• Probability space $(\Omega, \mathcal{A}, P)$

→ We need **three** ingredients to build a probability space:

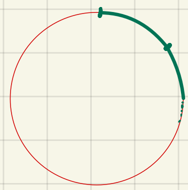
What is a set?

1. Ω, the sample space:

   - in mathematical terms it's a **set**
   - in probabilistic terms it's the set of all possible outcomes of any experiment

Example:

- all points on the edge of a unite circle



2. $\mathcal{A}$, the event space: collection of events

$\mathcal{A}$ is a σ-algebra, a class of subsets of Ω such that:

- Contains both empty set and Ω:

$$\{\emptyset, \Omega\} \in \mathcal{A}$$

- It's closed under complement:

complementary

$$A \in \mathcal{A}, \quad A^c \in \mathcal{A}$$

What countable means?

- It's closed under countable union:

$$\bigcup_i A_i \in \mathcal{A}$$

if $A_1, A_2, \dots, A_i$ is a finite or countable collection of elements of $\mathcal{A}$

$\Downarrow$

This third requirement implies that $\mathcal{A}$ is closed under countable intersection:

$$\bigcap_i A_i = \left(\bigcup_i A_i^c\right)^c \in \mathcal{A}$$

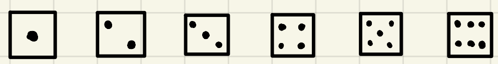
in other terms, it's verified the Morgan Law

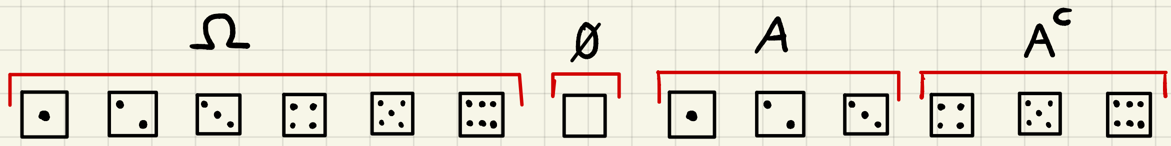
- From a probabilistic point of view

  $\mathcal{A}$ is a collection of the events that you are interested to evaluate: in other terms, the events which are meaningful for the problem you investigate.

Example:

Let's take a six-faced dice:





✱ Complete with the discussion of the Borel σ-algebra

3. P is a probability measure: a function from $\mathcal{A}$ to $[0;1]$. $P: \mathcal{A} \to \mathbb{R}$

that satisfeies those 3 properties

i) $0 \leq P(A) \leq 1$

ii) $P(\Omega) = 1$

iii) $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$

this is also called
δ- additivity

if $A_1, A_2, \dots, A_i \in \mathcal{A}$ and
$A_i \wedge A_j \neq \emptyset \quad \forall i \neq j$

disjoint

⊛ If P satisfies (i) and (ii), but is replaced by $P(\phi) = 0$, then P is called a measure.

⇒ It's important to remind that a probability measure can be a measure but a measure does not implies a probability measure.

Can you do an example?

Example of measure:

The Lebesgue measure is the measure m on the Borel δ-field such that

$$m(a, b] = b - a$$, called length

Note:

real line

$$m(\mathbb{R}) = +\infty$$ → so m is not a probability measure

$$m\{x\} = 0 \quad \forall x \in \mathbb{R}$$

point

Let see some consequences of the definition of probability measure

1) $P(A^c) = 1 - P(A)$

proof:

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

2) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

proof

3) $P(A \cup B \cup C) = P(A) + P(B) + P(C) + P(A \cap B \cap C) - P(A \cap B) - P(A \cap C) - P(B \cap C)$

proof

4) If $B \supset A$, then $P(B) \geq A$

proof

$$P(B) = P\left(A \cup (B-A)\right)$$

$$= P(A) + P(B-A)$$

$\downarrow$

this is ≥ 0

• Conditional probability

Now, if $\boxed{P(B) > 0}$ then, by definition,

$$P(A|B) = \frac{P(A \wedge B)}{P(B)};$$

If $P(B) = 0$, we not define $P(A|B)$.

def.

A partition of Ω it's any collection of subsets of Ω such that the union of the subsets is Ω and the subsets are pairwise disjoint.

Theorem of total probability

If $B_1, B_2, \dots, B_n$ is a finite or countable partition of Ω, then:

What is a partition?

$$P(A) = \sum_n P(A \mid B_n) \cdot P(B_n)$$

proof

$$P(A) = P(A \wedge \Omega) = P\left[A \wedge \left(\bigcup_n B_n\right)\right]$$

$$= P\left[\bigcup_n (A \wedge B_n)\right]$$

→ multypling and dividing by $P(B_n)$

$$= \sum_n P(A \wedge B_n) \cdot \frac{P(B_n)}{P(B_n)}$$

$$= \frac{\sum_n P(A \wedge B_n)}{P(B_n)} \cdot P(B_n)$$

$$= \sum_n P(A \mid B_n) \cdot P(B_n)$$

Why the theorem of the total probability it's so important?

→ Suppose we want to evaluate two probability measures:

1) $P(A)$

2) $P(A|B)$

We can easily say that the 2) is easier to calculate cause we know $P(B)$; instead, we don't have any information about $P(A)$.

- Let's make an experiment:

  - Suppose to have an urne with a white and a black ball:

$$\omega = \omega_1 + \omega_2$$

Let A be the event:

$$A = \{ \text{white ball at the } 2^\circ \text{ trial} \}$$

What is $P(A) = ?$

Let's define the possible events that may occure in the first trial B_1 and B_2

$$B_1 = \{\text{white ball at } 1° \text{ trial}\}$$

$$B_2 = \{\text{black ball at } 1° \text{ trial}\}$$

Now, with the theorem of total probability,

$$P(A) = P(A \mid B_1) \cdot P(B_1) + P(A \mid B_2) \cdot P(B_2)$$

$$= \frac{n_1 - 1}{n-1} \cdot \frac{n_1}{n} + \frac{n_1}{n-1} \cdot \frac{n_2}{n}$$

$$= \frac{\cancel{n_1^2} - n_1 + n_1 \cdot n - \cancel{n_1^2}}{n(n-1)} = \frac{n_1 \cancel{(n-1)}}{n \cancel{(n-1)}}$$

$$= \frac{n_1}{n}$$

- This experiment can be generalized

$$A = \{\text{white ball at the } j° \text{ trial}\}$$

$$j = 1, \dots, n$$

Once again, $P(A) = \dfrac{n_1}{n}$

• Definitions of probability

- Now we want to define what is a probability: there are 3 main definitions.

1) Classical definition of probability by Laplace

2) The frequentist definition

3) The subjective definition

[the 2) will be skipped in this course]

1) Laplace definition

$$P(A) = \frac{\text{number of cases where } A \text{ is true}}{\text{total number of cases}}$$

Is this definition good enough? No.

Shortcomings:

1) We can apply this definition only to little situations

2) We have to assume that all cases are equipossible.

$\Downarrow$ In general, this definition is a tautology

3) Subjective definition by De Finetti

$P(A)$ = price that a coherent person allows to pay for winning 1 provided A is true.

What do we mean with the term "coherent"?

Suppose that $P(A)=2$: if A is true the person lose 1; if A is false the person lose 2. So no person that is coherent accept to bet on this type of event.

- Shortcoming:

  -) not everyone give the same probability of an event even if are all "coherent", but there is no reason that one opinion ($P(A)$) is more accurate than others.

(*) The current opinion is that it's impossible to give a proper definition

- So, what can we do?

- We can give an assiomatic definition of probability

$\Downarrow$

(we don't give a definition but we say it's properties)

So, $P(A)$ it's defined by Kolmogorov's assioms.

- Consequence of the theorem of total probability

- Bayes theorem

If $B_1, B_2, \dots, B_n$ is a finite or countable partition, and $P(A) > 0$, then:

$$P(B_i \mid A) = \frac{P(A \wedge B_i)}{P(A)} = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_j P(A \mid B_j) \cdot P(B_j)}$$

✱ to complete

- Notion of independence

Let $\xi \subset \mathcal{A}$ be a collection of events.

