

PROBA

David Wiedemann

Table des matières

1	Some historical models	5
1.1	Laplace Model	5
2	Basic Formalism	6
2.1	Measure spaces : A notion of area	6
2.2	Probability spaces	7
2.3	Basic properties	7
2.4	Measurable and measure preserving maps	8
3	Probability spaces	8
3.1	Discrete probability spaces	8
3.1.1	Symmetric simple random walk	9
3.2	Continuous probability spaces	9
3.3	Borel σ -algebra	9
3.4	Probability Measures on \mathbb{R}^n	10
3.5	Probability measures on $(\mathbb{R}, \mathcal{F}_E)$	10
3.6	Probability measures on \mathbb{R}^n	12
3.7	Product probability measures on $\mathbb{R}^n, \mathbb{R}^{\mathbb{N}}$	13
3.7.1	Product σ -algebra	13
3.8	Product measures	13
3.9	Infinite product spaces	13
4	Random Variables	14
4.1	Independence of RV	17
4.2	Example of continuous random variables	17
4.3	Transformation of random variables	18
5	Random Vectors	18
5.1	Conditional law	20

6	Mathematical Expectation	21
6.1	Expected value for general random variables	22
6.2	Variance and covariance	25
6.3	Moments of a random variable	26
6.4	Moment Generating functions	27
6.5	MGF vs independence	28
7	Limit Theorems	28
7.1	Sequences of events	28
7.2	Sequences of random variables	30
7.3	Convergence in law vs almost sure convergence	31
8	Law of large numbers	32
8.1	Central limit theorem	33

List of Theorems

1	Definition (Laplace Model)	5
1	Proposition	5
2	Proposition	5
2	Definition (Intermediate model)	5
3	Definition (Geometric probability)	6
4	Definition (Measure space)	6
5	Definition (Probability space)	7
3	Lemme	7
6	Definition	8
4	Lemme (Push-Forward measure)	8
5	Proposition	8
6	Lemme	8
7	Proposition	9
7	Definition (Borel sigma-algebra)	9
8	Theorème (Existence of Lebesgue-measure)	10
9	Theorème (Uniforme Measure)	10
8	Definition (Cumulative distribution function)	10
10	Theorème	10
11	Lemme	11
12	Corollaire	12
13	Theorème (Dynkin)	12
9	Definition (Joint c.d.f.)	12
14	Theorème	13
10	Definition (Product algebra)	13
11	Definition	13

16	Lemme	14
17	Proposition	14
12	Definition (Random Variables)	14
13	Definition (Equality of RV)	14
14	Definition	14
19	Proposition	15
20	Lemme	15
15	Definition	15
21	Lemme	15
16	Definition	15
22	Proposition	15
17	Definition	16
23	Proposition	16
24	Proposition	16
18	Definition (Independence of RV)	17
25	Proposition	17
19	Definition (Random variables with density)	17
26	Théorème (Version of central limit theorem)	18
27	Lemme	18
28	Proposition	18
20	Definition (Random Vector)	18
30	Lemme	18
31	Proposition	19
32	Lemme	19
33	Proposition	19
34	Corollaire	19
21	Definition (Random vectors with density)	19
35	Proposition	20
36	Corollaire	20
22	Definition	20
23	Definition (Conditional law of rv with density)	20
37	Lemme	20
24	Definition (Expectation for discrete r.v.)	21
39	Proposition	21
40	Corollaire	21
41	Proposition	22
42	Proposition	22
43	Proposition	23
44	Proposition	23
45	Proposition	24
46	Proposition	24

25	Definition (Variance)	25
48	Proposition (Chebyshev inequality)	26
26	Definition (Covariance)	26
27	Definition (Moment)	26
49	Proposition	26
50	Theorème (Stone- Weierstrass)	27
28	Definition	27
53	Theorème (MGF determines the law)	27
29	Definition (MGF for random vectors)	27
54	Theorème	28
55	Proposition	28
56	Proposition	28
57	Lemme	28
58	Theorème (Borel-Cantelli I)	29
60	Theorème (Borel-Cantelli II)	29
30	Definition (Almost sure convergence)	30
31	Definition (Convergence in law)	30
61	Lemme	30
63	Proposition	31
64	Lemme	31
32	Definition (Convergence in probability)	32
65	Theorème (Weak LLN)	32
66	Theorème (Strong LLN)	33
69	Theorème (Central limit theorem)	34
70	Proposition (Lindeberg Exchange principle)	34

1 Some historical models

1.1 Laplace Model

Definition 1 (Laplace Model)

Ω finite set, $|\Omega| = n$ is the set of outcomes.

We can observe whether $E \subset \Omega$ happens, and we define it's probability

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

Question

Why should this have any meaning/content ?

Proposition 1

Consider laplace model for n coin tosses \Rightarrow every sequence has probability 2^{-n}

Denote by H_n the number of heads in n tosses

$$\mathbb{P}\left(\left|\frac{H_n}{n} - \frac{1}{2}\right| > \epsilon\right) \rightarrow 0$$

More generally

Proposition 2

If you have a laplace model for some event E , and look at n repetitions, then

$$\forall \epsilon > 0 \mathbb{P}\left(\left|\frac{E_n}{n} - \mathbb{P}(E)\right| > \epsilon\right) \rightarrow 0$$

Limitations of Laplace Model

- All outcomes have equal probability ?
- Need $|\Omega| < \infty$, so what about infinite sets ?

What next ?

Definition 2 (Intermediate model)

Let Ω to be any set and $P : \Omega \rightarrow [0, 1]$, s.t. $\sum_{\omega \in \Omega} p(\omega) = 1$

Event : $E \subset \Omega$ and

$$\mathbb{P}(E) := \sum_{\omega \in E} p(\omega)$$

- More freedom
- If you take Ω finite, $p(\omega) = \frac{1}{|\Omega|} \Rightarrow$ Laplace model
- Price ? How to choose $p : \Omega \rightarrow [0, 1] \rightarrow$ collect data, do statistics
- keeps many nice properties

- For countable sets, this is equivalent to the standard model.
- For uncountable Ω ?
- Problem 1 : There is no function s.t.

$$p(\omega) > 0 \forall \omega \in \Omega \text{ and } \sum p(\omega) = 1$$

This intermediate model is in essence only for countable sets.

What about uncountable sets ?

- What about a random point in $[0, 1]$ or $[0, 1]^n$?

Intuitively, consider $[0, 1]$, then we can set

$$\mathbb{P}(A) = \text{length}(A)$$

Definition 3 (Geometric probability)

Take $f : \mathbb{R} \rightarrow (0, \infty)$ to be a riemann-integrable function with total mass 1.

For any $A \subset \mathbb{R}$, s.t. 1_A riemann-integrable, we set $\mathbb{P}(A) = \int_A f(x)dx$

- In general quite ok
BUT
- You would expect there is one framework for uncountable and countable sets.
- What about more complicated spaces (eg. space of continuous functions)
- $\mathbb{P}(\mathbb{Q})$ is undefined

2 Basic Formalism

2.1 Measure spaces : A notion of area

- Set + structure
- General setting to talk about area

Definition 4 (Measure space)

$(\Omega, \mathcal{F}, \mu)$ is called a measure space if :

- Ω is some set
- $\mathcal{F} \subset P(\Omega)$ called a σ -algebra
 - $\emptyset \in \mathcal{F}$
 - $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$
 - $F_1, F_2, \dots \in \mathcal{F}$, then $\bigcup_{i \geq 1} F_i \in \mathcal{F}$ each F is called a measurable set.
- $\mu : \mathcal{F} \rightarrow [0, \infty)$ called the measure
 - $\mu(\emptyset) = 0$

— If F_1, \dots , are disjoint sets of the σ -algebra, then

$$\mu\left(\bigcup_{i \geq 1} F_i\right) = \sum_{i \geq 1} \mu(F_i)$$

— Defined by Borel 1898 and Lebesgue 1901-1903

2.2 Probability spaces

Given by Kolmogorov in 1933

Definition 5 (Probability space)

A triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space if it is a measure space and $\mathbb{P}(\Omega) = 1$

Interpretation

- Ω state space/universe
- \mathcal{F} is the set of events you can observe/have access to
- $\mathbb{P}(E)$ is the probability of E

Lemme 3

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

- $\Omega \in \mathcal{F}$
- $F_1, F_2 \in \mathcal{F} \Rightarrow F_1 \setminus F_2 \in \mathcal{F}$
- $F_1, \dots \in \mathcal{F} \Rightarrow \bigcap F_i \in \mathcal{F}$
- $F_1, F_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{i \geq 1} F_i \in \mathcal{F}$

Let us compare this definition with the prior ones

- Ω finite set, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(F) = \frac{|F|}{|\Omega|}$ this is a probability space and a laplace model.
- For Ω countable, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega)$
- The really new part is \mathcal{F} which restricts the sets we can measure

Lecture 2: ...

