])}{\lambda^t}.$$

Show the previous results!

Again independence:

Theorem

If X and Y are two independent r.v. with finite expectation, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Do the proof of the previous result.

Definition

Let X and Y be r.v. with finite expectation. The covariance between X and Y is defined as

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

When $Cov(X,Y) = 0$, we say that X and Y are uncorrelated.

Remark

Be careful: uncorrelation does NOT imply independence.

Example

Analyze the case when (X,Y) has density given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)}$$

and take $\rho = 0$.

Example

Show that if X and Y are r.v. taking only the values 0 and 1 and if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then X and Y are independent.

Proposition

Let $X_1, \cdots, X_n$ *be integrable r.v. such that* $Cov(X_i, X_j) = 0$ *for* $i \neq j$. *Then*

$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i).$$

Example

Let X and Y be r.v. with finite variance: show that if $Var(X) \neq Var(Y)$, then $X+Y$ and $X-Y$ are not independent r.v.

Example

Let X and Y be r.v. with finite variance: show that if X and Y are independent, then

$$Var(XY) = Var(X)Var(Y) + (\mathbb{E}[X])^2 Var(Y) + (\mathbb{E}[Y])^2 Var(X).$$

10th Lecture: Convergence of sequences of r.v.

Notion of convergence: almost everywhere

Recall that we have seen that if $\{X_n\}_{n\geq 1}$ is a sequence of r.v. then $\lim_{n\to\infty} X_n$ is a r.v.
The notion of convergence is of convergence to a finite limit: if we say $\{X_n\}_{n\in\mathbb{N}}$ converges in $\Lambda \in \mathcal{F}$, this means that for all $\omega \in \Lambda$ we have that the sequence $\{X_n(\omega)\}_{n\in\mathbb{N}}$ converges. When $\Lambda = \Omega$ we say the convergence holds everywhere.

Definition (Almost everywhere convergence)

The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge almost everywhere to X iff there exists a null set N such that

$$\forall \omega \in \Omega \setminus N : \lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \text{finite.}$$

Convergence in probability

Theorem

A sequence of r.v. $\{X_n\}_{n\in\mathbb{N}}$ *converges a.e. to* X *iff for every* $\epsilon > 0$ *we have that*

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| \leq \epsilon \quad \text{for all} \quad n \geq m) = 1$$

or equivalently

$$\lim_{m\to\infty} \mathbb{P}(|X_n - X| > \epsilon \quad \text{for some} \quad n \geq m) = 0 \tag{4}$$

Exercise: do the proof of the theorem.

Definition (Convergence in probability)

The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in probability to X, iff for every $\epsilon > 0$ it holds that $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$.

Convergence a.e implies convergence in probability

Theorem

Convergence a.e. to X implies convergence in probability to X.

Definition (Convergence in $\mathbb{L}^p$, $0 < p < \infty$)

The sequence $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in $\mathbb{L}^p$ to X, iff $X_n \in \mathbb{L}^p$, $X \in \mathbb{L}^p$ and

$$\lim_{n\to\infty} \mathbb{E}[(|X_n - X|)^p] = 0. \tag{5}$$

Definition

We say that X is dominated by Y if $|X| \leq Y$ a.e. and that the sequence is dominated by Y, if this is true for any n with the same Y. Moreover, if above Y is constant we say that X or X_n is uniformly bounded.

Above we can suppose $X = 0$ since the definitions hold for $X_n - X$.

Convergence in $\mathbb{L}^p$ implies convergence in probability

Theorem

Convergence in $\mathbb{L}^p$ implies convergence in probability. The converse is true if the sequence is dominated by some $Y \in \mathbb{L}^p$.

- Convergence in probability does not imply convergence in $\mathbb{L}^p$ and convergence in $\mathbb{L}^p$ does not imply convergence a.e.
- Convergence a.e. does not imply convergence in $\mathbb{L}^p$.

Theorem (Scheffé's Theorem)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of r.v. with densities $f_1, f_2, \cdots$ and let X be a r.v. with density f. If $\lim_n f_n = f$, holds a.e. then $\lim_{n\to\infty} \int_{\mathbb{R}} |f_n - f| dx = 0$.

Exercise: do the proof of the results above.

The $\lim inf_n E_n$ and the $\limsup_n E_n$

Definition

Let $\{E_n\}_{n\in\mathbb{N}}$ be a sequence of subsets of Ω. The $\limsup_n E_n$ and the $\liminf_n E_n$ are defined by

$$\limsup_n E_n = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} E_n \quad \text{and} \quad \liminf_n E_n = \cup_{m=1}^{\infty} \cap_{n=m}^{\infty} E_n \tag{6}$$

Remark

- *Note that a point belongs to* $\limsup_n E_n$ *iff it belongs to infinitely many terms of the sequence* $\{E_n\}_{n\in\mathbb{N}}$ *and belongs to* $\liminf_n E_n$ *iff it belongs to all the terms of the sequence from a certain point on. (Show this!)*
- *Also note that* $(\limsup_n E_n^c)^c = \liminf_n E_n$.
- *The event* $\limsup_n E_n$ *occurs iff the events* E_n *occur i.o.*
- *If each* $E_n \in \mathcal{F}$*, then* $\mathbb{P}(\limsup_n E_n) = \lim_{m\to\infty} \mathbb{P}(\cup_{n=m}^{\infty} E_n)$.

11th Lecture: Borel-Cantelli's lemmas and weak convergence

Borel-Cantelli's Lemma

Lemma (Borel-Cantelli - the convergent part)

For $\{E_n\}_{n\in\mathbb{N}}$ *arbitrary events, if* $\sum_{n=1}^{\infty}\mathbb{P}(E_n)<\infty$*, then* $\mathbb{P}(E_n \ i.o.)=0$*.*

We can rephrase the first theorem above:

Theorem

A sequence of r.v. $\{X_n\}_{n\geq 1}$ *converges a.e. to* 0 *iff* $\forall\epsilon>0$ $\mathbb{P}(\{|X_n|>\epsilon\} \ i.o.)=0$*.*

Theorem

If $\{X_n\}_{n\geq 1}$ *converges in probability to* X*, then there exists a sequence* $\{n_k\}$ *of integers growing to* ∞ *such that* $X_n\to X$ *a.e. This means that convergence in probability implies converges a.e. along a subsequence.*

Exercise: do the proof of the results.

If we add independence then we have

Lemma (Borel-Cantelli - the divergent part)

For independent events $\{E_n\}_{n\in\mathbb{N}}$, *if* $\sum_{n=1}^{\infty}\mathbb{P}(E_n)=\infty$, *then* $\mathbb{P}(E_n \text{ i.o.})=1$.

** The previous result also holds with pairwise independence. (We will not see the proof in this case!)*

Remark

Removing the independence assumption, the result is false. Take $E_n=A$ *with* $0<\mathbb{P}(A)<1$.

Corollary

For independent events $\{E_n\}_{n\in\mathbb{N}}$, *then* $\mathbb{P}(E_n \text{ i.o.})=0 \quad \text{or} \quad 1$ *if* $\sum_{n=1}^{\infty}\mathbb{P}(E_n)<\infty \quad \text{or} \quad \sum_{n=1}^{\infty}\mathbb{P}(E_n)=\infty$

Exercise: do the proof of the results above.

The weak convergence

If a sequence of r.v. $\{X_n\}_{n\geq 1}$ converges to some limit, does the sequence of probability distribution measures $\{\mu_n\}_{n\in\mathbb{N}}$ converges in some sense? Is is true that $\lim_n \mu_n(A)$ exists for any $A \in \mathcal{B}$? NO!!! And if $\{\mu_n\}_{n\in\mathbb{N}}$ converges in some sense, is the limit necessarily a probability measure? NO!!!

Definition

A p.m. μ in $(\mathbb{R},\mathcal{B})$ with $\mu(\mathbb{R})\leq 1$ is a called a subprobability measure.