Say that ξ is independent if:

$$P(A_1 \wedge \dots \wedge A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$$

whenever $A_1, \dots, A_n \in \xi$ and $A_1, \dots, A_n$ are distinct

If $\xi = \{A, B\}$, A is independent from B,

then $P(A \wedge B) = P(A) \cdot P(B)$

Remarks:

1) If $P(B) = 0 \quad \vee \quad P(B) = 1$, then

$A \perp B \quad \forall A \in \mathcal{A}$

proof

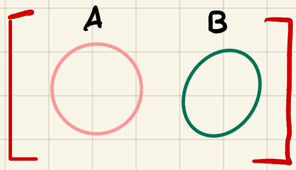
- if $P(B) = 0$; $P(A \wedge B) = 0 = P(A) \cdot \overbrace{0}^{P(B)} = P(A) \cdot P(B)$
- if $P(B) = 1$; $P(A \wedge B) = P(A) = P(A) \cdot \overbrace{1}^{P(B)} = P(A) \cdot P(B)$

- Suppose:

$$\begin{cases} \bullet\ \Omega = \mathbb{R} \\ \bullet\ A = \{\pi\} \\ \bullet\ B = \{x \in \mathbb{R} : \sin(x) = 0\} \end{cases}$$

If P is such that $P\{\pi\} = 0$, then I am obliged to say that $A \perp B$

- Suppose $A \wedge B = \emptyset$



$\downarrow$

However, if $P(B) = 0$ or $P(B) = 1$ we must consider that $A \perp B$

- Note:

- if $A \perp B$, then
  - $A \perp B^c$
  - $A^c \perp B$
  - $A^c \perp B^c$

- if $P(B) > 0$,

$$A \perp B \overset{\text{implies}}{\longleftrightarrow} P(A \mid B) = P(A)$$

- if ξ has more than 2 elements,

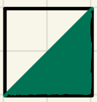
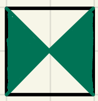
to prove that ξ is independent,
it is not enough to show that

$$P(A \wedge B) = P(A) \cdot P(B) \quad \forall A, B \in \xi$$
with $A \neq B$

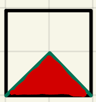
In fact:

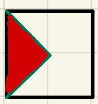
- Example

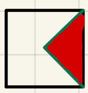
Suppose $\Omega = [0, 1]^2$ (the unit squared)
and $\xi = \{A, B, C\}$ where

$A =$  $B =$  $C =$ 

Suppose now that "P" is the area, then:

$P(A \wedge B) =$  $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(A \wedge C) =$  $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(B \wedge C) =$  $= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

but

$$P(A \wedge B \wedge C) = \square = \emptyset \neq \frac{1}{8} = P(A) \cdot P(B) \cdot P(C)$$

- Notion of random variable

Given the probability space and the measurable space $(S, \mathcal{B})$, a random variable is a measurable map $X : \Omega \to S$

?

What does it means map?

· What does it means "measurable"?

A measurable space is a pair $(S, \mathcal{B})$ such that S is a set and $\mathcal{B}$ is δ-field of subsets of S

Can you define a δ-field?

A Function $f : \Omega \to S$ is said to be measurable:

if $f^{-1}(B) \in \mathcal{A}$ $\forall B \in \mathcal{B}$ when $f^{-1}(B)$ denote the following set:

$$\left[f^{-1}(B) = \{ \omega \in \Omega : f(\omega) \in B \} \right]$$

So, if we take $f^{-1}(B) \in \mathcal{A}$ $\forall B \in \mathcal{B}$

$$\begin{cases} X : \Omega \to S \\ X^{-1}(B) \in \mathcal{A} \ \forall B \in \mathcal{B} \end{cases}$$

X is a random variable

⊕ if $S = R$ and $\mathcal{B} = \beta$ (the Borel δ-field)

X is said to be a real or univariate random variable.

⊕ if $S = R^n$ and $\mathcal{B} = \beta$ (the Borel δ-field on R^n)

X is said to be an n-variate random variable.

What is the intuitive meaning of a r.v.?

Given the probability space $(\Omega, \mathcal{B}, P)$:

first, we make the experiment and observe some $\omega \in \mathbb{R}$; then we make some measurament on ω and we obtain $X(\omega) \in S$

Let's make an example

Denote with Ω the people in the room and with ω a single person.

Let $X(\omega)$ be the measurament on height, we can say that $S = \mathbb{R}$

$$\begin{cases} \Omega = \text{people in the room} \\ \omega = \text{person} \\ X(\omega) = \text{measuvament of the height} \\ S = \mathbb{R} \end{cases}$$

So, if we make n measurament

$$X(\omega) = \{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\}$$

and $S = R^n$

Why the r.v. has to be measurable?

Given $(\Omega, \mathcal{B}, P)$ and a r.v. X, namely a measurable function $X : \Omega \to S$, we obtain another probability space $(S, \mathcal{B}, \nu)$ where

Why we obtain another prob. space?

ν is defined by:

$$\nu(B) = P[X^{-1}(B)] \quad \forall B \in \mathcal{B}$$

So, the probability ν is called the distribution of X and we write

$$X \sim \nu \qquad (\text{"}X \text{ has distribution } \nu\text{"})$$

$$\nu(B) = P\{\omega \in \Omega : X(\omega) \in B\}$$

- Since we want to define

$$\nu(B) = P[X^{-1}(B)] \quad \forall B \in \mathcal{B}$$

the set $X^{-1}(B)$ must belong to the domain $\mathcal{A}$ of P: we need that

$$[X^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}]$$

$\Downarrow$

[and this is exactly the notion of misurability (page 23)]

$\Rightarrow$ Now, let's suppose X is not measurable, then $X^{-1}(B) \notin \mathcal{A}$ for some $B \in \mathcal{B}$: so, in this case we can't define

$$\nu(B) = P[X^{-1}(B)]$$

because $X^{-1}(B)$ does not belong to the domain $\mathcal{A}$ of P

- Notation:

We will usually write:

$$\begin{cases} X^{-1}(B) = (X \in B) \\ \nu(B) = P(X \in B) \end{cases}$$

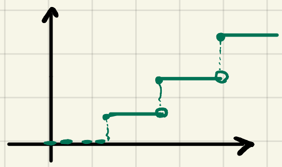
- Distribution function

• Let X be a real (univariate) random variable, then it's distribution function is defined by

$$F(x) = P[X \leq x] \quad \forall x \in \mathbb{R}$$

Then, $F: \mathbb{R} \to \mathbb{R}$ and satisfies this 3 properties

a) F is not decreasing



$$x \leq y \to F(x) \leq F(y)$$

b) F is right continuos, i.e.,

Why right?

$$F_X(x) = \lim_{y \to x^+} F(y) \quad \forall x \in \mathbb{R}$$

c) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$

So, if $F: \mathbb{R} \to \mathbb{R}$ is any function satisfying conditions a), b) and c), then we can say there exist a real random variable X such that

$$F(X) = P(X \leq x) \quad \forall x \in \mathbb{R}$$

→ to sum-up:

Given the $(\Omega, \mathcal{B}, P)$ and a measure, we can define:

$$\nu(B) = P(X \in B) \quad \forall B \in \mathcal{B}$$

that it's distributed in $(S, \mathcal{B}, \nu)$

And if $S = \mathbb{R}$ and $\mathcal{B} = \beta$,

$$F(X) = P(X \leq x) \quad \forall x \in \mathbb{R}$$

which we called distribution function.

And,

$$F(X) = \nu(-\infty ; x] \quad \forall x \in \mathbb{R}$$

- Titolo

Given F, let's evaluate $\nu\{x\}$

- $\nu\{x\} = \nu(-\infty, x] - \nu(-\infty, x)$

and by the previous theorem

$$= F(x) - \nu\left(\bigcup_{x=1}^{\infty} \left(-\infty ; x - \frac{1}{n}\right) \right)$$

* to complete

• theorem:

for $\forall$ distribution function F, $\exists!$, $a, b, c \geq 0$, $a+b+c=1$, such that:

$$F = a \cdot F_1 + b \cdot F_2 + c \cdot F_3$$

F_1 is a discrete distribution function

F_2 is a singular continuos function

F_3 is an absolute function

it may be that:

$a=1, b=0, c=0$ so F is a discrete density func.

$a=0, b=1, c=0$ so F is a singular continuos func.

$a=0, b=0, c=1$ so F is an absolute function

→ With this theorem we can define any distribution function

- Discrete case

We can say that **F** is discrete if $\exists$ (exist) $B \subset \mathbb{R}$ that is finite or countable such that

$$\nu(B) = 1$$

- For instance:

- if F is binomial of parameters m and B then $B = \{0, 1, 2, \dots, n\}$

- if F is a Poisson of parameter λ,

$$B = \{0, 1, 2, \dots, n\}$$

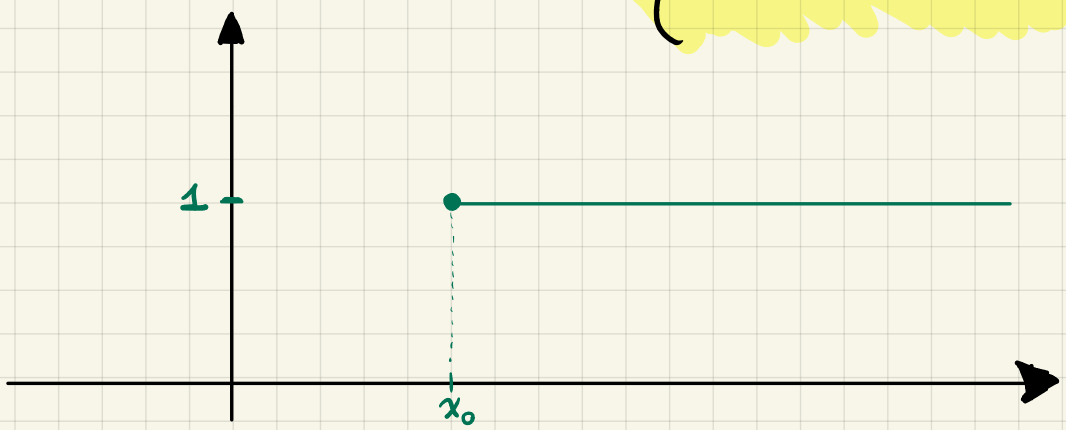
From now on, we will say indifferently

F is discrete, ν is discrete, X is discrete

- The form of distribution function

If $\nu(x_0) = 1$, then $B = \{x_0\}$, so

$$F(x) = \nu(-\infty; x] = \begin{cases} 0 & \text{if } x < x_0 \\ 1 & \text{if } x \geq x_0 \end{cases}$$



Such a ν is said to be degenerate.