Wed 29 Sep

2.3 Basic properties

- $F_1, F_2, \dots \in \mathcal{F}$ disjoint

$$\mu\left(\bigcup F_i\right) = \sum \mu(F_i)$$

- $F_1 \subset F_2 \in \mathcal{F} \Rightarrow \mu(F_1) \leq \mu(F_2)$
- $F_1 \subset F_2 \subset \dots \in \mathcal{F}$

$$\mu(F_n) \rightarrow \mu\left(\bigcup F_i\right)$$

— $F_1, F_2, \dots, \mathcal{F}$

$$\mu(\bigcup F_i) \leq \sum \mu(F_i)$$

In addition, in probability spaces

— $\mathcal{P}(F^c) = 1 - \mathcal{P}(F)$

— $F_1 \supset F_2 \supset \dots \Rightarrow \mathcal{P}(F_n) \rightarrow \mathcal{P}(\bigcap F_i)$

2.4 Measurable and measure preserving maps

Definition 6

Let $(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ two measure spaces.

$f : \Omega_1 \rightarrow \Omega_2$ is called measurable if for every $F \in \mathcal{F}_2$, $f^{-1}(F) \in \mathcal{F}_1$

A measurable function $f : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ is called measure preserving if $\forall F \in \mathcal{F}_2 \mu_1(f^{-1}(F)) = \mu_2(F)$.

Lemme 4 (Push-Forward measure)

Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), (\Omega_2, \mathcal{F}_2)$ be two measure spaces, and f measurable, then $\mathbb{P}_2(F) = \mathbb{P}_1(f^{-1}(F))$ is a probability measure.

3 Probability spaces

— Discrete probability spaces : Ω countable

— Continuous probability spaces : Ω uncountable.

3.1 Discrete probability spaces

Does introducing a σ -algebra \mathcal{F} enlargen the generality?

Proposition 5

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a discrete probability space, $\exists \Omega_2$ countable, $\mathbb{P}_2 : \mathcal{P}(\Omega_2) \rightarrow [0, 1]$ s.t. $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$ is a probability space and $\exists f : (\Omega_1, \mathcal{F}_1) \rightarrow (\Omega_2, \mathcal{F}_2)$ is measure preserving

Still \mathcal{F} is useful :

— can sequentially study a model/situation by taking $\mathcal{F}_1 \subset \mathcal{F}_2 \dots$

Lemme 6

There is no shift-invariant probability measure on $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$

Preuve

$$\mathbb{P}(\mathbb{Z}) = \mathbb{P}(\bigcup_n \{n\}) = \sum \mathbb{P}(\{n\}) = \infty$$

\Rightarrow cannot treat everyone on an equal ground!

□

3.1.1 Symmetric simple random walk

A simple walk of length n s.t. $|s_n - s_{n-1}| = 1$.

Let Ω be the set of all walks of length n , and consider $(\Omega, P(\Omega), \mathbb{P})$.

What is the probability that S hits 0?

What does it look like, what is its max?

3.2 Continuous probability spaces

Can we define a probability measure on S^1 s.t. $(S^1, P(S^1))$ that is rotation invariant?

Similarly to the countable case, but not the same as Ω is uncountable and setting $P(\{\omega\}) = 0$ gives no contradiction.

Proposition 7

You can not.

Preuve

Idea : decompose S^1 into countable many sets A_n st $\bigcup A_n = S^1$, they are disjoint and rotations of each other.

$\forall x \in S^1$, define S_x as $\{\dots, T^{-2}x, T^{-1}x, x, Tx, \dots\}$.

Note that either $S_x = S_y$ or $S_x \cap S_y = \emptyset$.

Lecture 3: Measurable maps

Wed 06 Oct

3.3 Borel σ -algebra

- Cannot define shift-invariant probability measure on $([0, 1], \mathcal{P}([0, 1]))$.
- What σ -algebra to choose on (X, τ) ?
- Want to know the size of all open-sets

Definition 7 (Borel sigma-algebra)

On (X, τ) the borel σ -algebra \mathcal{F}_τ is the smallest σ -algebra containing τ .

This is well defined because, given a collection of σ -algebras, their intersection is too.

Two nice properties

- Continuous functions on a Borel σ -algebra are also measurable.

Preuve

Suffices to check that $f^{-1}(U) \in \mathcal{F}_{\tau_1}$ for $U \in \tau_2$ but this is immediate since f is continuous.

In (\mathbb{R}^n, τ_E) , the Borel σ -algebra \mathcal{F}_E is generated by $(a_1, b_1) \times \dots \times (a_n, b_n)$.
 \mathcal{F}_E is the smallest σ -algebra containing open intervals. \square

3.4 Probability Measures on \mathbb{R}^n

Theorème 8 (Existence of Lebesgue-measure)

There exists a unique measure λ on $(\mathbb{R}^n, \mathcal{F}_E)$ s.t. $\lambda((a_1 \times b_1) \times \dots \times (a_n, b_n)) = \prod_i |b_i - a_i|$

Theorème 9 (Uniform Measure)

There exists a unique \mathbb{P} measure on $([0, 1]^n, \mathcal{F}_E)$ with the same property.

Both λ and \mathbb{P} are shift-invariant in fact only shift invariant measures on \mathbb{R} (up to a constant)

Preuve

Consider the case of $(\mathbb{R}^n, \mathcal{F}_E)$ and $f_r : x \rightarrow x + \tau, \tau \in \mathbb{R}^n$.

- f_r continuous \Rightarrow measurable
- $\tilde{\mathbb{P}}(A) = \mathbb{P}(f^{-1}(A))$ is a probability measure
- All boxes have the same measure \square

3.5 Probability measures on $(\mathbb{R}, \mathcal{F}_E)$

We saw that we can put a uniform measure on $[0, 1]$.

All probability measures on $(\mathbb{R}, \mathcal{F}_E)$

1. $\mathbb{P} : \mathcal{F}_E \rightarrow [0, 1]$
2. These are actually only characterized by $\mathbb{P}((-\infty, x))$

Definition 8 (Cumulative distribution function)

$F : \mathbb{R} \rightarrow [0, 1]$ is called a c.d.f if

- F is non-decreasing
- $F(x_n) \rightarrow 0$ then $x_n \rightarrow -\infty$
- $F(x_n) \rightarrow 1$ if $x_n \rightarrow 1$
- F is right-continuous.

Theorème 10

Given a probability measure \mathbb{P} on $(\mathbb{R}, \mathcal{F}_E)$, then $f(x) := \mathbb{P}((-\infty, x))$ is a c.d.f

Given a c.d.f, there exists a unique probability measure s.t. $\mathbb{P}((-\infty, x)) = F(x)$

Preuve

Given \mathbb{P} on $(\mathbb{R}, \mathcal{F}_E)$.

Let's show that $F(x) = \mathbb{P}((-\infty, x))$ is a c.d.f.

- $x < y \quad F(x) = \mathbb{P}((-\infty, x)) \leq \mathbb{P}((-\infty, y)) = F(y)$
- $x_n \rightarrow -\infty \quad F(x_n) = \mathbb{P}((-\infty, x_n)) \rightarrow \mathbb{P}(\bigcap_n (-\infty, x_n)) = 0$
- $x_n \rightarrow \infty \Rightarrow F(x_n) \rightarrow 1$ is similar
- Also for right continuous $x_n \rightarrow x$, we have that $[x_n, \infty) \subset [x_{n+1}, \infty)$

How do we construct \mathbb{P} given F ?