Definition (Weak convergence)

A sequence of subprobability measures μ in $(\mathbb{R},\mathcal{B})$ is said to converge weakly to a subprobability measure μ iff there exists a dense subset D of $\mathbb{R}$ such that $\forall a,b\in D, a<b$, $\lim_{n\to\infty}\mu_n((a,b])=\mu((a,b])$. We will use the notation $\mu_n \to^v \mu$ and μ is the weak limit (UNIQUE) of $\{\mu_n\}_{n\in\mathbb{N}}$.

Definition

An interval (a,b) is said to be a continuity interval of μ is a,b are not atoms of μ (that is $\mu((a,b)) = \mu([a,b])$).

Lemma

Let $\{\mu_n\}_{n\in\mathbb{N}}$ and μ be subprobability measures. The following propositions are equivalent:

1. *For every finite interval (a,b) and $\epsilon > 0$, there exists an $n_0(a,b,\epsilon)$ such that if $n \geq n_0$, then* $\mu((a+\epsilon, b-\epsilon)) - \epsilon \leq \mu_n((a,b)) \leq \mu((a-\epsilon, b+\epsilon)) + \epsilon$.
2. *for every continuity interval $(a,b]$ of μ we have that* $\lim_{n\to\infty} \mu_n((a,b]) = \mu((a,b])$.
3. $\mu_n \to^v \mu$.

Exercise: do the proof of the lemma.

12th Lecture: Helly's extraction theorem

Helly's extraction theorem

Recall that given any sequence of real numbers in a subset of $[0,1]$, there is a subsequence which converges and the limit is an element of that set. (This means that $[0,1]$ is sequentially compact). The set of subprobability measures is sequentially compact wrt the weak convergence.

Theorem (Helly's extraction theorem)

Given any sequence of subprobability measures, there exists a subsequence that converges weakly to a subprobability measure.

Exercise: do the proof of the theorem.

Uniqueness of the limit

Definition

Given F_n and F subdistribution functions, we say that F_n converges weakly to F and we write $F_n \to^v F$ if $\mu_n \to^v \mu$, where μ_n and μ are the subprobability measures of F_n and F, respectively.

Theorem

If every weakly converging subsequence of a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ of subprobability measures converges to the same μ, then $\mu_n \to^v \mu$.

Theorem

Let μ_n and μ be subprobability measures. Then $\mu_n \to^v \mu$ iff for all $f \in C_k$ (or C_0) we have that $\int_{\mathbb{R}} f(x)\mu_n(dx) \to_{n\to\infty} \int_{\mathbb{R}} f(x)\mu(dx)$.

Exercise: do the proof of the results above.

13th Lecture: Convergence in distribution

Convergence in distribution

Definition (Convergence in distribution)

A sequence of r.v. $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in distribution to F iff the corresponding sequence of distribution functions $\{F_n\}_{n\in\mathbb{N}}$ converges weakly to the distribution function F.

If X is a r.v. with distribution function F, we will say that $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X.

Theorem

Let F_n and F be the distribution functions of the r.v. X_n and X. If $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability, then $F_n \to^n F$. (Convergence in probability implies convergence in distribution).

Exercise: do the proof of the results above.

Lemma

Let $c \in \mathbb{R}$. Then $\{X_n\}_{n\in\mathbb{N}}$ converges to c in probability iff $\{X_n\}_{n\in\mathbb{N}}$ converges to c in distribution.

When the limit is a constant, the convergence in probability is equivalent to the convergence in distribution.
Note that it is not true that if $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X and $\{Y_n\}_{n\in\mathbb{N}}$ converges in distribution to Y, the sum $\{X_n+Y_n\}_{n\in\mathbb{N}}$ converges in distribution to $X+Y$. But see the special case $Y=0$!

Theorem

If $\{X_n\}_{n\in\mathbb{N}}$ converges in distribution to X and $\{Y_n\}_{n\in\mathbb{N}}$ converges in distribution to 0, then:

- $\{X_n+Y_n\}_{n\in\mathbb{N}}$ *converges in distribution to* X
- $\{X_nY_n\}_{n\in\mathbb{N}}$ *converges in distribution to* 0

Exercise: do the proof of the results above.

Limit Theorems

The law of large numbers has to do with partial sums of a sequence of r.v. $\{X_n\}_{n\geq 1}$.

$$S_n := X_1 + \cdots + X_n$$

Weak/strong Law

Depends on whether $\frac{S_n - \mathbb{E}[S_n]}{n} \to_{n\to\infty} 0$, in probability or a.e. (needs $\mathbb{E}[S_n]$ finite!)

Remark

We have seen above that if a sequence converges to 0 *in* $\mathbb{L}^2$ *then it converges to* 0 *in probability and then it converges a.e. to* 0 *along a subsequence.*

Weak Law of Large Numbers (bounded second moments)

Theorem (Chebychev)

If $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of uncorrelated r.v. whose second moments have a common bound, then* $\frac{S_n-\mathbb{E}[S_n]}{n} \to_{n\to\infty} 0$ *in* $\mathbb{L}^2$ *and therefore also in probability.*

Theorem (Rajchman)

If $\{X_n\}_{n\in\mathbb{N}}$ *is a sequence of uncorrelated r.v. whose second moments have a common bound, then* $\frac{S_n-\mathbb{E}[S_n]}{n} \to_{n\to\infty} 0$ *a.e.*

Exercise: do the proof of the results above.

14th Lecture: Law of Large Numbers

Equivalent sequences (Kintchine)

Definition

Two sequences of r.v. $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are said to be equivalent iff $\sum_{n\geq 1}\mathbb{P}(X_n \neq Y_n) < \infty$.

Theorem

If $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are equivalent then $\sum_{n\geq 1}(X_n - Y_n)$ converges a.e. Moreover, if $a_n \to +\infty$, then $\frac{1}{a_n}\sum_{j=1}^{n}(X_j - Y_j)$ converges a.e. to 0.

Theorem (Weak Law of Large Numbers of Kintchine)

If $\{X_n\}_{n\geq 1}$ is a sequence of pairwise independent and identically distributed r.v. with finite mean m, then $\frac{S_n}{n} \to m$ in probability.

Exercise: do the proof of the results.

Strong Law of Large Numbers

Theorem (Kolmogorov's Inequality)

Let $\{X_n\}_{n\geq 1}$ *be independent r.v. with* $\mathbb{E}[X_n]=0$ *for every* n *and* $\mathbb{E}[X_n^2]=\sigma^2(X_n)<\infty$. *Then, for every* $\varepsilon>0$ *it holds that*

$$\mathbb{P}\Big(\max_{1\leq j\leq n}|S_j|>\varepsilon\Big)\leq\frac{\sigma^2(S_n)}{\varepsilon^2}.$$

Theorem

Let $\{X_n\}_{n\geq 1}$ *be independent r.v. with* $\mathbb{E}[X_n]<\infty$ *and suppose that* $\exists A>0$ *s.t.* $|X_n-\mathbb{E}[X_n]|\leq A<\infty,\quad\forall n\in\mathbb{N}$. *Then,* $\forall\varepsilon>0$*:* $\mathbb{P}\Big(\max_{1\leq j\leq n}|S_j|\leq\varepsilon\Big)\leq\frac{(2A+4\varepsilon)^2}{\sigma^2(S_n)}$.

Exercise: do the proof of the results.

Strong Law of Large Numbers

Lemma (Kronecker's Lemma)

Let $\{x_k\}_{k\geq 1}$ a sequence of real numbers, $\{a_k\}_{k\geq 1}$ a sequence of strictly positive real numbers $\uparrow \infty$. If $\sum_{n\geq 1}\frac{x_n}{a_n}<\infty$, then $\frac{1}{a_n}\sum_{j=1}^{n} x_j \to 0$.