Similarly a r.v. X is degenerate for some $x_0 \in \mathbb{R}$

$$P(X = x_0) = 1$$

So, more in general, if $\nu\{a, b, c\} = 1$ then $B = \{a, b, c\}$

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \nu(a) & \text{if } a \leq x < b \\ \nu(a) + \nu(b) & \text{if } b \leq x < c \\ 1 & \text{if } x \leq c \end{cases}$$

- Singular distribution function

F is a singolar distribution function continuos when:

$$\begin{cases} F \text{ is continuous} \\ F' = 0 \\ m \text{ almost everywhere} \end{cases}$$

$\Rightarrow$ m a.e.

$$M : \left\{ x \in \mathbb{R} : F \text{ is not differentiable at } x, \text{ or it exist but it's not } F'(x) \neq 0 \right\} = 0$$

$m \to$ Lebesgue measure

Set of points

$\Rightarrow$ if F is discret, F is not continuous but $F' = 0$ m a.e.

- Absolute continuous function

$\exists$ a function $f: \mathbb{R} \to \mathbb{R}^+$, $f \geq 0$ and f is integrable such that:

$$F(x) = \int_{-\infty}^{x} \underbrace{f(t)}_{\text{the density function}} \, dt \quad \forall x \in \mathbb{R}$$

Notes:

- it's a Lebesgue integral (not Reimann)
- $\nu\{x\} = 0 \quad \forall x \in \mathbb{R}$
- $F' = f$ but only m a.e.

$$m\left\{x \in \mathbb{R} : F'(x) \neq f(x)\right\} = 0$$

- Consequence

- If f_1 and f_2 are both derivates of F, then

$$f_1 = F = f_2 \quad m \text{ a.e.}$$

→ So the density is m a.e. unique

$$\Downarrow$$

Example:

Let ϕ be the density of a $N(0,1)$

$$\phi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \quad ;$$

define

set of rational numbers

$$f(x) = \begin{cases} \phi(x) & \text{if } x \neq Q \\ \log(2 + sen(x)) & \text{if } x \in Q \end{cases}$$

Since Q is countable, $m(Q) = 0$.

Hence, $m\{x : f(x) \neq \phi(x)\} = m(Q) = 0$

that is $f = \phi$ m a.e.

is still a density of $N(0;1)$

Remember:

- A countable set has always a Lebesgue measure $= 0$

- We can say that:

$$\nu\{x\} \leq \nu(x-\varepsilon, x]$$

$$\nu\{x\} \leq F(x) - F(x-\varepsilon]$$

$$\nu\{x\} \leq \int_{x-\varepsilon}^{x} f(t)\, dt$$

$$\nu\{x\} \leq \lim_{\varepsilon \to 0} \int_{x-\varepsilon}^{x} f(t)\, dt = 0$$

$$\nu\{x\} = 0$$

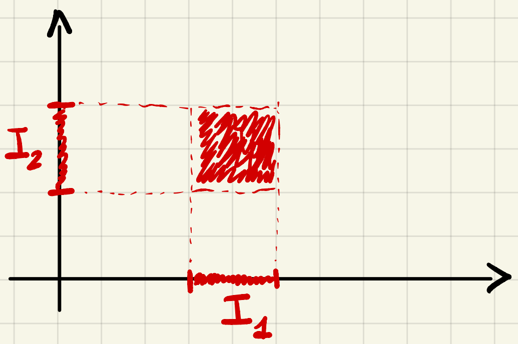
- Absolutely continuous $\xrightarrow{\text{implies}}$ continuous

A simple continuous that is not an absolute continuous is, for example, the singular r.v.

$$\begin{cases} F \text{ is continuous} \\ F' = 0 \quad m \text{ a.e.} \end{cases}$$

- The Borel δ-field on R^n, denoted as $\beta(R^n)$, is the smallest δ-field on R^n including all the form $I_1 \times I_2 \times \dots \times I_n$ where I_j is an interval for each j.

- If $n = 2$



The Lebesgue measure on $\mathbb{R}^n$ is the measure m_n on $\beta(R^n)$ such that:

$$m_n(I_1 \times \dots \times I_n) = m(I_1) \cdot \dots \cdot m(I_n)$$

Intuition: if $A \in \beta(R^n)$ is such that the area of A makes sense, then

$$m_n(A) = \underset{\text{(volume)}}{\text{area}}(A)$$

• Let $\underline{X} = (x_1, \dots, x_n)$ be a n-variate random variable;

let $\nu(B) = P(X \in B), \forall B \in \beta(R^n)$ the distribution of X.

$\Downarrow$

Such a ν is a probability measure in $\beta(R^n)$.

- Distribution function of X (n-variate r.v.)

$$F(X_1, \dots, X_n) = \nu\left[(-\infty, x_1], \dots, (-\infty, x_n]\right]$$

$$= P(X_1 \leq x, \dots, X_n \leq x), \quad \forall (X_1, \dots, X_n) \in \mathbb{R}^n$$

$$F : R^n \to R$$

→ X is discrete

$\exists\ B \subset \mathbb{R}^n$, with B finite or countable,

$P(X \in \mathbb{B}) = 1$

→ X is absolutely continuous

$\exists\ f : R^n \to R$, $f \geq 0$, finite and integrable, such that:

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n)\, dt_1 \dots dt_n$$

theorem

if and only if:

$$P(X \in A) = 0 \quad \forall A \in \mathcal{B}(\mathbb{R}^n)$$

such that

$$m_n(A) = 0$$

What are $\perp$ random variables?

Given any collection ξ of real random variables, we say that ξ is independent, if $P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$

$\forall A_1, \dots, A_n \in \underset{\text{Borel-set}}{\underset{\downarrow}{\mathcal{B}}}$, $\forall X_1, \dots, X_n \in \xi$ all distinct

- Important case:

ξ is independent and identically distributed (i.i.d)

$$\begin{cases} 1) & \xi \text{ is independent.} \\ 2) & \text{all the elements of } \xi \text{ have the same distribution.} \end{cases}$$

- Again, let $\mathcal{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$

  if $\mathcal{X}$ is absolutely continuous, then all marginals of $\mathcal{X}$ are still continuous.

  → However, it may be that x_i are absolutely continuous $\forall i$, but $\mathcal{X}$ is not absolutely continuous.

- Finally, if $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are $\perp$, then $\mathcal{X}$ is absolutely continuous if and only if $\mathcal{X}_i$ are absolutely continuous $\forall i$.

- What about the density?

  - If f is the density of $\mathcal{X}$, the density of a marginal can be obtained by integrating out all variables not involved in the marginal.

- For instance, the density of $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ with $k < n$ is

$$f(x_1, \ldots, x_k) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_k, \overbrace{x_{k+1}, x_n}^{n-k}) \, dx_1 \ldots dx_n$$

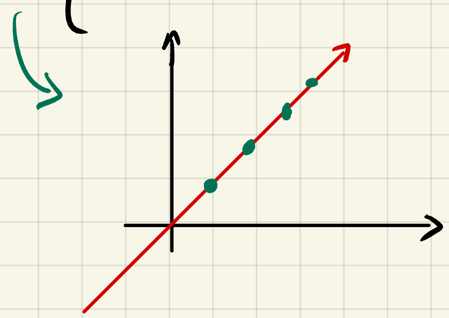
Let suppose $n=2$ and $X_2 = X_1 \sim N(0,1)$

the vector $\underline{X} = (X_1, X_2)$ is not absolutely continuous.

Why not?

In fact, we can take

$$A = \{(x,y) : y = x\}$$



$$P[(X_1, X_2) \in A] = 1$$

but $\underbrace{m_2(A) = 0}$

- If $X_1, \dots, X_n$ are $\perp$ and X_i is absolutely continuous $\forall i$, then X is absolutely continuous and the density of X is:

$$f(x_1, \dots, x_n) = f(x_1) \cdot \ldots \cdot f(x_n)$$

- Multivariate normal distributions

- Let $X = (X_1, \ldots, X_n)$ and $X \sim N_{(n)}(\vec{\mu}, \vec{\Sigma})$, where $\mu \in \mathbb{R}^n$ and Σ is a simmetric and non-negative matrix $n \times n$, if

complex number

$$E\left(e^{i \cdot t' \cdot x}\right) = e^{i \cdot t' \cdot \mu - \frac{1}{2} \cdot t' \cdot \Sigma \cdot t} \quad \forall t \in \mathbb{R}^n$$

t' → transpose

where

$$t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

are column vectors.