Trick using push-forward measure.

Define $f : (0, 1) \rightarrow \mathbb{R}$, define

$$f(x) = \inf_{y \in \mathbb{R}} \{F(y) \geq x\}$$

□

Define $\mathbb{P}(A) := \mathbb{P}_U(f^{-1}(A)) \forall A \in \mathcal{F}_E$ Why is f measurable?

If f is increasing $\Rightarrow f$ is measurable

Lecture 4: ...

Wed 13 Oct

Each c.d.f gives rise to a unique \mathbb{P} .

A priori $\mathbb{P}_1 = \mathbb{P}_2$ means $\forall F \in \mathcal{F}_E \mathbb{P}_1(F) = \mathbb{P}_2(F)$.

We show that it suffices to show that $\mathbb{P}_1((-\infty, x]) = \mathbb{P}_2((-\infty, x]) \forall x \in \mathbb{R}$.

Lemme 11

Given $(\mathbb{R}, \mathcal{F}_E, \mathbb{P})$ then $\forall B \in \mathcal{F}_E, \forall \epsilon > 0$ one can find disjoint intervals I_1, \dots, I_n s.t. $\mathbb{P}(B \Delta (I_1 \cup \dots \cup I_n)) < \epsilon$

Preuve

Consider the collection H of all subsets $H \in \mathcal{F}_E$ s.t. the property above holds.

We know that H contains all intervals, hence $\sigma(H) = \mathcal{F}_E$.

So we only need to show that H is a σ -algebra

1. $\emptyset \in H$: Know that $\forall x (-\infty, x] \in H$

2. If $B \in H \Rightarrow B^C \in H$.

Given $\epsilon > 0$, choose I_1, \dots, I_n s.t. $\mathbb{P}(B \Delta (I_1 \cup \dots)) < \epsilon$, but $(B \Delta A) = B^C \Delta A^C$, hence

$$\mathbb{P}(B^C \Delta (I_1 \cup \dots)) < \epsilon$$

3. $H_1, \dots \in H$, we want $\bigcup_i H_i \in H \exists n \in \mathbb{N}$

$$\mathbb{P}((\bigcup_{i=0}^m H_i) \Delta (\bigcup_i H_i)) < \frac{\epsilon}{2}$$

$\forall i = 1, \dots, m$, we have disjoint $I_{i,1}, \dots, I_{i,m_i}$ s.t.

$$\mathbb{P}(H_i \Delta (I_{i,1} \cup \dots)) < \frac{\epsilon}{2m}$$

Now use that

$$(\bigcup_{i=1}^m H_i) \Delta (\bigcup_{i=1}^m \bigcup_{j=1}^{m_i} I_{i,j}) \subseteq \bigcup_{i=1}^m (H_i \Delta \bigcup_{j=1}^{m_i} I_{i,j})$$

Finally, we can write a finite union of disjoint intervals

□

Corollaire 12

$\mathbb{P}_1, \mathbb{P}_2$ probability measure on $(\mathbb{R}, \mathcal{F}_E)$, then $\mathbb{P}_1 = \mathbb{P}_2$ as soon as

$$\mathbb{P}_1((-\infty, x]) = \mathbb{P}_2((-\infty, x])$$

or

$$\mathbb{P}_1(x, y) = \mathbb{P}_2(x, y)$$

Preuve

Notice $(-\infty, x)$ can be written as

$$(-\infty, x) = \left(\bigcup_n (x, x+n) \right)^C$$

So it suffices to prove the first point.

Observe, for all intervals $\mathbb{P}_1(I) = \mathbb{P}_2(I)$ since

$$\mathbb{P}_i(y, x) = \mathbb{P}_i(-\infty, x) - \mathbb{P}_i(-\infty, y)$$

The condition holds for B if $\forall \epsilon > 0$, we can pick I_1, \dots, I_n s.t.

$$\mathbb{P}_1(B \Delta (I_1 \cup \dots)) < \epsilon$$

and

$$\mathbb{P}_2(B \Delta (I_1 \cup \dots)) < \epsilon$$

So we need to check again that this is a σ -algebra and we are done.

Now we can conclude that

$$|\mathbb{P}_1(B) - \mathbb{P}_1(I_1 \cup \dots)| = |\mathbb{P}_1(B) - \mathbb{P}_2(I_1 \cup \dots)| < \epsilon$$

and

$$|\mathbb{P}_2(B) - \mathbb{P}_1(I_1 \cup \dots)| = |\mathbb{P}_2(B) - \mathbb{P}_2(I_1 \cup \dots)| < \epsilon$$

□

An abstract uniqueness result follows from a similar strategy.

Theorème 13 (Dynkin)

\mathbb{P}_1 and \mathbb{P}_2 two probability measures on (Ω, \mathcal{F}) , suppose $\mathbb{P}_1(H) = \mathbb{P}_2(H)$ for all $H \in \mathcal{H} \subset \mathcal{F}$ and

- $\sigma(\mathcal{H}) = \mathcal{F}$
- $H_1 \in \mathcal{H}, H_2 \in \mathcal{H} \Rightarrow H_1 \cap H_2 \in \mathcal{H}$

Then $\mathbb{P}_1 = \mathbb{P}_2$

3.6 Probability measures on \mathbb{R}^n **Definition 9 (Joint c.d.f.)**

$F : \mathbb{R}^n \rightarrow [0, 1]$

- F non-decreasing in each coordinate
- $F(x_1, \dots, x_n) \rightarrow 1$ if all $x_i \rightarrow -\infty$
- right-continuous

Theorème 14

Joint c.d.f $\iff \mathbb{P}$ on $(\mathbb{R}^n, \mathcal{F}_E)$

3.7 Product probability measures on $\mathbb{R}^n, \mathbb{R}^{\mathbb{N}}$

- Related to independence
- Natural mathematically

2 steps

- product σ -algebra
- product measure

3.7.1 Product σ -algebra

Definition 10 (Product algebra)

Let $(\Omega_i, \mathcal{F}_i)_{i \geq 1}$ measurable spaces, then the product σ -algebra \mathcal{F}_π on $\prod_i \Omega_i$ is the σ -algebra generated by sets $F = E_1 \times \dots \times E_n \times \Omega_{n+1} \times \dots$, $E_i \in \mathcal{F}_i$

Remarque

- Projections are measurable
- In fact, product σ -algebra s.t. all projections are measurable

Notice on \mathbb{R}^n , we now have two ways to define a σ -algebra.

- Take (\mathbb{R}^n, τ_E) and induce a Borel σ -algebra
- Take n copies of $(\mathbb{R}, \mathcal{F}_E)$ and consider \mathcal{F}_π on \mathbb{R}^n

3.8 Product measures

Definition 11

Given $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)_{i \geq 1}$ probability spaces \mathbb{P}_π on $(\prod_i \Omega_i, \mathcal{F}_\pi)$ is called the product measure of \mathbb{P}_i .

If $\forall n \geq 1$, all sets $E = E_1 \times E_2 \times \dots \times E_n \times \Omega_{n+1} \times \dots$

$$\mathbb{P}_\pi(E) = \prod_{i=1}^n \mathbb{P}_i(E_i)$$

Lecture 5: Conditional probability

Wed 20 Oct

3.9 Infinite product spaces

Case of $(\mathbb{R}, \mathcal{F}_E, \mathbb{P}_i)_{i \geq 1}$.

Space of infinite fair coin tosses

We want the infinite product of $(\{0, 1\}, P(\{0, 1\}), \mathbb{P})$.