Theorem

Let $\varphi:\mathbb{R}\to\mathbb{R}$ be a positive, even and continuous function, such that as $|x|$ increases: $\frac{\varphi(x)}{|x|}\uparrow$ and $\frac{\varphi(x)}{x^2}\downarrow$. Let $\{X_n\}_{n\geq 1}$ be a sequence of independent r.v. with $\mathbb{E}[X_n]=0$ for every n and let $0<a_n\uparrow+\infty$. If $\sum_{n\geq 1}\frac{\mathbb{E}[\varphi(X_n)]}{\varphi(a_n)}<\infty$, then $\sum_{n\geq 1}\frac{X_n}{a_n}$ converges a.e.

Exercise: do the proof of the theorem.

Strong Law of Large Numbers (Kolmogorov)

Theorem

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed r.v., then

$$\mathbb{E}[|X_1|] < \infty \Rightarrow \frac{S_n}{n} \to \mathbb{E}[X_1] \quad a.e.$$

$$E[|X_1|] = \infty \Rightarrow \limsup_n \frac{|S_n|}{n} = +\infty \quad a.e.$$

Exercise: do the proof of the theorem.

15th Lecture: Characteristic functions

Characteristic functions

Definition

For any r.v. X with probability measure μ and distribution function F, the characteristic function of X is defined as the function $\varphi : \mathbb{R} \to \mathbb{C}$ given by

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int_\Omega e^{itX} d\mathbb{P} = \int_\mathbb{R} e^{itx}\mu(dx) = \int_\mathbb{R} e^{itx} dF(x).$$

Note that the real and imaginary parts of φ_X are given, respectively by

$$Re\varphi(t) = \int_\mathbb{R} \cos(tx)\mu(dx) \quad \text{and} \quad Im\varphi(t) = \int_\mathbb{R} \sin(tx)\mu(dx).$$

Properties of the characteristic function

- $\forall t \in \mathbb{R}$: $|\varphi(t)| \leq 1 = \varphi(0)$.
- $\forall t \in \mathbb{R}$: $\varphi(t) = \overline{\varphi(-t)}$.
- φ is uniformly continuous.
- If φ_X is the c.f. for a r.v. X, then

$$\varphi_{aX+b}(t) = \varphi_X(at)e^{itb} \quad \text{and} \quad \varphi_{-X}(t) = \overline{\varphi_X(t)}.$$

- If $\{\varphi_n\}_{n\geq 1}$ is a sequence of characteristic functions, $\lambda_n \geq 0$ with $\sum_{n\geq 1} \lambda_n = 1$, then $\sum_{n\geq 1} \lambda_n \varphi_n$ is a characteristic function.
- If $\{\varphi_n\}_{n\geq 1}$ is a sequence of characteristic functions, then $\prod_{j=1}^{n} \varphi_j$ is a characteristic function.

Exercise: do the proof of the properties.

The sum of random variables

Let $S_n = X_1 + \cdots + X_n$. Then $\varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{X_j}(t)$.
What can we say about the distribution of S_n?

Definition

The convolution of two distribution functions F_1 and F_2 is the distribution function F defined on $x \in \mathbb{R}$ as $F(x) = \int_{\mathbb{R}} F_1(x-y)dF_2(y)$. In this case we use the notation $F = F_1 * F_2$.

Theorem

Let X and Y be two independent r.v. with distribution functions F_X and F_Y respectively. Then $X+Y$ has distribution function $F_X * F_Y$.

Exercise: do the proof of the theorem.

The convolution

Definition

The convolution of two probability density functions f_1 and f_2 is the probability density function f defined on $x \in \mathbb{R}$ as $f(x) = \int_{\mathbb{R}} f_1(x-y) f_2(y) dy$. In this case we also use the notation $f = f_1 * f_2$.

Theorem

The convolution of two absolutely continuous distribution functions F_1 and F_2 with densities f_1 and f_2, is absolutely continuous with density $f = f_1 * f_2$.

Exercise: do the proof of the theorem.

And what is the probability measure corresponding to $F_1 * F_2$?
We shall denote this measure by $\mu_1 * \mu_2$.

Theorem

For each $B \in \mathcal{B}$ we have that

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} \mu_1(B - y)\mu_2(dy),$$

where the set $B - y = \{x - y : x \in B\}$. Moreover, for each $\mathcal{B}$-measurable funtion g integrable wrt $\mu_1 * \mu_2$, we have that

$$\int_{\mathbb{R}} g(u)\mu_1 * \mu_2(du) = \iint_{R^2} g(x+y)\mu_1(dx)\mu_2(dy).$$

Exercise: do the proof of the theorem.

Theorem

The sum of a finite number of independent r.v. corresponds to the convolution of their distribution functions and to the product of their characteristic functions.

Lemma

If φ is a characteristic function, then $|\varphi|^2$ is a characteristic function.

Example

For $X \sim \text{Ber}(p)$ we have that $\varphi_X(t) = e^{it}p + (1-p)$.

For $X \sim \text{Poisson}(\lambda)$ we have that $\varphi_X(t) = e^{\lambda}(e^{it} - 1)$.

For $X \sim U[-a,a]$ we have that $\varphi_X(t) = \frac{sin(at)}{at}$, if $t \neq 0$ and $\varphi_X(0) = 1$.

For $X \sim N(\mu, \sigma^2)$, we have that $\varphi_X(t) = e^{it\mu}e^{-\frac{\sigma^2 t^2}{2}}$.

Exercise: do the proof of the results above.

Applications with the convolution

Example

Exercises:

1) Show that for X and Y independent r.v. with $X \sim N(0,1)$ and $Y \sim N(0,1)$ we have that $X+Y \sim N(0,2)$.

2) Show that for X and Y independent r.v. with $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$ we have that $X+Y \sim Poisson(\lambda_1+\lambda_2)$.

3) Show that for X and Y independent r.v. with $X \sim B(n,p)$ and $Y \sim B(m,p)$ we have that $X+Y \sim B(n+m,p)$.

16th Lecture: Inversion formula and uniqueness

Inversion formula

The question now is: Given a characteristic function how can we find the correspondent distribution function or the distribution measure?

Theorem (The characteristic function determines the distribution)

If $x_1 < x_2$, then:

$\mu((x_1,x_2)) + \frac{1}{2}\mu(\{x_1\}) + \frac{1}{2}\mu(\{x_2\}) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1} - e^{itx_2}}{it}\varphi(t)\,dt.$

**Note that the integrand function is defined by continuity at $t = 0$.*

Remark

Note that if (x_1,x_2) is a continuity interval for μ, then the previous theorem says that $F(x_2) - F(x_1) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1} - e^{itx_2}}{it}\varphi(t)\,dt.$

Exercise: do the proof of the theorem.

Uniqueness of distribution

Theorem

If two probability measures (or two distribution functions) have the same characteristic function, then the probability measures (or the distribution functions) are the same.

Theorem

If $\varphi \in L^1(\mathbb{R})$, *then* $F \in C^1$ *and* $F'(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\varphi(t)dt$, *that is* φ *is the characteristic function of an absolutely continuous r.v.*

Corollary

If $\varphi \in L^1(\mathbb{R})$, *then* $p(x) \in L^1(\mathbb{R})$ *where* $p(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-ixt}\varphi(t)dt$ *and* $\varphi(t) = \int_{-\infty}^{\infty} e^{itx}p(x)\,dx$.

Exercise: do the proof of the results.

17th Lecture: Converging Theorems and Applications

The atoms of μ

Theorem

- *For each x_0 we have that*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itx_0}\varphi(t)\,dt=\mu(\{x_0\}).$$

- *It holds that* $\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\varphi(t)|^2\,dt=\sum_{x\in\mathbb{R}}(\mu(\{x\}))^2$.

Theorem

μ is atomless (F is continuous) iff $\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\varphi(t)|^2\,dt=0$.

Exercise: do the proof of the results above.

Symmetric random variables

Definition

A r.v. X is sym. around 0 iff X and $-X$ have the same distribution.

Remark

For a symmetric r.v. its distribution μ has the following property $\mu(B) = \mu(-B)$ for any $B \in \mathcal{B}$. Such probability measure is said to be symmetric around 0. Equivalently, for the distribution function, we have that for any $x \in \mathbb{R}$, $F(x) = 1 - F(-x^-)$.