Note:

Σ is only positive (≥ 0) definite and thus it may be that $\det(\Sigma) = 0$ (singular)

- If $\det(\Sigma) > 0$, it can be shown that X is absolutely continuous with density:

$$f(X) = (2\pi)^{-\frac{n}{2}} \cdot (\det(\Sigma))^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}\cdot(x-\mu)^t \cdot \Sigma^{-1} \cdot (x-\mu)}$$

$$\forall x \in \mathbb{R}^n$$

[- Note that if $n = 1$ and $\det(\Sigma) > 0$ we obtain the density of a univariate normal distribution]

- When we write

$$X \sim N_{(n)}(\mu, \Sigma)$$

$$E\left(e^{i \cdot t' \cdot x}\right) = e^{i \cdot t' \cdot \mu - \frac{1}{2} \cdot t' \cdot \Sigma \cdot t} \quad \forall\, t \in \mathbb{R}^n$$

$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}_{n \times 1}$ $\quad (\quad)_{1 \times n} \quad (\quad) \quad [\quad]_{n \times n}$

Called caractheristic function of X:

→ We distinguish 2 cases:

1) $\det(\Sigma) = 0 \rightarrow X$ degenerate r.v.

2) $\det(\Sigma) > 0 \rightarrow X$ absolutely continuous

$$\mu = E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

$$\Sigma = \begin{bmatrix} Var(X_1), & cov(X_1, X_2) & & \\ & Var(X_2) & & \\ & & \ddots & \\ & & & Var(X_n) \end{bmatrix}$$

→ If $n = 1$, Σ reduces to a 0 scalar and so $\det(\Sigma) = 0$: otherwise, if the scalar k is $\neq 0$, the $\det(k) \neq 0$.

So, in this case, we have that

$$P(X = \mu) = 1$$

• Linear combination

If $X \sim N_{(n)}(\mu, \Sigma)$ and $Y = a + b \cdot X$, where $a \in \mathbb{R}^n$ and B is a $m \times m$ matrix:

$$y \sim N_n(a + B \cdot \mu, B \Sigma B')$$

proof:

Let us evaluate

$$E(e^{i \cdot t' \cdot y}) = E\left[e^{i \cdot t'(a + B \cdot X)}\right]$$

$$= E\left[e^{i \cdot t' \cdot a} \cdot e^{i \cdot t' \cdot BX}\right]$$

$$= e^{i \cdot t' \cdot a} \cdot E\left[e^{i \, t' \cdot B \cdot X}\right]$$

$$= e^{i \cdot t' \cdot a} \cdot E\left[e^{i \cdot (B' \cdot t)' \cdot X}\right]$$

$$= e^{i \cdot t' \cdot a} \cdot e^{i \cdot (B' \cdot t)' \cdot \mu - \frac{1}{2} \cdot (B' \cdot t)' \cdot \Sigma \cdot (B' \cdot t)}$$

$$= e^{i t'(a + B \cdot \mu) - \frac{1}{2} t'(B \Sigma B') t}$$

• Moments

- Let X be a real r.v., our aim is to define $E(X)$. However, it may be that $E(X)$ does not exist. Precisely, we first define $E(|X|)$:

$$E(|X|) = \int_0^{+\infty} P(|X| > t)\, dt$$

there may be two results:

1) $= +\infty$ and in this case we say that $E(X)$ does not exist

2) $< +\infty$ and in this case can be evaluated according to the formula:

$$E(X) = \begin{cases} \displaystyle\sum_{x \in B} x \cdot P(X = x) & \rightarrow \text{if } X \text{ is discrete, where } B \text{ is finite or countable and } P(X \in B) = 1 \\ \displaystyle\int_{-\infty}^{+\infty} x \cdot f(x)\, dx & \rightarrow \text{if } X \text{ is absolutely continuous where } f \text{ is the density} \\ \displaystyle\int_0^{+\infty} P(X > t)\, dt & \rightarrow \text{if } X \geq 0 \end{cases}$$

Important: if X does not belongs to any category of the previous, there are formula for $E(X)$: but, they are beyond the scopes of this course

Examples:

- discrete case:

• formula $$E(X) = \sum_{x \in B} x \cdot P(X = x)$$

Suppose to take a Poisson

$$X \sim \text{Poisson}(\lambda) \quad , \quad B = \{0, 1, 2, \dots\}$$

$$E[X] = \sum_{j=0}^{+\infty} j \cdot P(X = j)$$

$$= \sum_{j=0}^{+\infty} j \cdot \frac{e^{-\lambda} \cdot \lambda^j}{j!}$$

- r.v. with no-mean

• If X is absolutely continuous, it can be shown that
$$E(|X|) = \int_0^{+\infty} P(|X| > t)\, dt = \int_0^{+\infty} |t| \cdot f(t)\, dt$$

Say that X is Cauchy if X is absolutely continuous with density:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \forall x \in \mathbb{R}$$

Thus:

$$E(|x|) = \int_{-\infty}^{\infty} |x| \cdot f(x)\, dx = \int_{-\infty}^{+\infty} |x| \cdot \frac{1}{\pi} \cdot \frac{1}{1+x^2}\, dx$$

$$= \frac{1}{\pi} \cdot \int_{-\infty}^{+\infty} \frac{|x|}{1+x^2}\, dx = \frac{2}{\pi} \cdot \int_{0}^{+\infty} \frac{x}{1+x^2}\, dx$$

cause the function it's even

$$= +\infty$$

So it has no mean

- **Note:** from now on, we will assume that all r.v. has the mean.

→ The Following formulas in the next page are very useful since they not depends on however the r.v. is discrete, absolutely continuous or whatever.

- Properties

1) Linearity: $E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$

2) Positivity: if $X \geq 0$, $E[X] \geq 0$

3) if $P(X = a) = 1$, then $E[X] = a$

4) if $P(A) > 0$ and $X > 0$ on A,

then $E(X \cdot 1_A) > 0$

$\Downarrow$

Here, 1_A is called the indicator function of the event A and is defined as:

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

and

$$X \cdot 1_A = \begin{cases} X & \text{on } A \\ 0 & \text{on } A^c \end{cases}$$

Exercise: Let's evaluate the expected value of 1_A

$$E[1_A] = \sum x_i \cdot P(X_i) = 0 \cdot P(X_i) + 1 \cdot P(X_i) = P(X_i)$$
$$= P(1_A = 1) = P(A)$$

Suppose we want to forecast an event and a is our guess: then,

$X - a$, can be regarded as the "error".

If $E(X^2) < +\infty$,

$$E[(X - E(X))^2] = \min_{a \in \mathbb{R}} E[\underbrace{(X - a)^2}_{\text{"error"}}]$$

f attains its minimum at $a = E(X)$

- X is a real random variable, in order to say if $E(X)$ exist we evaluate:

$$E(|X|) = \int_0^{+\infty} P(|X| > t)\, dt = \begin{cases} +\infty : \text{non exist} \\ < \infty : \text{exist} \end{cases}$$

So, $f(a) = E\left[(X-a)^2\right] = E\left\{\left[(X - E(X)) + (E(X) - a)\right]^2\right\}$

$$= E\left[(X - E(X))^2\right] + \left[E(X) - a\right]^2$$

has the minimum value when $E(X) = a$

But also we can take

$$g(a) = E\left[|X - a|\right] = E\left\{|\text{error}|\right\}$$

attains its median $a = \text{median}(X)$

- If $E(|X|^s) < +\infty$,

  we can define the s-moment of order s.

  $$E(X^s).$$

  In particular, if $E(X^2) < +\infty$,

  we can define the variance as follows:

  $$Var(X) = E\left[(X - E(X))^2\right] = E(X^2) - E(X)^2$$

- Properties of the variance

1. Linear combination

$$Var(a + b \cdot X) = E\left[\left(\cancel{a} + b \cdot X - (\cancel{a} + bE(X))\right)^2\right]$$

$$= b^2 \cdot Var(X)$$

2. Degenerate

$$X \text{ is degenerate} \longleftrightarrow Var(X) = 0$$

- Covariance

- If we have two r.v. X and Y, and if $E[X \cdot Y] < +\infty$, $E[|X|] < +\infty$, $E[|Y|] < +\infty$ we can define the covariance as follows:

$$Cov(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

$$= E[(X \cdot Y)] - E(X) \cdot E(Y)$$

- Correlation coefficient

- In addition, if $E(X^2) < +\infty$, $E(Y^2) < +\infty$, $Var(X) > 0$ and $Var(Y) > 0$ we can also define

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

- If $E[(X)^\alpha] < +\infty$ for some $\alpha > 0$, then $E[(X)^\beta] < +\infty \ \forall \beta \in (0, \alpha]$.