We use the uniform measure $([0, 1], \mathcal{F}_E, \mathbb{P})$, for $x \in [0, 1]$, $x = 0.x_1x_2\dots$, we send $f : x \rightarrow (x_1, x_2, \dots)$

Lemme 16

f as defined above is measurable

Preuve

Note that

— \mathcal{F}_π generated by $F_1 \times \dots, F_n \times \{0, 1\} \times \{0, 1\}$ with $|F_i| = 1$

— \mathcal{F}_E is generated by sets of the forme $(2^{-n}j, 2^{-n}(j+1))$. □

Moreover, $(j2^{-n}, (j+1)2^{-n})$ is in correspondence with $F_1 \times \dots \times F_n \times \{0, 1\} \times \dots$

Proposition 17

There exists a product probability measure on $(\{0, 1\}^{\mathbb{N}}, \mathcal{F}_\pi)$

Preuve

Consider $f : ([0, 1], \mathcal{F}_E) \mapsto (\{0, 1\}^{\mathbb{N}}, \mathcal{F}_\pi)$.

We define \mathbb{P}_π as the pushforward of \mathbb{P}_U under f □

Lecture 6: Random Variables

Wed 27 Oct

4 Random Variables

Definition 12 (Random Variables)

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

Then $X : \Omega \mapsto \mathbb{R}$ measurable as a map $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{F}_E)$ is called a (real) random variable.

The pushforward measure $\mathbb{P}_X(F) = \mathbb{P}(X^{-1}(F)) \forall F \in \mathcal{F}_E$ is called the law of X

Remarque

There is a more general notion of $(\Omega_2, \mathcal{F}_2)$ valued random variable.

Definition 13 (Equality of RV)

X, Y two random variables are called equal in law if

$$\mathbb{P}_X(F) = \mathbb{P}_Y(F) \forall F \in \mathcal{F}_E$$

Definition 14

X is a R.V. we call the c.d.f. of \mathbb{P}_X F_X

$$F_X(s) = \mathbb{P}_X(X \leq s)$$

Proposition 19

Each R.V. X gives rise to a unique c.d.f. $F_X(s) = \mathbb{P}_X(X \leq s)$ and conversely, each c.d.f. gives rise to a unique law of a probability measure

Preuve

Follows directly from the proposition relating probability measures and c.d.f. \square

Lemme 20

1. $\mathbb{P}_X < s = F(s^-)$
2. $\mathbb{P}_X(X = s) = F(s) - F(s^-)$
3. $\mathbb{P}_X(X \in (a, b)) = F(b^-) - F(a)$

Definition 15

X a R.V., $s \in \mathbb{R}$.

If $F(s) - F(s^-) > 0 \iff \mathbb{P}_X(X = s) > 0$, then s is a atom of X

Lemme 21

A R.V. can have at most countably many atoms or in other words, a c.d.f. can have at most countably many jumps.

Definition 16

X a R.V.

If F_X increases by jumps, we call X a discrete R.V.

If F_X is cts, we call X a cts R.V.

Proposition 22

X a R.V. Then we can write $F(X) = aF_Y + bF_Z$ s.t. $a + b = 1$ and Y discrete, Z cts R.V.

Preuve

If F_X is discrete or cts, we are done.

$\exists S = \{s_1, s_2, \dots\}$ s.t. $F_X(s_i) - F_X(s_i^-) > 0$ iff $s_i \in S$.

Consider

$$\hat{F}_Y(s) = \sum 1_{\{S \geq s_i\}} (F(s_i) - F(s_i^-))$$

and

$$\hat{F}_Z(s) = F_X(s) - \hat{F}_Y(s)$$

We now show that \hat{F}_Z continuous.

Finally, define

$$F_Y(s) = \frac{\hat{F}_Y(s)}{\hat{F}_Y(\infty)}$$

and similarly

$$F_Z(s) = \frac{\hat{F}_Z(s)}{\hat{F}_Z(\infty)}$$

\square

Lecture 7: Example of RV

Wed 03 Nov

Geometric R.V.

Let $S = \mathbb{N}$ and $0 < p \leq 1$.

$$\mathbb{P}(X = k) = (1 - p)^{k-1}p$$

Corresponds to first succes if success rate is p .

Definition 17

We call a rv with support \mathbb{N} memoryless if

$$\mathbb{P}(X > k + l | X > k) = \mathbb{P}(X > l)$$

Proposition 23

Geo(p) is memoryless and every memoryless RV with support on \mathbb{N} is a geometric rv.

Preuve

$$\mathbb{P}(X > k + l | X > k) \mathbb{P}(X > k) = \mathbb{P}(X > k + l) = (1 - p)^{k+l}$$

But also $\mathbb{P}(X > l) = (1 - p)^l$

$$\mathbb{P}(X > k + l | X > k) = (1 - p)^l$$

Now suppose X is a memoryless RV with $\mathbb{P}(X > 1) > 0$, then

$$\mathbb{P}(X > l + 1 | X > 1) = \frac{\mathbb{P}(X > l + 1)}{\mathbb{P}(X > 1)} = \mathbb{P}(X > l) \quad \square$$

Inductively, it follows that $\mathbb{P}(X > l) = \mathbb{P}(X > 1)^l$

Poisson RV

Define

$$\mathbb{P}(Poi(\lambda) = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Proposition 24

$$Ber(n, \frac{\lambda}{n}) \mapsto Poi(\lambda) \text{ as } n \rightarrow \infty$$

in the sense that $\forall k \in \mathbb{N}$

$$\mathbb{P}(\text{Ber}(n, \frac{\lambda}{n}) = k) \rightarrow \mathbb{P}(\text{Poi}(\lambda) = k)$$

Preuve

$$\begin{aligned} \mathbb{P}(\text{Bin}(n, \frac{\lambda}{n}) = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \mathbb{P}(\text{Bin}(n, \frac{\lambda}{n}) = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \left(\frac{n!}{(n-k)! n^k}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

4.1 Independence of RV

Definition 18 (Independence of RV)

$(X_i)_{i \geq 1}$ RV defined on $(\Omega, \mathcal{F}, \mathbb{P})$ are called mutually independent if $\forall J \subset \{1, 2, \dots\}$ finite ad $\forall E_j \in \mathcal{F}_E \forall j \in J$.

$$\mathbb{P}\left(\bigcap_{j \in J} \{X_j \in E_j\}\right) = \prod \mathbb{P}(X_j \in E_j)$$

Proposition 25

$(X_i)_{i \geq 1}$ RV with laws \mathbb{P}_{X_i} then we can find a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and RV \tilde{X}_i s.t.

- $X_i \simeq \tilde{X}_i$
- (\tilde{X}_i) are mutually independent.

Preuve

Consider the product probability space of (\mathbb{P}_{X_i}) i.e. $(\mathbb{R}^n, \mathcal{F}_\pi, \mathbb{P}_\pi)$.

Let \tilde{X}_i be the projection on the i -th coordinate.

Are (\tilde{X}_i) independent?

$$\mathbb{P}_\pi\left(\bigcap_{j \in J} \{\tilde{X}_j \in E_j\}\right) = \mathbb{P}_\pi\left(\bigcap_i F_i\right)$$

□

With $F_i = \mathbb{R}$ if $i \notin J$ and $F_i = E_i$ if $i \in J$

4.2 Example of continuous random variables

We will mainly work with a subclass of continuous rv :

Definition 19 (Random variables with density)

We call a continuous rv X with c.d.f. F_x a r.v. with densite if $\exists f : \mathbb{R} \rightarrow [0, \infty)$ which is integrable, $\int_{\mathbb{R}} f = 1$ st

$$\mathbb{P}(X \leq t) = F_X(t) = \int_{-\infty}^t f_X(s) ds$$

Lecture 8: rv with density

Wed 10 Nov

The gaussian random variable describes sums of independent errors

Theorème 26 (Version of central limit theorem)

Let X_1, \dots be iid random variables s.t. $\mathbb{P}(|X_I| < C) = 1$ for some $C > 0$ and such that $-X_i$ and X_i have the same law.

Then $S_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}} \rightarrow \mathcal{N}(0, \sigma^2)$ in the sense that $\mathbb{P}(S_n \in (a, b)) \rightarrow \mathbb{P}(\mathcal{N}(0, \sigma^2) \in (a, b))$

4.3 Transformation of random variables

Lemme 27

If $\phi : \mathbb{R} \rightarrow \mathbb{R}$ continuous and r.v. on $(\omega, \mathcal{F}, \mathbb{P})$ then $\phi(X)$ is also r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$

Preuve

Need to check that $\phi \circ X$ is measurable.