Theorem

A r.v. X or a p.m. μ is symmetric iff its characteristic function is real-valued (for all t.)

Exercise: do the proof of the theorem.

Convergence theorems

Theorem (Lévy's converging Theorem)

Let $\{\mu_n\}_{n\geq 1}$ be probability measures on $\mathbb{R}$ with characteristic function $\{\varphi_n\}_{n\geq 1}$.

- *If μ_∞ is a probability measure on $\mathbb{R}$ and $\mu_n \to^v \mu_\infty$, then $\varphi_n(t) \to_{n\to\infty} \varphi_\infty(t)$, where φ_∞ is the characteristic function of μ_∞.*
- *If $\varphi_n(t) \to_{n\to\infty} \varphi_\infty(t)$ for all $t \in \mathbb{R}$, and $\varphi_\infty(t)$ is continuous at $t = 0$, then*
    - *$\mu_n \to^v \mu_\infty$ where μ_∞ is a probability measure,*
    - *φ_∞ is the characteristic function of μ_∞.*

Exercise: do the proof of the theorem.

Corollary

If $\{\mu_n\}_{n\geq 1}$ and μ are probability measures with characteristic functions $\{\varphi_n\}_{n\geq 1}$ and φ, then $\mu_n \to^v \mu_\infty$ iff $\varphi_n(t) \to_{n\to\infty} \varphi(t)$, for all $t \in \mathbb{R}$.

Exercise: do the proof of the corollary.

Example

Exercises:

1) Take μ_n which gives mass $1/2$ to 0 and to n and analyze it.

2) Take μ_n as Uniform in $[-n, n]$ and analyze it.

Applications

Theorem

If F has finite absolute moment of order k, with $k \geq 1$, then φ has the following expansion around a neighbourhood of $t = 0$:

$$\varphi(t) = \sum_{j=0}^{k} \frac{i^j}{j!} m^j t^j + o(|t|^k)$$

$$\varphi(t) = \sum_{j=0}^{k-1} \frac{i^j}{j!} m^j t^j + \frac{\theta_k}{k!} \mu^k |t|^k,$$

where m^j is the moment of order j, μ^k is the absolute moment of order k and $\theta_k \leq 1$.

Applications: limiting laws

Below $\{X_n\}_{n\geq 1}$ are i.i.d. r.v. with distribution function F and $S_n = \sum_{j=1}^{n} X_j$.

Theorem (The weak law of large numbers)

If F has finite mean $m < \infty$, then $\frac{S_n}{n} \to m$ in probability.

Theorem (The central limit theorem)

If F has finite mean $m < \infty$ and variance σ^2 such that $0 < \sigma^2 < +\infty$, then $\frac{S_n - mn}{\sigma\sqrt{n}} \to \Phi$ in distribution, where Φ is the distribution function of $N(0,1)$.

Exercise: do the proof of the results above.

18th Lecture: Conditional expectation

Definition (Conditional probability)

Given a set $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ we define $\mathbb{P}_A(\cdot)$ in the following way:

$$\mathbb{P}_A(E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(A)}.$$

$\mathbb{P}_A$ is a probability measure and it is called the conditional probability with respect to A. The expectation with respect to this probability is called the conditional expectation wrt A:

$$\mathbb{E}_A[X] = \int_\Omega X(\omega)\mathbb{P}_A(d\omega) = \frac{1}{\mathbb{P}(A)} \int_A X(\omega)\mathbb{P}(d\omega).$$

Definition

If we take now a partition of Ω that is $(A_n)_{n\geq 1}$ with $\Omega = \cup_{n\geq 1} A_n$, $A_n \in \mathcal{F}$ and $A_n \cap A_m = \varnothing$ if $m \neq n$, then given a set $E \in \mathcal{F}$ we have that

$$\mathbb{P}(E) = \sum_{n\geq 1} \mathbb{P}(E \cap A_n) = \sum_{n\geq 1} \mathbb{P}_{A_n}(E)\mathbb{P}(A_n).$$

Definition

As above we have that (if $\mathbb{E}[X]$ is finite)

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int \cup_{n\geq 1} A_n X(\omega)\mathbb{P}(d\omega)$$
$$= \sum_{n\geq 1} \int_{A_n} X(\omega)\mathbb{P}(d\omega) = \sum_{n\geq 1} \mathbb{P}(A_n)\mathbb{E}_{A_n}[X].$$

Example

Suppose that we have a card deck with 52 cards and that we take one out and it is spades. What is the probability of taking another card of the deck and that it is also spades?

Wald's equation

Theorem

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d.r.v. with finite mean. For $k \geq 1$ let $\mathcal{F}_k$ be the σ-algebra generated by X_j with $j = 1, \cdots, k$. Suppose that N is a random variable taking positive integer values such that for all $k \geq 1$ we have that $\{N \leq k\} \in \mathcal{F}_k$ and $\mathbb{E}[N] < \infty$. Then $\mathbb{E}[S_N] = \mathbb{E}[X_1]\mathbb{E}[N]$.

To prove it note that

$$\mathbb{E}[S_N] = \int_\Omega S_N \mathbb{P}(d\omega) = \int_{\{N\geq 1\}} S_N \mathbb{P}(d\omega) = \sum_{k\geq 1} \int_{\{N=k\}} S_N \mathbb{P}(d\omega)$$

$$= \sum_{k\geq 1} \sum_{j=1}^{k} \int_{\{N=k\}} X_j \mathbb{P}(d\omega) = \sum_{j\geq 1} \sum_{k\geq j} \int_{\{N=k\}} X_j \mathbb{P}(d\omega)$$

$$= \sum_{j\geq 1} \int_{\{N\geq j\}} X_j \mathbb{P}(d\omega) = \sum_{j\geq 1} \Big(\mathbb{E}[X_j] - \int_{\{N\leq j-1\}} X_j \mathbb{P}(d\omega) \Big).$$

Now we note that the set $\{N \leq j-1\}$ and the r.v. X_j are independent (remember that $\{N \leq j-1\} \in \mathcal{F}_{j-1}$ and note the definition of $\mathcal{F}_{j-1}$), therefore we get

$$\mathbb{E}[S_N] = \sum_{j\geq 1} \mathbb{E}[X_j]\mathbb{P}(N \geq j) = \mathbb{E}[X_1] \sum_{j\geq 1} \mathbb{P}(N \geq j) = \mathbb{E}[X_1]\mathbb{E}[N].$$

To justify that we can interchange summations we have to repeat the computations taking $|X_j|$ and we will see that we get the result $\mathbb{E}[|X_1|]\mathbb{E}[N]$ which is finite by hypothesis.

Now, let X be a discrete r.v. and let $A_n = \{X = a_n\}$. Given an integrable r.v. Y we define the function $\mathbb{E}[Y|\mathcal{G}]$ in Ω as

$$\mathbb{E}[Y|\mathcal{G}] = \sum_{n\geq 1} \mathbf{1}_{A_n}(\cdot)\mathbb{E}[Y|A_n],$$

this means that $\mathbb{E}[Y|\mathcal{G}]$ is a discrete r.v. that takes the value $\mathbb{E}[Y|A_n]$ on the set A_n.

We can rewrite the expression above as

$$\mathbb{E}[Y] = \sum_{n\geq 1} \int_{A_n} \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega) = \int_{\Omega} \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$$

Analogously for any $A \in \mathcal{G}$, A is a union of subcolletion of the A_n's, so that, for every $A \in \mathcal{G}$ we have that

$$\int_A Y\mathbb{P}(d\omega) = \int_A \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$$

Attention to the measurability of the functions involved.

Now, we suppose that we have two functions φ_1 and φ_2 both $\mathcal{G}$ measurable and such that

$$\int_A Y\mathbb{P}(d\omega) = \int_A \varphi_1\mathbb{P}(d\omega) = \int_A \varphi_2\mathbb{P}(d\omega).$$

If we take the set $A = \{\omega \in \Omega : \varphi_1(\omega) > \varphi_2(\omega)\}$, then $A \in \mathcal{G}$ and we conclude that $\mathbb{P}(A) = 0$. Repeating the argument exchanging φ_1 with φ_2 we conclude that $\varphi_1 = \varphi_2$ a.e.