Example:

If $E(X^5) < +\infty$, we are sure that $E(X^2) < +\infty$.

→ Given the definition of independence of two r.v. we can say:

- if $X \perp Y \longrightarrow Cov(X, Y) = 0$

$\not\longleftarrow$

$\downarrow$

In Fact, $E(X \cdot Y) = E(X) \cdot E(Y)$ so that

$$Cov(X, Y) = E(X) \cdot E(Y) - E(X) \cdot E(Y) = 0$$

- But why a null covariance does not implies independence?

→ Let's do an example:

$$X \sim N(0, 1) \quad \text{and} \quad Y = X^2$$

So X and Y are clearly dependent

$$\mathrm{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

but $E(X) = 0$ and $Y = X^2$

$$= E(X \cdot Y) - 0 \cdot E(Y)$$

$$= E[X \cdot X^2]$$

$$= E[X^3]$$ # X is absolutely continuous

$$= \int_{-\infty}^{+\infty} x^3 \cdot f(x)\, dx$$

$\Downarrow$

this integral is odd so

$$= \underline{0}$$

$\rightarrow$ Now, let $Y = X^2$ again and evaluate

$$P(|X| \leq 1 \wedge Y > 1)$$

to complete

- Variance of X_j

- $$Var\left(\sum_{j=1}^{n} a_j \cdot X_j\right) = \sum_{j=1}^{n} a_i^2 \cdot Var(X_i) + \sum_{i \neq j} a_i \cdot a_j \cdot Cov(X_i, X_j)$$

- in the special case where $X_i \perp X_j$

$$Var\left(\sum_{j=1}^{n} a_j \cdot X_j\right) = \sum_{j=1}^{n} a_j^2 \cdot Var(X_j)$$

- For example

$$V(X+Y) = V(X) + V(y) + 2\,Cov(X,y)$$

$$V(X-Y) = V(X) + V(y) - 2\,Cov(X,y)$$

- **Note:** if $X \perp Y$

$$Var(X+Y) = Var(X-Y)$$

- Corr. coefficient with standardized r.v.

- Given a real r.v. X, take the standardization

$$Z = \frac{X - E(X)}{\sqrt{Var(X)}}$$

$$Z \sim (\mu = 0 \ ; \ \sigma = 1)$$

- Now, $\rho(X, Y) = \dfrac{Cov(X, Y)}{\sqrt{\sigma(X) \cdot \sigma(Y)}}$

$$= Cov\left(\frac{X - E(X)}{\sqrt{Var(X)}} \ ; \ \frac{Y - E(Y)}{\sqrt{Var(X)}} \right)$$

Chebicev inequality

$$P\left[\,|X| \geq c\,\right] \leq \frac{E(|X|^{\alpha})}{c^{\alpha}} \qquad \forall\ c, \alpha > 0$$

Important case:

if $\alpha = 2$ and $X = y - E(y)$

$$P\left[\,|y - E(y)| \geq c\,\right] \leq \frac{E\left[\,|(y - E(y)|^{2}\,\right]}{c^{2}}$$

$$\Downarrow$$

$$P\left[\,|y - E(y)| \geq c\,\right] \leq \frac{Var(y)}{c^{2}}$$

And we can say also:

$$\left[\,c^{\alpha} \cdot P\left[|x| \geq c\right]\,\right] \leq E(|X|^{\alpha})$$

$$\Downarrow$$

$$E(c^{\alpha})$$

We can prove this:

- proof

Can you see why?

$$E(|X^{\alpha}|) \geq E\left\{|X^{\alpha}| \cdot 1(|X| \geq c)\right\}$$

So,

$$|X^\alpha| \geq C^\alpha \quad \text{on the set } \{|x| \geq c\}$$

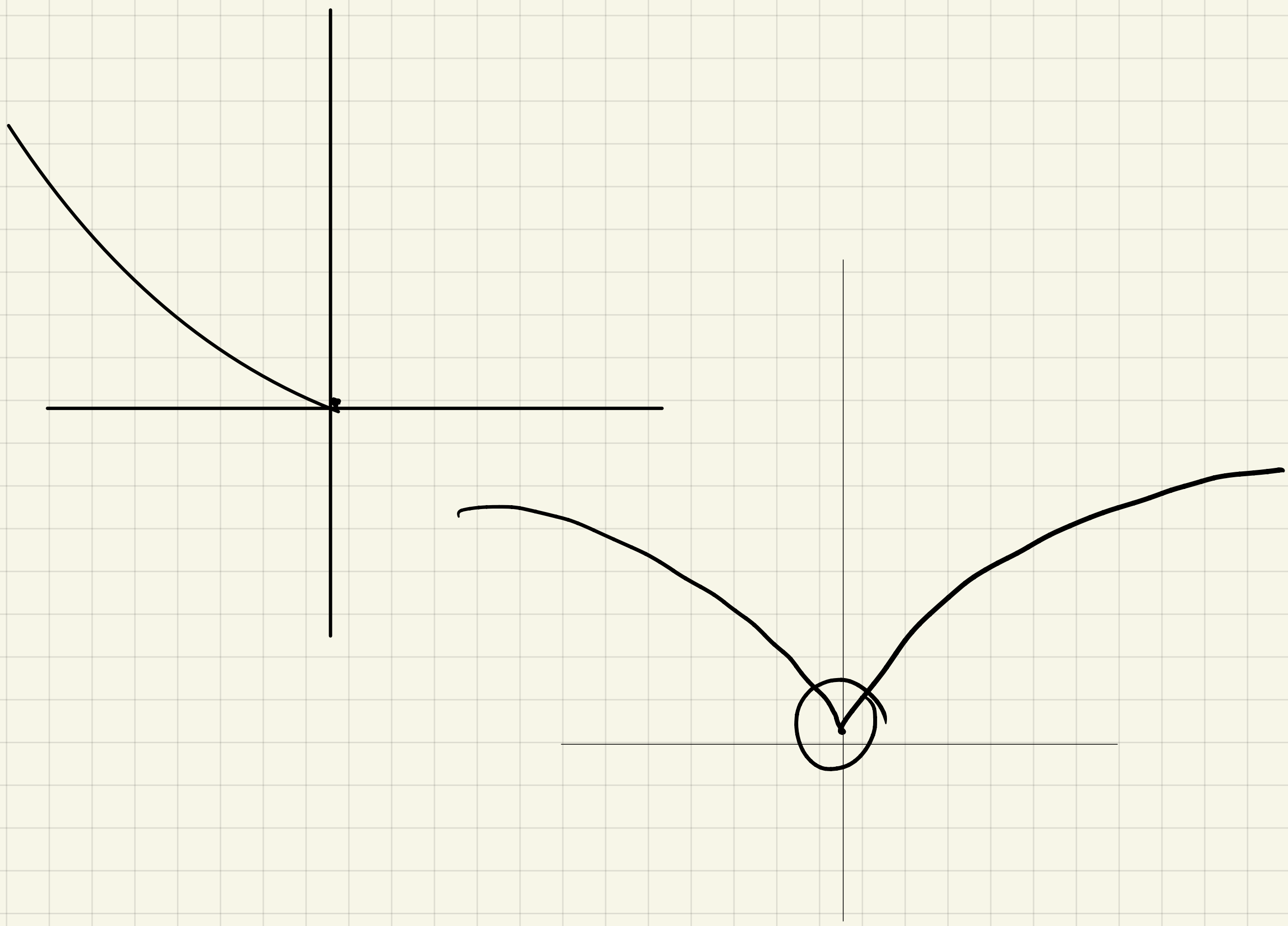
Then,

$$E\left[|X^\alpha|\right] \geq E\left\{c^\alpha \cdot \mathbb{1}(|x| \geq c)\right\}$$

Since C^α it's a constant

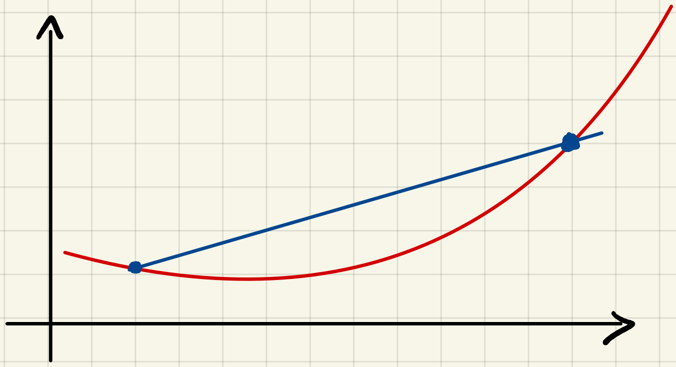
$$E\left[|X^\alpha|\right] \geq c^\alpha \cdot E\left[\mathbb{1}(|x| \geq c)\right]$$

$$\geq c^\alpha \cdot P(|X| \geq c)$$



Brief note on the convex function

def: a real f. is called convex if the line segment between any two points on the graph of the f. lies above the graph between two points.