- Continuous functions are measurable
- Composition of measurable functions is measurable □

Proposition 28

Let X be a continuous random variable with density $f_X : \mathbb{R} \rightarrow [0, \infty)$.

Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ bijective, continuously differentiable with $\phi'(x) \neq 0 \forall x \in \mathbb{R}$.

Then $\phi(X)$ is a r.v. with density given by

$$f_{\phi(X)}(x) = \frac{1}{\phi' \circ \phi^{-1}(X)} f_X(\phi^{-1}(x))$$

5 Random Vectors

Definition 20 (Random Vector)

$(\Omega, \mathcal{F}, \mathbb{P})$ X_1, \dots, X_n random variables then $\overline{X} = (X_1, \dots, X_n)$ is called a random vector.

Remarque

Marginal laws on their own do not describe the behavior of \overline{X} .

Lemme 30

If $X_1, \dots, X_n : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{F}_E)$ measurable \overline{X} is measurable from $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^n, \mathcal{F}_E)$

Preuve

Suffices to check that \bar{X}^{-1} of each $E = F_1 \times \dots \times F_n$ is in \mathcal{F}

$$\bar{X}^{-1}(E) = \bigcap X_i^{-1}(F_i) \in \mathcal{F}$$

Hence we can define

$$\mathbb{P}_{\bar{X}}(E) := \mathbb{P}(\bar{X}^{-1}(E)) \forall E \in \mathcal{F}_E$$

Which is a probability law on $(\mathbb{R}^n, \mathcal{F}_E)$ called the joint law of \bar{X} □

Proposition 31

The joint law of a random vector \bar{X} is uniquely characterised by the joint cdf

Preuve

Restatement of probability measure on \mathbb{R}^n are in correspondence with joint cdf □

Lemme 32

X_1, \dots, X_n random variables (Ω, \mathcal{F}) are independent if and only if

$$F_{\bar{X}}(x_1, \dots, x_n) = \prod_i F_i(x_i)$$

Transformations of random vectors**Proposition 33**

\bar{X} is a \mathbb{R}^n valued random vector and $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuous then $\phi(\bar{X})$ is a \mathbb{R}^m valued random vector with values in \mathbb{R}^m

Corollaire 34

Let X_1, \dots, X_n be random variables on Ω then $\sum a_i X_i$ is a random variable

Definition 21 (Random vectors with density)

Let $\bar{X} = (X_1, \dots, X_n)$ random vector then

$$f_{\bar{X}} : \mathbb{R}^n \rightarrow [0, \infty)$$

Riemann integrable with $\int_{\mathbb{R}^n} f_{\bar{X}}(y) dy = 1$ if $\forall [a_0, b_1] \times \dots [a_n, b_n] = B$ we have

$$\mathbb{P}(\{X_1 \in [a_0, b_1]\} \cap \dots) = \int_B f_{\bar{X}}(y) dy$$

Gaussian vectors

Let $\bar{\mu} \in \mathbb{R}^n$ and C positive definite $n \times n$ matrix called covariance matrix then the density

$$f_{\bar{\mu}, C}(\bar{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det C^{\frac{1}{2}}} \exp(-2(\bar{x} - \bar{\mu})^T C^{-1}(\bar{x} - \bar{\mu}))$$

Lecture 9: Expectation

Wed 17 Nov

Proposition 35

\bar{X} random vector on \mathbb{R}^n , $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ a C^1 -diffeomorphism with $J = \det D\Phi \neq 0$, then $\Phi(\bar{X})$ is a random vector with density

$$f_{\Phi(\bar{X})}(\bar{y}) = \frac{1}{|J_\Phi(\Phi^{-1}(\bar{y}))|} f_{\bar{X}}(\Phi^{-1}(\bar{y}))$$

Corollaire 36

X, Y independent r.v. with density f_X, f_Y , the density of $X + Y$ is then equal to

$$f_{X+Y}(y) = \int_{\mathbb{R}} f_X(x) f_Y(y-x) dx$$

Preuve

$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2, (x_0, y_0) \rightarrow (x_0, x_0 + y_0)$.

Φ is a nice diffeo with $J = 1$, then

$$f_{X, X+Y}(x_0, y_0) = f_{X,Y}(x_0, y_0 - x_0) = f_X(x_0) f_Y(y_0 - x_0) \quad \square$$

5.1 Conditional law

Definition 22

Let (X, Y) be a discrete random vector. Let S_X be the support of X and S_Y the support of Y , then $\forall x \in S_X$ the conditional law of Y given $X = x$ defined as

$$\forall y \in S_Y : \mathbb{P}(Y = y | X = x) = \frac{\mathbb{P}\{X = x\} \cap \{Y = y\}}{\mathbb{P}(X = x)}$$

What about continuous r.v.?

In general, no recipe, but all is good for r.v. with density

Definition 23 (Conditional law of rv with density)

Let (X, Y) be r.v. with density, suppose $f_X(x_0) > 0$, then the conditional density of Y given $X = x_0$

$$f_{Y|X}(y) = \frac{f_{X,Y}(x_0, y)}{f_X(x_0)}$$

defines a density of a random variable.

Lemme 37

$(X, Y) \sim N(\mu, C)$, then the conditional law of Y given $X = x_0$ is again a gaussian.

6 Mathematical Expectation

Definition 24 (Expectation for discrete r.v.)

Let X be a discrete r.v. with support S_X , we call X integrable if

$$\sum_{s \in S_X} |s| \mathbb{P}(X = s) < \infty$$

and if X is integrable, we define

$$\mathbb{E}[X] = \sum_{s \in S_X} s \mathbb{P}(X = s)$$

to be the expectation.

Remarque

$\mathbb{E}[X]$ only depends on \mathbb{P}_X , does not determine \mathbb{P}_X

Proposition 39

Let X, Y be integrable and discrete r.v.

— Linearity

$$\mathbb{E}[\lambda X + \beta Y] = \lambda \mathbb{E}X + \beta \mathbb{E}Y$$

—

$$|\mathbb{E}X| \leq \mathbb{E}|X|$$

Preuve

$$\mathbb{E}[X + Y] = \sum_{s \in S_{X+Y}} s \mathbb{P}[X + Y = s]$$

Notice

$$\begin{aligned} \mathbb{P}(X + Y = s) &= \sum_{x \in S_X} \sum_{y \in S_Y} \mathbb{P}(\{X = x_0\} \cap \{Y = y_0\}) 1_{s=x_0+y_0} \\ \mathbb{E}[X + Y] &= \sum_{x_0} \sum_{y_0} \sum_s s 1_{x_0+y_0=s} \mathbb{P}(\{X = x_0\} \cap \{Y = y_0\}) = \mathbb{E}X + \mathbb{E}Y \quad \square \end{aligned}$$

Corollaire 40

Let X, Y be integrable r.v. s.t.

$$\mathbb{P}(X \geq Y) = 1 \implies \mathbb{E}X \geq \mathbb{E}Y$$

Lecture 10: Expectation

Wed 24 Nov

6.1 Expected value for general random variables

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$.

We discretize X , define

$$\forall n \geq 1, X_n(\omega) = 2^{-n} \lfloor 2^n X(\omega) \rfloor$$

X_n is a random variable.

Exo that

$$X_n(\omega) \leq X(\omega) \leq X_n(\omega) + 2^{-n}$$

Proposition 41

Let X be a random variable, we say that X is integrable if

$$\mathbb{E}|X_1| < \infty$$

and then, $\forall n \geq 1, \mathbb{E}|X_n| < \infty$.

In which case, the limite

$$\lim_{n \rightarrow +\infty} \mathbb{E}X_n$$

exists and we define $\mathbb{E}|X| = \lim \mathbb{E}|X_n|$

Preuve

The property above implies

$$X_n(\omega) - 1 \leq X_1(\omega) \leq X_n(\omega) + 1 \quad \forall \omega \in \Omega$$

To show that the limit exists, we show that the sequence $(\mathbb{E}X_n)$ is cauchy.