This means that $\mathbb{E}[Y|\mathcal{G}]$ is unique up to a equivalence and we are going to denote $\mathbb{E}_{\mathcal{G}}[Y]$ or $\mathbb{E}[Y|\mathcal{G}]$ to denote that class. The results holds for any σ-algebra.

Theorem

If $\mathbb{E}[|Y|] < \infty$ and $\mathcal{G}$ is a σ-algebra contained in $\mathcal{F}$, then, there exists a unique equivalence class of integrable r.v. $\mathbb{E}[Y|\mathcal{G}]$ belonging to $\mathcal{G}$ such that for any $A \in \mathcal{G}$ it holds that $\int_A Y\mathbb{P}(d\omega) = \int_A \mathbb{E}[Y|\mathcal{G}]\mathbb{P}(d\omega)$.

Conditional expectation

Definition (Conditional expectation)

Given an integrable r.v. Y and a σ-algebra $\mathcal{G}$, the conditional expectation $\mathbb{E}_{\mathcal{G}}[Y]$ of Y with respect to $\mathcal{G}$ is any one of the equivalence class of r.v. on Ω such that:

1. it belongs to $\mathcal{G}$;
2. it has the same integral as Y over any set in $\mathcal{G}$.

Note that for $Y = \mathbf{1}_\Lambda$ with $\Lambda \in \mathcal{F}$ we write $\mathbb{P}(\Lambda|\mathcal{G}) = \mathbb{E}[\mathbf{1}_\Lambda|\mathcal{G}]$ and this is the conditional probability of Λ relatively to $\mathcal{G}$. This is any one of the equivalence class of r.v. belonging to $\mathcal{G}$ and satisfying

$$\forall B \in \mathcal{G} : \mathbb{P}(B \cap \Lambda) = \int_B \mathbb{P}(\Lambda|\mathcal{G})\mathbb{P}(d\omega).$$

Conditional expectation

Theorem

Let Y and ZY be integrable r.v. and let $Z \in \mathcal{G}$. Then

$$\mathbb{E}[YZ|\mathcal{G}] = Z\mathbb{E}[Y|\mathcal{G}], \quad a.e.$$

Exercise: do the proof of the theorem.

Let us note that $\mathbb{E}[X|\mathcal{T}] = \mathbb{E}[X]$, where $\mathcal{T}$ is the trivial σ-algebra, that is $\mathcal{T} := \{\varnothing, \Omega\}$.

Properties of the conditional expectation

Let X and X_n be integrable r.v.

1. If $X \in \mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = X$ a.e., this is true also if $X = a$ a.e.,
2. $\mathbb{E}[X_1 + X_2|\mathcal{G}] = \mathbb{E}[X_1|\mathcal{G}] + \mathbb{E}[X_2|\mathcal{G}]$,
3. If $X_1 \leq X_2$ then $\mathbb{E}[X_1|\mathcal{G}] \leq \mathbb{E}[X_2|\mathcal{G}]$,
4. $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$,
5. If $X_n \uparrow X$, then $\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}]$,
6. If $X_n \downarrow X$, then $\mathbb{E}[X_n|\mathcal{G}] \downarrow \mathbb{E}[X|\mathcal{G}]$,
7. If $|X_n| \leq Y$, Y is integrable and $X_n \to X$, then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$,
8. $\mathbb{E}[|XY||\mathcal{G}]^2 \leq \mathbb{E}[X^2|\mathcal{G}]\mathbb{E}[Y^2|\mathcal{G}]$. (Cauchy-Schwarz inequality)

Exercise: do the proof.

19th Lecture: Fundamental Theorems

Jensen's inequality

Theorem (Jensen's inequality)

If φ is a convex function on $\mathbb{R}$ and X and $\varphi(X)$ are integrable r.v., then for each $\mathcal{G}$:

$$\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}].$$

Exercise: do the proof.

Note that when $\Lambda = \Omega$, the defining relation for the conditional expectation says that

$$\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]|\mathcal{T}] = \mathbb{E}[Y|\mathcal{T}] = \mathbb{E}[\mathbb{E}[Y|\mathcal{T}]|\mathcal{G}]$$

This can be generalized and it is called the tower law.

Tower Law

Theorem (Tower law)

If Y is integrable and $\mathcal{F}_1 \subset \mathcal{F}_2$, then:

- $\mathbb{E}[Y|\mathcal{F}_1] = \mathbb{E}[Y|\mathcal{F}_2]$ *iff* $\mathbb{E}[Y|\mathcal{F}_2] \in \mathcal{F}_1$.
- $\mathbb{E}[\mathbb{E}[Y|\mathcal{F}_2]|\mathcal{F}_1] = \mathbb{E}[Y|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_1]|\mathcal{F}_2]$

As a particular case we note that

$$\mathbb{E}[\mathbb{E}[Y|X_1, X_2]|X_1] = \mathbb{E}[Y|X_1] = \mathbb{E}[\mathbb{E}[Y|X_1]|X_1, X_2].$$

Tower Law: proof of the theorem

Let us prove the theorem now.

We start with the first assertion. Let start by assuming that $\mathbb{E}[Y|\mathcal{F}_1] = \mathbb{E}[Y|\mathcal{F}_2]$, then by 1) in page 128 we have that $\mathbb{E}[Y|\mathcal{F}_2] \in \mathcal{F}_1$. Now let us assume that $\mathbb{E}[Y|\mathcal{F}_2] \in \mathcal{F}_1$. Then, for $A \in \mathcal{F}_1$, 2) in page 128 holds, from where the result follows.

Now let us prove the second assertion. Note that $\mathbb{E}[Y|\mathcal{F}_1] \in \mathcal{F}_2$, and from the first assertion applied to $\mathbb{E}[Y|\mathcal{F}_1]$ we conclude the second equality. Let us prove the first equality now. For that purpose note that if $\Lambda \in \mathcal{F}_1$ then $\Lambda \in \mathcal{F}_2$, so that

$$\int_\Lambda \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_2]|\mathcal{F}_1]\mathbb{P}(d\omega) = \int_\Lambda \mathbb{E}[Y|\mathcal{F}_2]\mathbb{P}(d\omega) = \int_\Lambda Y\mathbb{P}(d\omega).$$

Moreover, $\mathbb{E}[\mathbb{E}[Y|\mathcal{F}_2]|\mathcal{F}_1] \in \mathcal{F}_1$ so that, both properties defining the conditional expectation are verified and we are done.

Conditional independence

Let $\mathcal{F}$ be a σ-algebra and let $\{\mathcal{F}_\alpha\}_{\alpha\in A}$, where A is a index set, be contained in $\mathcal{F}$.

Definition

The collection $\{\mathcal{F}_\alpha\}_{\alpha\in A}$ is said to be conditionally independent to a σ-algebra $\mathcal{G}$ iff for any finite collection of sets $A_1,\cdots,A_n$ with $A_j\in\mathcal{F}_j$ and with $\alpha_j's$ distinct indices of A we have

$$\mathbb{P}\Big(\cap_{j=1}^n A_j|\mathcal{G}\Big)=\prod_{j=1}^n\mathbb{P}(A_j|\mathcal{G}).$$

Note that if $\mathcal{G}=\mathcal{T}$ then the previous condition is just the usual independence.