On mathematical terms:

If the function it's twice differentiable, then we can say that

$$f \text{ is convex} \longleftrightarrow f'' > 0$$

Jensen's inequality

If $f : I \to \mathbb{R}$ is convex, the $P(X \in I) = 1$ and $E[|X|] < +\infty$, then

$$E[f(x)] \geq f[E(X)]$$

We can also say that the inequality is strict such that:

$$E[f(x)] > f[E(X)]$$

when those assumptions are verified:

- X non degenerate
- f strictly convex

- Example

$n(t, X)$

$E(n(t, X))$

to complete

Fourier transform

Laplace transform

$E(e^{tX})$

Characteristic function

Let $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ be a n-variate r.v.

By definition, the characteristic function of X is $\phi(t) : \mathbb{R}^n \longrightarrow \mathbb{C}$

$$\phi(t) = E\left[e^{i \cdot t' \cdot x}\right]$$

$$= E\left[e^{i \cdot \sum_{j=1}^{n} t_j \cdot x_j}\right] \quad \forall\, t \in \mathbb{R}^n$$

Here i is the complex number such that $i^2 = -1$.

We also recall that

$$e^{i \cdot \alpha} = \sqrt{\cos(\alpha)^2 + i^2 \cdot \sin(\alpha)^2} \qquad \forall\, \alpha \in \mathbb{R}$$

Important fact

We can say that $X \sim Y$ if and only if $\phi_{(X)} = \phi_{(Y)}$

$\Rightarrow$ From now on, for semplicity, $n = 1$ so that X is a real random variable.

Properties

1) If $X \perp Y$, then

$$\phi_{X+Y} = E\left[e^{it(X+Y)}\right] = E\left[e^{itX} \cdot e^{itY}\right]$$

$$= E\left(e^{itX}\right) \cdot E\left(e^{itY}\right)$$

$$= \phi_X(t) \cdot \phi_Y(t) \quad \forall t \in \mathbb{R}$$

$e^{x^2} = 2x \cdot e^{x^2}$

So, $\phi_{X+Y} = \phi_X \cdot \phi_Y$

$E(e^{itX})$

$E[iX \cdot e^{itX}]$

2) If $E\left(|X|^J\right) < +\infty$, then $\phi_X \in C^J$

$\Downarrow$

the J-th derivative of ϕ exist and is continuous

So, $\phi^{(k)}(t) = i^k \cdot E\left(X^k \cdot e^{it \cdot X}\right)$ for $k = 1, \dots, J$

Hence, $\phi^{(k)}(0) = i^k \cdot E(X^k)$

Conversely, if ϕ has the J-th derivative at 0 and J is even, then

$$E\left(|X|^J\right) < +\infty$$

3) Let X and $X_1, \dots, X_n$ be real r.v.

Then,

$$X_n \xrightarrow{d} X$$

$$\Updownarrow$$

$$\phi_X(t) = \lim_{n \to +\infty} \phi_{X_n}(t) \quad \forall\, t \in \mathbb{R}$$

4) Inversion theorem

If $P(X = a) = P(X = b) = 0$, then

$$P(a < X \leq b) = \frac{1}{2\pi i} \cdot \lim_{c \to +\infty} \int_{-\infty}^{+\infty} \frac{e^{-ita} - e^{-itb}}{t} \cdot \phi(t)\, dt$$

$$\Downarrow$$

$$= F(b) - F(a)$$

- We have ϕ and we wonder about F

→ As a consequence of the previous theorem we obtain this fact:

if $\int_{-\infty}^{\infty} |\phi(t)|\, dt < +\infty$, then X is absolutely continuous with density:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \cdot \phi(t)\, dt$$

• If $X \sim N(0,1)$, then $\phi(t) = e^{-\frac{t^2}{2}}$

$E[e^{it'X}]$

Knowing this fact, we can derive the $\phi_y(t)$ of a normal distribution:

$\frac{y-\mu}{\sigma} = Z$ Standard Gaussian

$$y \sim N(\mu, \sigma^2) \text{ and } \underline{y = \mu + \sigma X}$$

so that

$$\phi_y(t) = E\left(e^{ity}\right) = E\left(e^{it(\mu + \sigma X)}\right)$$

$$= E\left(e^{it\mu} \cdot e^{it\sigma X}\right)$$

$$= e^{it\mu} \cdot E\left(e^{it\sigma X}\right)$$

$$= e^{it\mu} \cdot \phi_X(t \cdot \sigma)$$

$$= e^{it\mu} \cdot e^{-\frac{1}{2}(t \cdot \sigma)^2}$$

$$= e^{it\mu - \frac{1}{2}(t\sigma)^2}$$

To sum up,

$$y \sim N(\mu, \sigma^2) \implies \phi_y(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}$$

→ The characteristic function always exist and satisfies

$$|\phi(t)| \leq 1 \quad \forall t \in \mathbb{R}$$

Suppose $Z = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim N_n(\mu, \Sigma)$

$$\phi_Z(t) = E\left(e^{i\, \underbrace{t' \cdot Z}}\right) \quad \forall\, t \in \mathbb{R}^n$$

$$\sim N\left(\overset{a\mu}{t'\mu}\, ,\, \overset{B \Sigma B'}{\underline{t' \Sigma t}}\right)$$

$$= \phi(1) = e^{i \cdot 1 \cdot t'\mu - \frac{1}{2} 1^2 \cdot t' \Sigma t}$$

$$= e^{i t'\mu - \frac{1}{2} t' \Sigma t}$$

Moment generating function

If X is any real r.v., the moment generating function is

$$\Psi(t) = E(e^{tX}) \quad \forall\, t \in \mathbb{R}$$

- Shortcomings:

If $t = 0$, $\Psi(t) = 1$ but for $t \neq 0$ it may be that

$$\Psi(t) = +\infty$$

[for example in the Cauchy distribution]

For instance, if

$$\Psi(t) < +\infty \qquad \forall\, t \in (-\varepsilon, \varepsilon)$$

where $\varepsilon > 0$, then

$$E[|X|^n] < +\infty \quad \forall\, n.$$

Conditional distributions

Let (X, Y) be a bivariate random variable.

Our goal is to introduce the notion of conditional distribution of Y given X

Definition:

The conditional distribution of Y given X is a map

$$P[(X, Y) \in C \mid X = x]$$

defined for all $x \in \mathbb{R}$ and $C \in \mathcal{B}(\mathbb{R}^2)$ that satisfies 3 properties:

a) For fixed $x \in \mathbb{R}$,

$$C \mapsto P[(X, Y) \in C \mid X = x]$$

is a probability measure on $\mathcal{B}(\mathbb{R}^2)$

b) $P[(X, Y) \in C] = E_{X}\left\{P[(X, Y) \in C \mid X = x]\right\}$

$\forall\, C \in \mathcal{B}(\mathbb{R}^2)$

Expectation is taken with respect to x

c) $P[(x, y) \in C \mid X = x] = P[(x, Y) \in C \mid X = x]$

Intuition behind:

Given true that $X = x$, what is the joint distribution of X and Y?

Important remark:

Even if we write

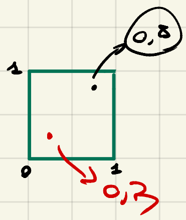
$$P[(X, Y) \in C \mid X = x]$$

it is not necessary true that

$$P[(X, Y) \in C \mid X = x] = \frac{P[X = x \wedge (X, Y) \in C]}{P(X = x)}$$

In fact, it may be that

$$P(X = x) = 0$$



$$Y \sim Bin(50, 0{,}8)$$

$$P(Y = 3) = \binom{50}{3} \cdot p^3 \cdot (1-p)^{47}$$

$Y \mid X = x$

$X \mid Y = y$

$[-2;2]$

$$P(X \in A) = \mathbb{1}_{x \text{ in } A}$$

$$P\left(y \in \{0, 1, \dots n\}\right)$$

$$= E_x \{$$

Order statistics

If $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ is any n-variate r.v., the corresponding order statistics.

- $Y = \begin{pmatrix} X_{(1)} \\ \vdots \\ X_{(n)} \end{pmatrix}$ where $X_{(1)} \leq \ldots \leq X_{(n)}$

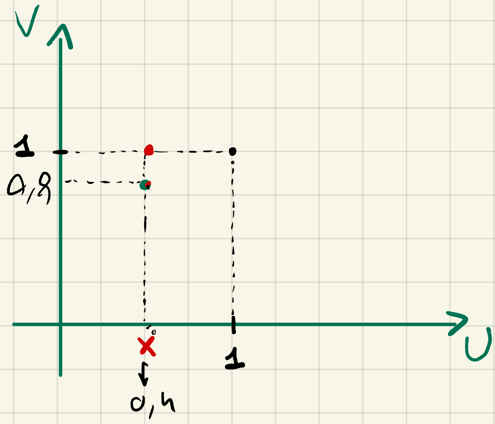
General problem: given distribution X, find the distribution Y

- If $X_1, \ldots, X_n$ are iid and X_1 is absolutely continuous with density g, then Y is still absolutely continuous with density:

$$f(x_1, \ldots, x_n) = \begin{cases} n! \prod_{i=1}^{n} g(x_i) & \text{if } x_1 < x_2 < \ldots < x_n \\ 0 & \text{otherwise} \end{cases}$$

Exercise

Let U, V be i.i.d. with U uniform $(0,1)$.
Find the conditional distribution of $\max(U,V)$ given $\min(U,V)$.