Take $n \in \mathbb{N}, m \geq n$

$$|\mathbb{E}X_n - \mathbb{E}X_m| = |\mathbb{E}(X_n - X_m)|$$

But now $X_n - X_m \leq 2^{-n+1}$, and hence $\mathbb{E}|X_n - X_m| \rightarrow 0$

□

Proposition 42

Properties of \mathbb{E}

- linear
- $\mathbb{E}X \leq \mathbb{E}|X|$
- $\mathbb{P}(X \leq Y) = 1 \Rightarrow \mathbb{E}X \leq \mathbb{E}Y$

Preuve

By discretization

□

Proposition 43

X a r.v. with density, then X is integrable iff

$$\int_{\mathbb{R}} |y| f_X(y) dy < \infty$$

and then

$$\mathbb{E}X = \int_{\mathbb{R}} y f_X(y) dy$$

Preuve

Take X_n the discretization of X , we have

$$\mathbb{E}X_n = \sum_{k \in \mathbb{Z}} k 2^{-n} \mathbb{P}(X_n = k 2^{-n})$$

Then

$$\mathbb{P}(X_n = k 2^{-n}) = \int_{k 2^{-n}}^{(k+1) 2^{-n}} f_X(y) dy$$

$$\mathbb{E}X_n = \sum_{k \geq 2} \int_{k 2^{-n}}^{(k+1) 2^{-n}} k 2^{-n} f_X(y) dy$$

$$\mathbb{E}X_n \leq \sum_{k \in \mathbb{Z}} \int_{k 2^{-n}}^{(k+1) 2^{-n}} y f_X(y) dy$$

and

$$\mathbb{E}X_n \geq \sum_{k \in \mathbb{Z}} \int_{k 2^{-n}}^{(k+1) 2^{-n}} (y - 2^{-n}) f_X(y) dy$$

and hence

$$\mathbb{E}X_n \rightarrow \int_{\mathbb{R}} y f_X(y) dy$$

□

Proposition 44

Take \bar{X} some random vector in \mathbb{R}^n and

$$\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$$

measurable.

— \bar{X} is discrete, then $\Phi(\bar{X})$ is also discrete, so if

$$\Phi(\bar{X})$$

integrable, then

$$\mathbb{E}\Phi(\bar{X}) = \sum_{s \in S_{\Phi(\bar{X})}} s \mathbb{P}(\Phi(\bar{X}) = s) = \sum_{x \in S_X} \Phi(x) \mathbb{P}(\bar{X} = x_0)$$

Similarly if \bar{X} is a random vector with density, $\Phi(\bar{X})$ is integrable and "nice", then

$$\mathbb{E}\Phi(\bar{X}) = \int_{\mathbb{R}^n} \Phi(x) f_{\bar{X}}(x) dx$$

Preuve

Sketch for 2 :

Discretize Φ and write

$$\mathbb{E}\Phi_n = \sum_{k \in \mathbb{Z}} k 2^{-n} \mathbb{P}(\Phi_n = k 2^{-n})$$

□

Proposition 45

$X \sim Y$ iff

$\forall g : \mathbb{R} \rightarrow \mathbb{R}$ continuous and bounded

$$\mathbb{E}g(X) = \mathbb{E}g(Y)$$

Preuve

One direction is obvious.

In the other direction, note that

$$X \sim Y \iff \forall t \in \mathbb{R} F_X(t) = F_Y(t)$$

but

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{E}1_{X \leq t}$$

hence $X \sim Y$ iff

$$\forall t \in \mathbb{R} \mathbb{E}1_{X \leq t} = \mathbb{E}1_{Y \leq t}$$

We approximate $1_{X \leq t}$ by a continuous function.

□

Proposition 46

If X, Y independent, $g(X), h(Y)$ integrable, then

$$\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y)$$

Furthermore, if $\forall h, g$ continuous and bounded

$$\mathbb{E}h(X)g(Y) = \mathbb{E}h(X)\mathbb{E}g(Y)$$

then X and Y independent.

Preuve

X, Y are independent iff $F_{X,Y}(t) = F_X(t)F_Y(t)$, hence

$$\mathbb{E}1_{X \leq t_1, Y \leq t_2} = \mathbb{E}1_{X \leq t_1} \mathbb{E}1_{Y \leq t_2}$$

We then approximate as in the last proof.

For the second part, using the exercise sheet shows that

$$X, Y \text{ independent} \implies g(X)h(Y) \text{ independent}$$

so it suffices to show that $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.

We have to prove in two steps

- X, Y discrete
- via discretization X_n, Y_n .

Note that

$$\mathbb{E}XY = \sum_{S_{XY}} s\mathbb{P}(XY = s)$$

and

$$\begin{aligned} \mathbb{E}X\mathbb{E}Y &= \sum_{S_X} \sum_{S_Y} x_0 y_0 \mathbb{P}(X = x_0) \mathbb{P}(Y = y_0) \\ \mathbb{E}XY &= \sum_{S_X} \sum_{S_Y} xy \mathbb{P}(X = x, Y = y) \sum_{S_{X,Y}} 1_{s=xy} \\ &= \sum_{S_{X,Y}} \sum_{S_X} \sum_{S_Y} s 1_{s=xy} \mathbb{P}(X = x, Y = y) = \mathbb{P}(XY = s) \end{aligned} \quad \square$$

Lecture 11: Variance

Wed 01 Dec

6.2 Variance and covariance

Definition 25 (Variance)

Let X be an integrable random variable s.t. X^2 is also integrable, then we define the variance of X as

$$\text{var}(X) = \mathbb{E}((X - \mathbb{E}X)^2)$$

and it is well defined

Then $\sigma(X) = \sqrt{\text{var}(X)}$ is called the standard deviation.

Remarque

$$(X - \mathbb{E}X)^2 \leq |X|^2 + 2|X|\mathbb{E}X + (\mathbb{E}X)^2 < 2(|X|^2 + \mathbb{E}|X|^2)$$

Note that if $\text{Var}(X) = 0$, then $X = \mathbb{E}X$ almost surely. Furthermore

$$\mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}X) + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

Proposition 48 (Chebyshev inequality)

X integrable random variable with finite variance

$$\mathbb{P}(|X - \mathbb{E}X| > t) \leq \frac{\text{Var}X}{t^2}$$

Preuve

If Y is positive and integrable, we have

$$\mathbb{P}(Y > t) = \frac{\mathbb{E}Y}{t}$$

and we apply this to $Y = (X - \mathbb{E}X)^2$. □

Definition 26 (Covariance)

Let X, Y be integrable, with finite variance, then the covariance of X and Y is defined as

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$$

and we call

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X \text{Var}Y}}$$

the correlation.

As long as the variances are different from 0.

6.3 Moments of a random variable

Definition 27 (Moment)

Let X be a random variable such that $|X|^n$ is integrable.

Then we say that X admits a n-th moment $\mathbb{E}X^n$.

Why is this useful?

- We get control on the tails, if X admits a n-th moment, then $\mathbb{P}(|X|^n > t) \leq \frac{\mathbb{E}|X|^n}{t^n}$
- Sometimes determines the law

Proposition 49

Let X, Y be random variable such that $\mathbb{P}(X)(X \in [-C, C]) = \mathbb{P}(Y \in [-C, C]) = 1$ for some $C \in \mathbb{R}$.

Then $X \sim Y$ if and only if $\forall n \mathbb{E}X^n = \mathbb{E}Y^n$

Preuve

Notice that X, Y admit n-th moments for all n as they are bounded by C.

Clearly if X, Y are equal in law, then X^n, Y^n are equal in law.

Theorème 50 (Stone- Weierstrass)

Let g be a continuous function on $[-c, c]$.

Then $\forall \epsilon > 0$, there exists a polynomial P_ϵ s.t. $\sup_{x \in [-c, c]} |g(x) - P_\epsilon(x)| < \epsilon$

We will use this theorem to prove the proposition.

It suffices to show that $\forall g$ continuous and bounded $\mathbb{E}g(X) = \mathbb{E}g(Y)$.