Theorem

For each $\alpha \in A$*, let* $\mathcal{F}^{(\alpha)}$ *be the smallest* σ*-algebra containing all* $\mathcal{F}_\beta$ *with* $\beta \in A \backslash \{\alpha\}$*. Then, the* $\mathcal{F}_\alpha$*'s are conditionally independent relatively to a* σ*-algebra* $\mathcal{G}$ *iff for each* α *and* $A_\alpha \in \mathcal{F}_\alpha$ *we have*

$$\mathbb{P}\Big(A_\alpha | \mathcal{F}^{(\alpha)} \vee \mathcal{G}\Big) = \mathbb{P}(A_\alpha | \mathcal{G}),$$

where $\mathcal{F}^{(\alpha)} \vee \mathcal{G}$ *denotes the smallest* σ*-algebra containing* $\mathcal{F}^{(\alpha)}$ *and* $\mathcal{G}$*.*

Note that if in the previous theorem $\mathcal{G} = \mathcal{T}$ and $\mathcal{F}_\alpha$ is generated by a r.v. say X_α then we have

Corollary

Let $(X_\alpha)_{\alpha \in A}$ *be a collection of r.v. and for each* α *let* $\mathcal{F}^{(\alpha)}$ *be the* σ*-algebra generated by all the r.v. except by* X_α*. Then, the r.v.* X_α*'s are independent iff for each* α *and* $B \in \mathcal{B}$ *we have*

$$\mathbb{P}(X_\alpha \in B | \mathcal{F}^{(\alpha)}) = \mathbb{P}(X_\alpha \in B) \quad a.e.$$

Now, let X_1 and X_2 be two independent r.v. What happens if we condition $X_1 + X_2$ by X_1?

Theorem

Let X_1 and X_2 be two independent r.v. with probability measures μ_1 and μ_2, respectively. Then, for each $B \in \mathcal{B}$:

$$\mathbb{P}(X_1 + X_2 \in B | X_1) = \mathbb{P}(X_1 + X_2 \in B | \mathcal{F}_1) = \mu_2(B - X_1) \quad a.e.$$

where $\mathcal{F}_1$ is the $\sigma-$algebra generated by X_1.
More generally, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent r.v. with probability measures $(\mu_n)_{n\in\mathbb{N}}$ and let $S_n = X_1 + \cdots + X_n$. Then, for each $B \in \mathcal{B}$:

$$\mathbb{P}(S_n \in B | S_1, \cdots, S_{n-1}) = \mu_n(B - S_{n_1}) = \mathbb{P}(S_n \in B | S_{n-1}) \quad a.e.$$

Exercise: Prove all the results above.

Let us look quickly at the proof of the previous theorem.
Remember that $\mathbb{P}(X_1+X_2\in B|X_1)=\mathbb{E}[\mathbf{1}_{\{X_1+X_2\in B\}}|X_1]$. Now using the Theorem of page 76 we have that, for $\Lambda\in\mathcal{F}_1$ (note that this set is such that $\Lambda=X_1^{-1}(A)$, where $A\in\mathcal{B}$, to prove this use the trick with monotone classes, see the Theorem in page 4)

$$\int_\Lambda \mu_2(B-X_1)\mathbb{P}(d\omega)=\int_A \mu_2(B-x_1)\mu_1(dx_1)$$

$$=\int_A \mu_1(dx_1)\int_\Omega \mathbf{1}_{\{x_1+x_2\in B\}}\mu_2(dx_2)=\int\int_{\{x_1\in A,x_1+x_2\in B\}}\mu_1\times\mu_2(dx_1,dx_2)$$

$$=\int\int_{\{X_1\in A,X_1+X_2\in B\}}\mathbb{P}(d\omega)=\mathbb{P}(X_1\in A,X_1+X_2\in B)$$

$$=\int_\Lambda \mathbf{1}_{\{X_1+X_2\in B\}}\mathbb{P}(d\omega).$$

Since $\mu_2(B-X_1)\in\mathcal{F}_1$ and since the previous relation is true for any $\Lambda\in\mathcal{F}_1$, the result follows. As an exercise, prove the second assertion of the theorem.

Conditional distribution of X given a set A.

Given a r.v. X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and for an event A with $\mathbb{P}(A) > 0$ we define the conditional distribution of X given A as:

$$\mathbb{P}(X \in B|A) = \frac{\mathbb{P}((X \in B) \cap A)}{\mathbb{P}(A)}.$$

Exercise: Check that this gives a probability measure on the Borel σ-algebra.

Now, we can define the conditional distribution function of X given the set A on $x \in \mathbb{R}$ as

$$F_X(x|A) = \mathbb{P}(X \leq x|A) = \frac{\mathbb{P}((X \leq x) \cap A)}{\mathbb{P}(A)}$$

The conditional expectation of X given the set A is the expectation of the conditional distribution given by

$$\mathbb{E}[X|A] = \int x \, dF_X(x|A)$$

if it exists. As above, if we take now a partition of Ω that is $(A_n)_{n\geq 1}$ with $\Omega = \cup_{n\geq 1} A_n$, $A_n \in \mathcal{F}$ and $A_n \cap A_m = \varnothing$ if $m \neq n$, then

$$\mathbb{P}(X \in B) = \sum_{n\geq 1} \mathbb{P}(X \in B|A_n)\mathbb{P}(A_n).$$

Also for any x,
$F_X(x) = \mathbb{P}(X \leq x) = \sum_{n\geq 1} \mathbb{P}(X \leq x|A_n)\mathbb{P}(A_n) = \sum_{n\geq 1} F_X(x|A_n)\mathbb{P}(A_n)$
and analogously

$$\mathbb{E}[X] = \int x dF_X(x) = \sum_{n\geq 1} \mathbb{P}(A_n)\mathbb{E}[X|A_n].$$

1) Let $X \sim U[-1,1]$ and let $A = \{X \geq 0\}$. What is the conditional distribution of X given A?

Conditional distribution of X given a discrete r.v. Y

Let us suppose now that the partition is generated by a discrete r.v. Let Y be a discrete r.v. defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking the values $(a_n)_{n\in\mathbb{N}}$. Then the events $\{Y = a_n\}$ form a partition of Ω. In this case $\mathbb{P}(X \in B|Y = a_n)$ is called the conditional distribution of X given $Y = a_n$ and we have that

$$\mathbb{P}(X \in B|Y = a_n) = \sum_{n\geq 1} \mathbb{P}(X \in B|Y = a_n)\mathbb{P}(Y = a_n).$$

Also for any x,

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leq x) &= \sum_{n\geq 1} \mathbb{P}(X \leq x|Y = a_n)\mathbb{P}(Y = a_n) \\ &= \sum_{n\geq 1} F_X(x|Y = a_n)\mathbb{P}(X = a_n) \end{aligned}$$

and analogously $\mathbb{E}[X] = \int x dF_X(x) = \sum_{n\geq 1} \mathbb{P}(Y = a_n)\mathbb{E}[X|Y = a_n]$.

Note that for B fixed we have that $\mathbb{P}(X \in B|Y = a_n)$ is a function of a_n let us say $g(a_n)$. Defining $g(y) = \mathbb{P}(X \in B|Y = y)$ we have that $\mathbb{P}(X \in B) = \int \mathbb{P}(X \in B|Y = y)dF_Y(y) = \int g(y)dF_Y(y)$. Moreover,

$$F_X(x) = \int F_X(x|Y = y)dF_Y(y) \qquad \mathbb{E}[X] = \int \mathbb{E}[X|Y = y]dF_Y(y).$$

When X is integrable the function $\varphi(y) = \mathbb{E}[X|Y = y]$ is finite. In this case, the r.v. $\varphi(Y)$ is called the conditional expectation of X given Y: $\varphi(Y) = \mathbb{E}[X|Y]$. We note that $\mathbb{E}[X|Y = y]$ is the value of the random variable $\mathbb{E}[X|Y]$ when $Y = y$. The last formula can be interpreted as

$$\mathbb{E}[X] = \mathbb{E}[\varphi(Y)] = \mathbb{E}[\mathbb{E}[X|Y]].$$

2) Consider the following experience: a player tosses a fair coin n times obtaining k heads with $0 \leq k \leq n$. After that a second player tosses the same coin k times. Let X be the number of heads obtained by the second player. What is the expectation of X supposing that all the events are independent?