$U = 0,7$ | $U = 0,3$
$V = 0,4$ | $V = 0,8$

$x = \min(U,V)$
$y = \max(U,V)$
} Order statistics of (U,V)

$\Downarrow$

So, we can recall the previous theorem and (x,y) is absolutely continuous with density:

$$f(x,y) = \begin{cases} 2!\, g(x)\cdot g(y) & \text{if } x<y \\ 0 & \text{otherwise} \end{cases}$$

We also know that since $U \sim \text{Uniform}(0,1)$

$$g(x) = \begin{cases} 1 & \text{if } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$f(x,y) = \begin{cases} 2! & \text{if } 0<x<y<1 \\ 0 & \text{otherwise} \end{cases}$$

Since we want find $h(y|x) = \dfrac{f(x,y)}{f(x)}$,

we need the marginal density of X.

$$f_{(1)}(x) = \int_x^1 2!\, dy = 2(1-x) \quad \forall\, x \in (0,1)$$

Hence,

$$h(y|x) = \frac{f(x,y)}{f(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x}$$

$\forall\ 0<x<y<1$

Exercise

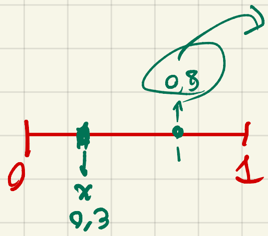
Suppose $X \sim \text{Uniform}[0,1)$.

$$Y \mid X = x \sim \text{Bin}(n, x) \quad \forall x \in (0,1)$$

- Calculate $X \mid Y = y$

$$P\left(Y \in \{0, 1, \ldots, n\}\right) =$$

$$= E_X\left\{P\left(Y \in \{0,1,\ldots,n\} \mid X = x\right)\right\}$$



$X = 0{,}3$

$p = 0{,}3 = 0{,}4\ ,\ 0{,}5$

$Y \in (0, 1, 2, 3,$

Convergence of random variables

Problem:

$X_1, X_2, \dots$ are real random variables

We want to investigate whether

$$X_n \longrightarrow X \quad \text{in some sense as } n \to +\infty$$

Almost sure convergence

$$X_n \xrightarrow{a.s.} X$$

- $P\{\omega \in \Omega : X_n(\omega) \not\longrightarrow X(\omega)\} = 0$
- $P\{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\} = 1$

Given $\omega \in \Omega$, $X(\omega)$ is a real number

and $(X_{(1)}(\omega), X_{(2)}(\omega), \dots)$ is a sequence of real numbers

LP convergence

Given $P > 0$,

$$X_n \xrightarrow{L_P} X \iff \begin{cases} E|X_n|^p < +\infty \quad \forall n \\ E|X|^p < +\infty \\ \lim_{n \to +\infty} E[|X_n - X|^p] = 0 \end{cases}$$

Convergence in probability

$$X_n \xrightarrow{P} X \iff \lim_{n \to +\infty} P[|X_n - X| > \varepsilon] = 0 \quad \forall \varepsilon > 0$$

Convergence in distribution

$$X_n \xrightarrow{d} X \iff F(x) = \lim_{n\to+\infty} F_n(x)$$

$\forall\, x \in \mathbb{R}$ such that F is a continuous function

Why F has to be continuous?

We can argue as follows:

- Suppose X_n and X are all degenerate r.v. and

$$X_n = \frac{1}{n} \quad \text{and} \quad X = 0$$

- $$F(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

- $$F_n(x) = \begin{cases} 1 & \text{if } x \geq 1/n \\ 0 & \text{if } x < 1/n \end{cases}$$

Hence, we have $F_n(x) \longrightarrow F(x) \quad \forall\, x \neq 0$

However, in the only point where F is not continuous, we have $F(0) = 1$ while $F_n(0) = 0 \quad \forall\, n$

is that $F_n(0) \not\longrightarrow F(0)$

to sum-up

$$\begin{tikzcd}
\text{Convergence in } L_p \arrow[rr, "/" marking] \arrow[dr, shift left] & & \text{Almost sure convergence} \arrow[dl, shift right] \\
& \text{Convergence in probability} \arrow[ul, "/" marking, shift left] \arrow[ur, "/" marking, shift right] \arrow[d, shift right] & \\
& \text{Convergence in distribution} \arrow[u, "/" marking, shift right] &
\end{tikzcd}$$

· **Remarks**

1) In general, convergence in distribution does not implies convergence in probability. However, in the special case where X is degenerate, we have

$$X_n \xrightarrow{P} X \iff X_n \xrightarrow{d} X$$

2) $X_n \xrightarrow{P} X$ does not implies $X_n \xrightarrow{a.s.} X$

However, if $X_n \xrightarrow{P} X$, there is a suitable subsequence $n_1 < n_2 < n_3 \dots$ such that

$$X_{n_J} \xrightarrow{a.s.} X \quad \text{as } J \to +\infty$$

3) If $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$ then

$$P(X = Y) = 1 \quad [a.s.]$$

As a consequence, suppose we know that $X_n \xrightarrow{P} X$, and we are interested in whether X_n converges a.s. or L_p.

Then, if

$$X_n \xrightarrow{a.s.} Y$$

$$X_n \xrightarrow{P} Y$$

$$X = Y \to a.s.$$

Let us prove a part of the "picture".

1) L_p convergence $\Rightarrow$ convergence in probability

to be completed

2) a.s. convergence $\not\Rightarrow$ Convergence in L_p

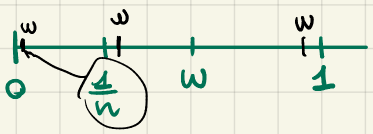
Let's take an example where a.s. conv. is true but not the L_p conv.

For example:

$$(\Omega, \mathcal{A}, P) = \left([0,1), \mathcal{B}[0,1), m\right)$$ where m is the Lebesgue measure

Let $X(\omega) = 0$, $X_n(\omega) = n^{1/p} \cdot \mathbb{1}_{[0,1/n)}(\omega)$

If $\omega \in (0,1)$, then $\frac{1}{n} < \omega$ for each sufficently large n



and then $X_n(\omega) = 0$ for each large n.

Instead, if $\omega = 0$

$$X_n(0) = n^{\frac{1}{p}} \cdot \mathbb{1}_{[0,1/n)}(0) = n^{1/p} \cdot 1 \longrightarrow +\infty$$

Hence,

$$P\left\{\omega : X_n(\omega) \not\longrightarrow X(\omega)\right\}$$

$$= P\left\{\omega : X_n(\omega) \not\longrightarrow 0\right\}$$

$$= P\{0\} = m\{0\} = 0. \qquad X_n \xrightarrow{a.s.} X$$

→ However,

$$E\left\{|X_n - X|^p\right\} = E\left\{|X_n{}^p|\right\}$$

$$= E\left\{\left(n^{\cancel{1/p}} \cdot \mathbb{1}_{[0,\frac{1}{n})}\right)^{\cancel{p}}\right\}$$

$$= n \cdot E\left\{\mathbb{1}^p_{[0,1/n)}\right\} = n \cdot P\left\{[0\,;\,1/n)\right\}$$

since it's a Lebesgue measure

$$= n \cdot \left[\frac{1}{n} - 0\right] = 1$$

Therefore, since

$$E\left\{|X_n - X|^p\right\} = 1, \text{ then } X_n \xrightarrow{L^p} X$$

Law of large numbers

Given a sequence $X_1, X_2, \dots$ of iid of real random variables: we say that (X_n) satisfies the law of large numbers if

$$\overline{X}_n \longrightarrow V \quad \text{for some r.v. } V$$

where $\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

- if $\overline{X}_n \xrightarrow{a.s.} V$, we speak of strong law of large numbers
- if $\overline{X}_n \xrightarrow{P} V$, we speak of weak law of large numbers

The r.v. V could be any, but an important case is when V is $E(X_n) = E(X_1)$ $\forall n$, and $V = E(X_1)$.

Heuristic interpretation

X_i is the value of some random quantities at time i.

$\overline{X}_n$ is the sample mean of the first n observations

So, it's natural to investigate whether $\overline{X}_n$ converges to a limit.

Most important S.L.L.N. - theorem

If $X_1, X_2, \dots, X_n$ is iid, then

$$\overline{X}_n \xrightarrow{a.s.} V \iff E(X_1) < +\infty$$

for some V.