Observe that \forall polynomial P , $\mathbb{E}P(X) = \mathbb{E}P(Y)$ by linearity of \mathbb{E} .

Given g continuous and bounded, notice that $\mathbb{E}g(X) = \mathbb{E}\tilde{g}(X)$.

This is because $\mathbb{P}(X \in [-C, C])$.

Now pick $\epsilon > 0$ and apply Stone-Weierstrass to \tilde{g} .

This gives P_ϵ s.t. $\sup_{x \in [-C, C]} |\tilde{g}(x) - P_\epsilon(x)| < \epsilon$.

Now

$$|\mathbb{E}g(X) - \mathbb{E}g(Y)| \leq |\mathbb{E}g(X) - \mathbb{E}P_\epsilon(X)| + |\mathbb{E}P_\epsilon(Y) - \mathbb{E}g(Y)| \quad \square$$

Remarque

— Holds more generally, but not always :

Sometimes moments don't exist

Sometimes moments grow too fast and don't characterise uniquely

6.4 Moment Generating functions**Definition 28**

Let X be a random variable s.t. $\exists c > 0$ s.t. $\forall t \in [-c, c]$, the function $\exp(tX)$ is integrable, then we define the moment generating function

$$M_x(t) = \mathbb{E}(\exp(tX))$$

Remarque

Might not exist even when all moments exist.

Theorème 53 (MGF determines the law)

If X, Y are random variables that admit MGF in some interval $[-c, c]$, then

$$X \sim Y \iff \forall (-t, t) M_X(t) = M_Y(t).$$

We can also define MGF for random vectors

Definition 29 (MGF for random vectors)

Let (X_1, \dots, X_N) be random vectors s.t. $\forall t$

$$\exp \langle t, \bar{X} \rangle$$

is integrable, then MGF of \bar{X} is defined as

$$M_{\bar{X}}(t) = \mathbb{E} \exp(\langle t, X \rangle)$$

Theorème 54

If X, Y are random vectors admitting MGF in some $(-c, c)^n$, then $X \sim Y \iff M_X(t) = M_Y(t)$

Lecture 12: Sequences of random variables

Wed 08 Dec

6.5 MGF vs independence**Proposition 55**

Let X, Y be r.v. s.t. M_X, M_Y exist for $t \in (-c, c)$.
 X, Y are independent iff

$$M_X(t)M_Y(s) = M_{X,Y}(t, s) \quad \forall t, s \in \mathbb{R}$$

Preuve

If X, Y are independent, then

$$\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(y)$$

And we apply to $\exp tX, \exp tY$.

Take \tilde{X}, \tilde{Y} independent s.t. $\tilde{X} \sim X, \tilde{Y} \sim Y$, then the moment generating function of \tilde{X}, \tilde{Y} agree. But then

$$M_{X,Y} = M_X M_Y = M_{\tilde{X}} M_{\tilde{Y}} = M_{\tilde{X}, \tilde{Y}}$$

Hence X, Y are independent. □

Proposition 56

If $\tilde{X} \sim N(\mu, C)$, then it's MGF is

$$M_{\tilde{X}} = \exp(\langle \mu, t \rangle + \frac{1}{2} t^T C t)$$

7 Limit Theorems**7.1 Sequences of events**

Let E_1, \dots be a sequence of events.

Define $F_1 = \{ \text{only finitely many } E_i \text{ occur} \}$.

Do we have $F_1 \in \mathcal{F}$? Yes.

Can we estimate the probability of these events

Lemme 57

Let E_1, \dots be mutually independent. Then at least one of E_i happens almost

surely (ie. $\mathbb{P}(\cup E_i) = 1$) if and only if

$$\prod (1 - \mathbb{P}E_i) \rightarrow 0$$

Preuve

Write

$$\mathbb{P}(\cup E_i) = 1 - \mathbb{P}(\cap E_i^c) \geq 1 - \mathbb{P}(\bigcup_{n \leq i \leq 1} E_i) \geq 1 - \prod (1 - \mathbb{P}E_i) \quad \square$$

Theorème 58 (Borel-Cantelli I)

Let E_1, \dots be events s.t. $\sum \mathbb{P}(E_i) < \infty$, then the probability that only finitely many happen.

Then a.s. only finitely many E_i occur

$$\mathbb{P}(F_3) = \mathbb{P}(\bigcap_{n \geq 1} \bigcup_i E_i) = 0$$

Preuve

We want to show that $\mathbb{P}(\bigcap_{n \geq 1} \bigcup_i E_i) = 0$.

So

$$\mathbb{P}(F_3) \leq \mathbb{P}(\bigcup_{i \geq n} E_i) \leq \sum_{i \geq n} \mathbb{P}(E_i) \rightarrow 0 \quad \square$$

Exemple

Let $X_1, X_2 \dots$ be Geometric random variables with value $\frac{1}{2}$, s.t. $\mathbb{P}(X_i > k) = 2^{-k}$.

Let $E_n = \{\max_{i \in [n]} X_i > n\}$.

Do infinitely many E_n occur?

$$\mathbb{P}(E) \subset \mathbb{P}(\bigcup \{X_i > n\}) \leq n 2^{-n}$$

Hence, the sum $\sum_{n \geq 1} \mathbb{P}(E_n) < \infty$.

So we can apply BC-I, and hence

$$\mathbb{P}(\text{infinitely many } E_i \text{ occur}) = 0$$

Is there a criteria for infinitely many events to happen?

Theorème 60 (Borel-Cantelli II)

Let $E_1 \dots$ be independent events s.t. $\sum_{i \geq 1} \mathbb{P}(E_i) = \infty$, then $\mathbb{P}(\bigcap_{n \geq 1} \bigcup_{i \geq n} E_i) = 1$

Preuve

We want

$$\mathbb{P}((\bigcap_{n \geq 1} \bigcup_{i \geq n} E_i)^c) \leq \sum_{n \geq 1} \mathbb{P}(\bigcap_{i \geq n} E_i^c)$$

We need that $\forall n$

$$\mathbb{P}\left(\bigcap_{i \geq n} E_i^c\right) = 0$$

We have

$$\mathbb{P}\left(\bigcap_{i \geq n} E_i^c\right) = \prod_{i \geq n} (1 - \mathbb{P}(E_i))$$

Now using that $1 - x \leq e^{-x}$, hence

$$\prod_{i \geq n} (1 - \mathbb{P}(E_i)) \leq \exp\left(-\sum_{i \geq n} \mathbb{P}(E_i)\right) \rightarrow 0 \quad \square$$

7.2 Sequences of random variables

Do we have infinitely many $X_i \geq 0$, does X_n converge, is X_n bounded? In all cases, we have to see that our questions make sense!

Definition 30 (Almost sure convergence)

Let X_1, X_2, \dots be random variables defined on the same probability space.

Then we say that X_i converges to X_0 almost surely if

$$\mathbb{P}(\{\omega : X_n(\omega) \rightarrow X_0(\omega)\}) = 1$$

Is the set $\{\omega : X_n(\omega) \rightarrow X_0(\omega)\}$ even measurable?

Definition 31 (Convergence in law)

Let X_1, \dots be random variable.