Conditional distribution: general case

Let us define now the conditional expectation for general r.v. X and Y. Before we defined the conditional distribution of X when Y was discrete, so that $\mathbb{P}(Y=y)=0$ for all $y\neq a_n$. But now we want to extend this to the continuous case in which the probability above is null for all $y\in\mathbb{R}$. How to do it? We define by approximation. Take I an interval containing y with size Δy and define

$$\mathbb{P}(X\in B|Y=y)\sim\mathbb{P}(X\in B|Y\in I)=\frac{\mathbb{P}(X\in B,Y\in I)}{\mathbb{P}(Y\in I)}.$$

If $\mathbb{P}(X\in B|Y\in I)$ has a limit when $\Delta y\to 0$ we call to the limit $\mathbb{P}(X\in B|Y=y)$:

$$\lim_{\Delta y\to 0}\mathbb{P}(X\in B|Y\in I)=\mathbb{P}(X\in B|Y=y).$$

Let us go back to the case in which X is discrete.
Then we have

$$
\begin{aligned}
F_{(X,Y)}(x,y) &= \mathbb{P}(X \leq x, Y \leq y) = \sum_{n:a_n \leq y} \mathbb{P}(X \leq x, Y = a_n) \\
&= \sum_{n:a_n \leq y} \mathbb{P}(X \leq x | Y = a_n)\mathbb{P}(Y = a_n) \\
&= \sum_{n:a_n \leq y} F_X(x | Y = a_n)\mathbb{P}(Y = a_n) \\
&= \int_{-\infty}^{y} F_X(x | Y = a) dF_Y(a).
\end{aligned}
$$

Note that in the discrete case, the joint distribution is like a composition of the marginal distribution of Y with the conditional distribution of X given Y. Let use then the last equality!

Formal definitions

Definition

Let X and Y be two r.v. defined on the same probability space. A function $\mathbb{P}(X \in B|Y = y)$ defined for each borelian B and $y \in \mathbb{R}$ is a (regular) conditional distribution for X given Y if:

1. for each y fixed, $\mathbb{P}(X \in B|Y = y)$ defines a probability measure in $\mathcal{B}$,
2. for any $B \in \mathcal{B}$ fixed, $\mathbb{P}(X \in B|Y = y)$ is a measurable function of y,
3. for any $(x, y) \in \mathbb{R}^2$ it holds that

$$\mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{y} F_X(x|Y = a)dF_Y(a).$$

$\mathbb{P}(X \in B|Y = y)$ is called the conditional probability of X belonging to B given that $Y = y$ and $F_X(\cdot|Y = y) = \mathbb{P}(X \leq \cdot|Y = y)$ is the conditional distribution of X given $Y = y$.

Formal definitions

Theorem

Let X and Y be two r.v. defined on the same probability space. There exists a (regular) conditional distribution for X given Y. In fact there exists only one in the sense that they are equal a.e.: that is, if $\mathbb{P}_1(X \in B|Y = y)$ and $\mathbb{P}_2(X \in B|Y = y)$ are conditional distributions for X given Y, then there exists a borelian B_0 such that $\mathbb{P}(Y \in B_0) = 1$ and $\mathbb{P}_1(X \in B|Y = y) = \mathbb{P}_2(X \in B|Y = y)$ for all $B \in \mathcal{B}$ and $y \in B_0$.

Theorem

For each $B \in \mathcal{B}$ fixed, the limit

$$\lim_{\Delta a \to 0} \mathbb{P}(X \in B|Y \in I) = \mathbb{P}(X \in B|Y = a)$$

exists a.e. Moreover, for each $B \in \mathcal{B}$ fixed, the limit is equal to $\mathbb{P}(X \in B|X = y)$ as given in the definition above, a.e.

1) What is the conditional distribution of Y given Y? Let us guess it. If it is given that $Y=y$, then $Y=y$! So the candidate is $\mathbb{P}(Y=y|Y=y)=1$ the distribution which gives weight 1 to the point y. Check that for $B=(q_1,q_2)$ with $q_i\in\mathbb{Q}$ it holds that

$$\mathbb{P}(Y\in B|Y=y)=\lim_{\Delta a\to 0}\mathbb{P}(Y\in B|Y\in I),$$

which proves the result.
Note however that if we take $B=\{y_0\}$ then

$$\mathbb{P}(Y=y_0|Y=y)=\lim_{\Delta a\to 0}\mathbb{P}(Y=y_0|Y\in I)=0!$$

This does not contradict our result but contradicts our intuition!
2) Given $Y=y$ what is the conditional distribution of $Z=g(Y)$? Recall that above we have seen that if $Y=y$, then $\mathbb{P}(Y=y|Y=y)=1$. Here it is analogous. In this case we have that $\mathbb{P}(g(Y)=g(y)|Y=y)=1$.

3) Let X be a symmetric r.v. around 0. What is the conditional distribution of X given the r.v. $|X|$? Given that $|X| = y > 0$, then $X = y$ or $-y$, there are no other possibilities and from symmetry we have that:

$$\mathbb{P}(X = y||X| = y) = \frac{1}{2} = \mathbb{P}(X = -y||X| = y), \quad y > 0,$$

and $\mathbb{P}(X = 0||X| = 0) = 1$.
Let us do it now in a different way. Suppose $y > 0$. Take $B = (q_1, a_2)$ with $q_i \in \mathbb{Q}$ and take $I \subset B$. Then

$$\mathbb{P}(X \in B||X| \in I) = \mathbb{P}(X \in I) = \frac{1}{2}\Big(\mathbb{P}(X \in I) + \mathbb{P}(X \in -I)\Big) = \frac{1}{2}\mathbb{P}(|X| \in I).$$

And

$$\mathbb{P}(X \in -B||X| \in I) = \mathbb{P}(X \in -I) = \frac{1}{2}\mathbb{P}(|X| \in I).$$

Since $I \subset B$ we have that

$$\mathbb{P}(X \in B||X| \in I) = \frac{1}{2} = \mathbb{P}(X \in -B||X| \in I).$$

Therefore,

$$\mathbb{P}(X \in B||X| = y) = \lim_{\Delta y \to 0} \mathbb{P}(X \in B||X| \in I) = \frac{1}{2},$$

$$\mathbb{P}(X \in -B||X| = y) = \lim_{\Delta y \to 0} \mathbb{P}(X \in -B||X| \in I) = \frac{1}{2}.$$

Taking B decreasing to $\{y\}$ we see that the conditional probability gives weight $1/2$ to each one of the points y and $-y$. The proof that $\mathbb{P}(X = 0||X| = 0) = 1$ can be reached by taking $B = (q_1, q_2)$ as above with $q_1 < 0 < q_2$.

4) Let X and Y be independent r.v. each one with law $N(0,\sigma^2)$ with $\sigma^2 > 0$. What is the conditional distribution of (X,Y) given $\sqrt{X^2+Y^2}$? For $z > 0$, $\sqrt{X^2+Y^2} = z$ iff (X,Y) is in the circle centered at $(0,0)$ with radius z. Therefore the conditional distribution is concentrated in that circle, that is, in the set of points of $\mathbb{R}^2$ given by
$\mathcal{C} := \{(x,y) : x^2+y^2 = z\}$.
Note that the joint density function of (X,Y) is given by

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)^2}{2\sigma^2}}.$$

Note that the density is constant on the circle $\mathcal{C}$. Therefore, before the experience all the points in the circle $\mathcal{C}$ had the same "chance" and our guess for the distribution is the uniform distribution on the circle, that is, for $B \in \mathcal{B}^2$ and $z > 0$:

$$\mathbb{P}((X,Y) \in B | \sqrt{X^2+Y^2} = z) = \frac{\text{"size of"}(B \cap \mathcal{C})}{2\pi z}.$$

Prove it!