Moreover, if $E(X_1) < +\infty$, then $V = E(X_1)$

namely

$$\overline{X}_n \xrightarrow{a.s.} E(X_1)$$

Example:

Let X_n be iid and let F be the distribution function of X_1

Define

$$F_n(t) = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq t\}} \quad \forall\, t \in \mathbb{R},\ \forall\, n = 1, 2, 3 \dots$$

"Empirical distribution function"

Then, $F_n(t) \xrightarrow{a.s.} F(t) \quad \forall\, t \in \mathbb{R}$

"$F_n(t)$ is a consistent estimator of $F(t)$"

Let's prove this Fact:

Define:

$$y_i = \mathbb{1}_{(x_i \leq t)} \quad , \text{ then } (y_1) \text{ is iid}$$

since (x_i) is iid.

And $E(y_1) = E\left(\mathbb{1}_{(x_1 \leq t)}\right) = P\left(x_1 \leq t\right)$

$= F(t)$

Therefore,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(x_i \leq t)} = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}_n \xrightarrow{a.s.} E(y_1)$$

- Another version of SLLN:

If $\sup_n E(X_n^2) < +\infty$, $Cov(X_i, X_j) \leq 0 \quad \forall\, i \neq j$

and $E(X_i) = E(X_1) \quad \forall\, i$, then

$$\bar{X}_n \xrightarrow{a.s.} E(X_1)$$

Once again, $\underline{V = E(X_1)}$.

Proving this result requires a certain effort, but the proof become very simple if we only show convergence in probability.

In fact, let's take $P\left[\left|\bar{X}_n - E(X_1)\right| > \varepsilon\right]$

Since $E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(x_i) = E(X_1)$

and by Chebichev

$$P\left(\left|\bar{X}_n - E(X_1)\right| > \varepsilon\right) \leq \frac{Var(\bar{X}_n)}{\varepsilon^2}$$

and $\frac{1}{\varepsilon^2} Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{\varepsilon^2}\left\{\frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) + \frac{1}{n^2}\sum_{i=j}^{n} Cov(X_i, X_j)\right\}$.

Since $Cov(X_i, X_j) \leq 0$, $\frac{1}{n^2}\sum_{i=j}^{n} Cov(X_i, X_j) \leq 0$

Hence, $P\left[(\overline{X}_n - E(X_1)) > \varepsilon\right] \leq \frac{1}{\varepsilon^2 \cdot n^2} \sum_{i=1}^{n} Var(X_i) \leq \frac{1}{\varepsilon^2 n^2} \cdot \sum_{i=1}^{n} E(X_i^2)$

Since, $\sup_n E(X_n^2) < +\infty$, there is a $c > 0$ such that

$E(X_i^2) \leq c \quad \forall i.$

Hence, $P\left[|\overline{X}_n - E(X_1)| > \varepsilon\right] = \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^{n} c = \frac{\cancel{n} \cdot c}{\varepsilon^2 \cdot n^{\cancel{2}}} = \frac{c}{\varepsilon^2 \cdot n}$

So that,

$$\lim_n P\left[|\overline{X}_n - E(X_1)| > \varepsilon\right] = \lim_n \frac{c}{\varepsilon^2 \cdot n} = 0$$

$\cdot$ SLLN where V is a general r.v. (not degen.)

If X_n is stationary and $E(X_1) < +\infty$ then

$$\overline{X}_n \xrightarrow{a.s.} V$$ where the r.v. may be non degenerate.

Stationary (X_n):

$$(X_2, X_3, X_4, \dots) \sim (X_1, X_2, X_3, \dots)$$

namely the distribution of the sequence does not change when the sequence is shifted on.

So, by induction we obtain that

$$(X_k, X_{k+1}, X_{k+2}, \dots) \sim (X_1, X_2, X_3, \dots) \quad \forall k \geq 1$$

Hence, $X_k \sim X_1 \ \forall k \geq 1$, namely if the (X_n) is stationary $\Rightarrow (X_n)$ is identically distributed.

Finally, let us see a version of the weak law of large numbers

Weak law of large numbers

If (X_n) is iid and the characteristic function of X_1 is differentiable at 0, then

$$\bar{X}_n \xrightarrow{P} a \ , \text{ for some number } a$$

⊛ Remarks

- In connection with Kolmogorov SLLN, we actually proved that $\bar{X}_n \xrightarrow{P} E(X_1)$, but in our proof we only assumed that the ϕ of X_1 is differentiable at 0.

- If $E[|X_1| < +\infty]$, then the characteristic function is differentiable at 0, but the converse is not true.

So, if (X_n) is iid. it may be that

the WLLN holds while the strong law of large number fails.

$\Downarrow$

Example

(X_n) is iid and X_1 is absolutely continuous with density:

$$f(x) = \begin{cases} \dfrac{c}{x^2 \cdot \log|x|} \cdot \mathbb{1} & \text{if } x \notin [-2;2] \text{ where} \\ & c \text{ is a suitable constant} \\ 0 & \text{if } x \in [-2;2] \end{cases}$$

In this case it can be shown that $\Phi'(c) = 0$ where Φ is the characteristic function of X_1, so that $\overline{X}_n \xrightarrow{P} 0$.

However, $E(X_1) = +\infty$ so that $\overline{X}_n$ does not converge almost surely to a limit.

As to c, $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, then

$$1 = \int_{-\infty}^{+\infty} f(x)\,dx = \int_{-\infty}^{-2} \frac{c}{x^2 \cdot \log|x|}\,dx + \int_{2}^{+\infty} \frac{c}{x^2 \cdot \log|x|}\,dx$$

$$= 2C \cdot \int_{2}^{+\infty} \frac{1}{x^2 \cdot \log|x|}\, dx$$

So that, $$c = \frac{1}{2 \cdot \int_{2}^{+\infty} \frac{1}{x^2 \cdot \log|x|}\, dx}$$

Central limit theorem

In general, a central limit theorem may be described as follows:

- we have a sequence (X_n) of real r.v.
- (X_n) satisfies the CLT if there are constants a_n and $b_n > 0$ such that

$$\sum_{i=1}^{n} \frac{X_i - a_n}{b_n} \xrightarrow{d} N(0,1)$$

- a_n and b_n arbitrary constants

- Important special case is the following

$$a_n = E\left(\sum_{i=1}^{n} X_i\right), \quad b_n = \sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}$$

→ Most popular of the CLT

If (X_n) is iid, $E(X_i^2) < +\infty$ and X_1 is not degenerate

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \xrightarrow{d} N(0,1)$$

- In the iid case, taking

$\mu = E(X_1)$ and $\sigma^2 = Var(X_1)$, we have that

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sigma\left(\sum_{i=1}^{n} X_i\right)} = \frac{\sum_{i=1}^{n} X_i - n\cdot\mu}{\sqrt{n\cdot\sigma^2}} = \frac{\sqrt{n}}{\sigma}\cdot\left(\bar{X}_n - \mu\right)$$

→ Another version of the CLT

If (X_n) is independent, $E(X_n) = 0 \ \forall n$ and

$$\frac{\sum_{i=1}^{n} E|X_i|^{\alpha}}{\left(\sum_{i=1}^{n} E(X_i^2)\right)^{\frac{\alpha}{2}}} \xrightarrow{\text{as } n\to+\infty} 0 \quad \text{for some } \alpha > 2.$$

then we can say that

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \xrightarrow{d} N(0,1)$$

Example:

Suppose (X_n) independent, $E(X_n) = 0 \;\forall n$,

$$\sum_{i=1}^{n} E(X_i)^2 \longrightarrow +\infty, \quad |X_i| \le c \;\forall i.$$

Under these conditions, what about convergence in distribution of

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \;?$$

Now, $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} E(X_i^2)$ [since $E(X_n) = 0$]

And we decide to let $\alpha = 3$ in the previous theorem;

$$\frac{\sum_{i=1}^{n} E|X_i|^3}{\left(\sum_{i=1}^{n} E(X_i^2)\right)^{3/2}} = \frac{\sum_{i=1}^{n} E\left(|X_i| \cdot X_i^2\right)}{\left(\sum_{i=1}^{n} E(X_i^2)\right)^{3/2}} \le \frac{c \cdot \sum E(X_i^2)}{\left(\sum_{i=1}^{n} E(X_i^2)\right)^{3/2}}$$

$$\le \frac{c}{\left(\sum_{i=1}^{n} E(X_i^2)\right)^{1/2}} \xrightarrow{\text{as } n \to +\infty} 0$$

Hence by the 2° version of the CLT, we conclude that

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \xrightarrow{d} N(0,1)$$

Example

Suppose (X_n) iid, $E(X_1)=0$, $E(X_1^2)=1$

What about the convergence in distribution of

$$\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} X_i^2}}\right) ?$$

We can write this fraction as follows:

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} X_i^2}} = \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \cdot \frac{1}{\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}}$$