We say that X_n converges to X_0 in law if

$$F_n(t) = \mathbb{P}(X_n \leq t)$$

converges to $F_X(t)$ as $n \rightarrow \infty$ for all t s.t. $\mathbb{P}(X = t) = 0$ ie. t is a continuity point of F_X

Lecture 13: LLN

Wed 15 Dec

Lemme 61

Let X_1, \dots be a sequence of r.v. $(X_n) \rightarrow X_0$ iff $\forall a < b$ continuity points of F_{X_0} ,

$$\mathbb{P}(X_n \in (a, b)) \rightarrow \mathbb{P}(X_0 \in (a, b))$$

Preuve

If a, b are continuity points, then

$$\mathbb{P}(X_0 \in (a, b)) = F_{X_0}(b) - F_{X_0}(a)$$

In the other direction, fix b a continuity point of F_{X_0} , observe $\forall a_n \rightarrow -\infty$ continuity points of F_{X_0} , now

$$F_{X_n}(b) - \mathbb{P}(X_n \leq b) \geq \mathbb{P}(X_n \in (a_k, b))$$

by assumption $\rightarrow \mathbb{P}(X_0 \in (a_k, b))$ which goes to $F_{X_0}(b)$ as $k \rightarrow \infty$ hence $\liminf F_{X_n}(b) \geq F_{X_0}(b)$

Now consider instead (b, c_k) , then similarly

$$\liminf \mathbb{P}(X_n > b) \geq \mathbb{P}(X_0 > b)$$

but $\mathbb{P}(X_0 < b) + \mathbb{P}(X_0 > b) = 1$ and

$$\liminf \mathbb{P}(X_n < b) + \liminf \mathbb{P}(X_n > b) \leq \liminf \mathbb{P}(X_n < b) + \mathbb{P}(X_n > b) = 1$$

□

Remarque

In fact a similar proof gives $\iff \forall (a, b)$

$$\liminf \mathbb{P}(X_n \in (a, b)) \geq \mathbb{P}(X_1 \in (a, b))$$

7.3 Convergence in law vs almost sure convergence

Convergence in law $\not\implies$ a.s. convergence.

Let X_1, X_2, \dots be i.i.d. $Ber(\frac{1}{2})$ random variables.

Now (X_i) converge in law to $Ber(\frac{1}{2})$ but they don't converge almost surely since $X_i(\omega)$ converges iff becomes constant.

Define $E_n = \{X_k \text{ constant on } [2^{n-1}, 2^n]\}$.

$$\mathbb{P}(E_n) = 2^{-2^{n-1}}$$

Since $\sum \mathbb{P}(E_n) < \infty$

$$B \subset I \implies \mathbb{P}(\{\inf \text{ many } E_i \text{ occur}\}) = 0$$

Proposition 63

If $X_0, X_1 \dots$ random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, if $(X_n) \rightarrow X_0$ almost surely then $(X_n) \rightarrow X_0$ in law.

Lemma 64

$\forall \epsilon > 0, \mathbb{P}(|X_n - X_0| > \epsilon) \rightarrow 0$.

Proof of proposition

Let t be a continuity point of F_{X_0} , hence

$$F_{X_0}(t) = \lim_{m \rightarrow +\infty} F_{X_0}(t + \frac{1}{m}) = \lim_{m \rightarrow -\infty} F_{X_0}(t - \frac{1}{m})$$

$$\begin{aligned} \text{Now } F_{X_n}(t) &= \mathbb{P}(X_n \leq t) = \mathbb{P}(X_n \leq tX_0 \leq t + \frac{1}{m}) + \mathbb{P}(X_n \leq tX_0 > t + \frac{1}{m}) \\ &\leq \mathbb{P}(X_0 \leq t + \frac{1}{m}) + \mathbb{P}(|X_0 - X_n| > \frac{1}{m}) \implies F_{X_n}(t) \leq F_{X_0}(t + \frac{1}{m}) + \mathbb{P}(|X_0 - X_n| > \frac{1}{m}) \end{aligned}$$

Similarly we get working with $F_{X_0}(t - \frac{1}{n})$

$$F_{X_n}(t) \geq F_{X_0}(t - \frac{1}{m}) - \mathbb{P}(|X_0 - X_n| > \frac{1}{m}) \quad \square$$

Preuve (Of lemma)

$\{X_n \text{ converges to } X_0\}$ has probability 1

$$\{X_n \rightarrow X_0\} \subset \bigcup_{m \in \mathbb{N}} \{|X_n - X_0| \leq \epsilon \forall n \geq m\}$$

So let $E_m = \{|X_n - X_0| \leq \epsilon \forall n \geq m\}$, now since $E_m \subset E_{m+1}$

$$1 = \mathbb{P}(\{X_n \rightarrow X_0\}) \leq \mathbb{P}(\bigcup E_m) = 1$$

since $E_m \subset E_{m+1}$, then $\lim \mathbb{P}(E_m) \rightarrow 1$.

But now $\mathbb{P}(|X_n - X_m| \geq \epsilon) \leq \mathbb{P}(E_m^c) \rightarrow 0$ \square

Definition 32 (Convergence in probability)

Let X_0, \dots be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

We say (X_i) converges in probability to X_0 if $\mathbb{P}(|X_n - X_0| > \epsilon) \rightarrow 0 \forall \epsilon > 0$

8 Law of large numbers

Let X_1, \dots be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, set

$$S_n = \frac{\sum_i X_i}{n}$$

How does S_n behave?

— In general this might be difficult, take for example

$$X_1 = X_2 = \dots$$

— In case of sufficient independence, there is an averaging effect $S_n \sim \mathbb{E}X_1$

Theorème 65 (Weak LLN)

Let X_1, X_2, \dots be i.i.d. integrable random variables, then

$$\mathbb{P}(|S_n - \mathbb{E}X_1| > \epsilon) \rightarrow 0$$

Theorème 66 (Strong LLN)*Let X_1, X_2, \dots be i.i.d. integrable*

$$S_n \rightarrow \mathbb{E}X_1$$

Clearly $SLLN \implies WLLN$

Remarque*Hypothesis can be weakened*

- Don't need identical distribution
- Don't need full independence

Remarque*Limit of S_n depends only on $\mathbb{E}X_1$* **Preuve (WLLN)***Assuming $\mathbb{E}X_1^2 < C$* *Idea : work with expectation and use Markov.*

$$\mathbb{P}(|S_n - \mathbb{E}X_1| > \epsilon) = \mathbb{P}(|S_n - \mathbb{E}X_1|^2 > \epsilon^2) \leq \frac{\mathbb{E}(S_n - \mathbb{E}X_1)^2}{\epsilon^2} = \epsilon^{-2} \mathbb{E}\left(\sum (X_i - \mathbb{E}X_1) \frac{1}{n}\right)$$

Now

$$= \frac{1}{n^2} \sum_{i,j} \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j) = \frac{1}{n^2} \sum_i \mathbb{E}(X_i - \mathbb{E}X_i)^2 \leq \frac{1}{n^2} nC \rightarrow 0 \quad \square$$

Lecture 14: Central limit theorem

Wed 22 Dec

Preuve (of SLLN)*We suppose $\mathbb{E}X_1^4 < C$.**Recall that $\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\epsilon}{n}$.**Instead of $\mathbb{E}(S_n - \mathbb{E}X_1)^2$, we look at $\mathbb{E}(S_n - \mathbb{E}X_1)^4$.**Define*

$$A_{ij} = \mathbb{E}((X_i - \mathbb{E}X_i)^2(X_j - \mathbb{E}X_j)^2)$$

*We can now apply Cauchy-Schwarz to see that $A_{ij} \leq C'$.**By Markov*

$$\mathbb{P}(|S_n - \mathbb{E}X_1| > \epsilon) = \mathbb{P}(|S_n - \mathbb{E}X_1|^4 > \epsilon^4) \leq \frac{c}{n^2} \quad \square$$

*Hence the event $\{|S_n - \mathbb{E}X_1| > \epsilon\}$ can only happen finitely many times.***8.1 Central limit theorem**

Theorème 69 (Central limit theorem)

Let X_1, \dots , be i.i.d. random variables with $\text{Var}X_1 = \sigma^2$, then

$$\frac{\sum_{i=1}^n (X_i - \mathbb{E}X_i)}{\sqrt{n}} \rightarrow \mathcal{N}(0, \sigma^2)$$

Observe that we can assume $\mathbb{E}X_i = 0$ by adding a constant and $\text{Var}X_1 = 1$ by scaling.

Preuve

The idea is that swapping X_i for $\mathcal{N}(0, 1)$ shouldn't influence the sum.

We will show we can swap all of them.

We prove it under the assumption that $\mathbb{E}|X_i|^3 < \infty$.

Proposition 70 (Lindeberg Exchange principle)

Given a sequence of random variables X_1, \dots with 0 mean and variance 1, then

$$S_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}, |g|, |g'|, |g''|, |g'''| < C \text{ then } \mathbb{E}(g(S_n) - g(X)) \rightarrow 0$$

Lindeberg exchange principle implies CLT.