20th Lecture: Discrete time martingales

Martingales

Let $(X_n)_{n\in\mathbb{N}}$ be independent r.v. with mean zero and let $S_n = \sum_{j=1}^n X_j$. Then

$$\begin{aligned}\mathbb{E}[S_{n+1}|X_1,\cdots,X_n] &= \mathbb{E}[X_1+\cdots+X_n+X_{n+1}|X_1,\cdots,X_n]\\ &= S_n+\mathbb{E}[X_{n+1}|X_1,\cdots,X_n] = S_n+\mathbb{E}[X_{n+1}]\\ &= S_n.\end{aligned}$$

Historically, the equation above gave rise to consider dependent r.v. which satisfy $\mathbb{E}[X_{n+1}|X_1,\cdots,X_n]=0$ and this opened a way to define a class of stochastic processes which are extremely useful - **the martingales**.

Martingales

Definition (smartingale: martingale, submartingale, supermartingale)

The sequence of r.v. and $\sigma-$algebras $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is said to be a martingale iff for each $n \in \mathbb{N}$ we have that

1. $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ and $X_n \in \mathcal{F}_n$, (this means that X_n is adapted to $\mathcal{F}_n$)
2. $\mathbb{E}[|X_n|] < \infty$ for each $n \in \mathbb{N}$, (this means that X_n is integrable)
3. for each $n \in \mathbb{N}$, we have that

$$X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n] \quad \text{a.e.} \quad (\textit{martingale})$$
$$X_n \leq \mathbb{E}[X_{n+1}|\mathcal{F}_n] \quad \text{a.e.} \quad (\textit{submartingale})$$
$$X_n \geq \mathbb{E}[X_{n+1}|\mathcal{F}_n] \quad \text{a.e.} \quad (\textit{supermartingale})$$

Examples

Example

Check that $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a (sub)martingale in each case below:

1. let $(X_n)_n$ be a sequence of independent r.v. with mean zero, $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ and $Y_n = S_n$,
2. let $(X_n)_n$ be a sequence of independent r.v. with mean one, $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ and $Y_n = \prod_{k=1}^n X_k$,
3. let X be an integrable r.v. and let $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$, $Y_n = \mathbb{E}[X|\mathcal{F}_n]$, (GOOD FOR CREATING MARTINGALES!)
4. let $(X_n)_n$ be a sequence of non-negative integrable r.v., $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ and $Y_n = S_n$, (sub)

Note that the condition for martingale implies that for $n < m$ we have that

$$X_n = \mathbb{E}[X_m|\mathcal{F}_n] \quad \text{a.e.}$$

Jensen's inequality

Theorem (Jensen's inequality)

Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a submartingal and let φ be an increasing convex function defined on $\mathbb{R}$. If $\varphi(X_n)$ is integrable for any n, then $(\varphi(X_n), \mathcal{F}_n)_{n\in\mathbb{N}}$ is also a submartingal.

Corollary

If $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingal then $(X_n^+, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingal. If $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingal, then $(|X_n|, \mathcal{F}_n)_{n\in\mathbb{N}}$ and $(|X_n|^p, \mathcal{F}_n)_{n\in\mathbb{N}}$ for $1 < p < \infty$ (if $X_n \in \mathbb{L}^p$) are also submartingales.

Exercise: Prove the theorem.

Martingales in Game theory

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d.r.v. taking the value 1 with probability p and -1 with probability $1-p$. The interpretation is that $X_n=1$ represents a success while $X_n=-1$ represents a failure of a player at the n-th time he is playing a game. Let us suppose that the player can win or lose a certain amount V_n at the n-th time he plays the game, that is, V_n is the amount of the bet at time n. Then, at time n the player possesses

$$Y_n=\sum_{i=1}^{n} V_iX_i=Y_{n-1}+V_nX_n.$$

It is quite natural to assume that the amount V_n may depend on the previous amounts, that is, of $V_1,\cdots,V_{n-1}$ and also of $X_1,\cdots,X_{n-1}$. In other words, let $\mathcal{F}_0=\{\varnothing,\Omega\}$ and $\mathcal{F}_n=\sigma(X_1,\cdots,X_n)$. Then, V_n is a function $\mathcal{F}_{n-1}$ measurable, that is, the sequence that determines the player's strategy is said to be predictable.

Martingales in Game theory (cont.)

Let $S_n = X_1 + \cdots + X_n$. Then

$$Y_n = \sum_{i=1}^{n} V_i \Delta S_i,$$

where $\Delta S_i = S_i - S_{i-1}$. Then, the sequence $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is said to be the transform of S by V.

From the player's point of view, the game is said to be fair (favorable or unfavorable) if at each step if $\mathbb{E}[Y_{n+1} - Y_n|\mathcal{F}_n] = 0 \quad (\geq 0 \quad \text{or} \quad \leq 0)$

We want to analyze in which conditions the game is fair? A simple computation shows that :

1. The game is fair if $p = 1 - p = 1/2$. $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale.
2. The game is favorable if $p > 1 - p$. $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingale.
3. The game is fair if $p < 1 - p$. $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a supermartingale.

Martingales in Game theory (cont.)

Let us now consider another strategy. Take $(V_n, \mathcal{F}_{n-1})_{n\geq 1}$ with $V_1 = 1$ and for $n \geq 1$ we have that $V_n = 2^{n-1}$ if $X_1 = -1, \cdots, X_{n-1} = -1$ and 0 otherwise.

Under this strategy, a player starts to bet 1 euro and doubles the bet in the next play if he had lost or leaves immediately the game in case he had won.

If $X_1 = -1, \cdots X_n = -1$, then the total loss after n plays is $\sum_{i=1}^n 2^{i-1} = 2^n - 1$.

Therefore, if $X_{n+1} = 1$ then $Y_{n+1} = Y_n + X_{n+1}V_{n+1} = -(2^n - 1) + 2^n = 1$.

Let $\tau := \inf\{n \geq 1 : Y_n = 1\}$, that is the first time that $Y_n = 1$. If $p = \frac{1}{2}$, then the game is fair and

$$\mathbb{P}(\tau = n) = \mathbb{P}(Y_n = 1, Y_k \neq 1, \forall k = 1, \cdots, n-1) = \Big(\frac{1}{2}\Big)^n.$$

Martingales in Game theory (cont.)

From where we conclude that

$$\mathbb{P}(\tau < \infty) = \mathbb{P}(\cup_{n\geq 1} \tau = n) = \sum_{n\geq 1} \Big(\frac{1}{2}\Big)^n = 1.$$

Moreover, $\mathbb{P}(Y_\tau = 1) = 1$ and $\mathbb{E}[Y_\tau] = 1$.
Therefore, even in a fair game, applying the strategy described above, a player can, in finite time, complete the game with success, that is, increase his capital in one unity: $\mathbb{E}[Y_\tau] = 1 > Y_0 = 0$. In game theory this type of system - double the bet after a loss and leave the game immediately after a win - is called a martingale.
We note however that $p = 1/2$, so that $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale and $\mathbb{E}[Y_n] = \mathbb{E}[Y_0] = 0$ for all $n \geq 1$. Above the same is not true for a random time (above we took the random time τ.)

Markov time

Definition

A r.v. τ which takes values in the set $\{0, 1, \cdots, \infty\}$ is said to be a Markov time wrt a σ-algebra $\mathcal{F}_n$ if for each $n \geq 0$ we have that $\{\tau = n\} \in \mathcal{F}_n$. When $\mathbb{P}(\tau < \infty) = 1$, the Markov time is said to be a stopping time.

If $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a sequence of r.v. and σ-algebras with $\mathcal{F}_n \in \mathcal{F}_{n+1}$, and if τ is a Markov time wrt $\mathcal{F}_n$, then we write $X_\tau = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{\tau=n\}}$. Note that since $\mathbb{P}(\tau < \infty) = 1$ we have that $X_\tau = 0$ in the set $\tau = \infty$. Prove that X_τ is a r.v.

Example (Prove it!)

Let $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ be a martingale (or submartingale) and τ a Markov time wrt $\mathcal{F}_n$. Then the stopping process $X^\tau = (X_{n\wedge\tau}, \mathcal{F}_n)$ is also a martingale (or submartingale).