II. Conditioning.

Given $(\Omega, \mathcal{F}, \mathbb{P})$ and $B \in \Omega$ with $\mathbb{P}[B] > 0$ there is a natural induced probability measure on B, $\mathcal{F}_B = \{A \in \mathcal{F} \mid A \subset B\}$:

$$\mathbb{P}_B[A] = \frac{\mathbb{P}[A]}{\mathbb{P}[B]} \quad \text{for } A \subseteq B.$$

$\mathbb{P}_B$ is actually the restriction to $\mathcal{F}_B$ of a probability measure on $(\Omega, \mathcal{F})$:

$$\mathbb{P}_B[A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$ conditional probability of A knowing B.
also denoted $\mathbb{P}[A \mid B]$

A and B independent $\iff \mathbb{P}[A \mid B] = \mathbb{P}[A]$
$\iff$ knowing that B happens gives no information on the probability of A.

More generally, we would like to define the conditional expectation of random variables, knowing the occurrences of a bunch of events $\rightsquigarrow$ conditional expectation w.r.t to a σ-subalgebra.

Chapter II. Conditioning.

1. Conditional expectation.

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space, $\mathcal{A} \subset \mathcal{F}$ σ-subalgebra.
$X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Definition There exists a unique variable $\tilde{X} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ s.t., for any $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$ r.v. $\mathcal{A}$-measurable and bounded,

$$\mathbb{E}[XZ] = \mathbb{E}[\tilde{X}Z]. \qquad (*)$$

One denotes $\tilde{X} = \mathbb{E}[X \mid \mathcal{A}]$: this is the conditional expectation of X knowing $\mathcal{A}$

Intuition : $\mathbb{E}[X|\mathcal{A}]$ is the function of occurrences of events in $\mathcal{A}$ which, knowing said occurrences, returns the corresponding expected value of X.

Proof (unicity) : suppose that $\tilde{X}^1$ and $\tilde{X}^2$ satisfy $(*)$

Take $Z = \mathbb{1}_{\tilde{X}^1 > \tilde{X}^2}$. Since $\tilde{X}^1$ and $\tilde{X}^2$ are $\mathcal{A}$measurable, Z is in $L^\infty(\Omega, \mathcal{A}, \mathbb{P})$, so :

$$\mathbb{E}[\underbrace{(\tilde{X}^1 - \tilde{X}^2)\mathbb{1}_{(\tilde{X}^1 > \tilde{X}^2)}}_{\geq 0}] = \mathbb{E}[\tilde{X}^1 Z] - \mathbb{E}[\tilde{X}^2 Z] = 0.$$

So, almost surely, $\mathbb{1}_{\tilde{X}^1 > \tilde{X}^2} = 0$, and by symmetry $\mathbb{1}_{\tilde{X}^1 < \tilde{X}^2} = 0$.

Hence, $\mathbb{P}[\tilde{X}^1 = \tilde{X}^2] = 1$.

Proof (L^2 variables). Suppose that X is square integrable.

The space $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a closed subspace of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Therefore, the orthogonal projection $\tilde{X} = \Pi_{L^2(\Omega, \mathcal{A}, \mathbb{P})}(X)$ is well defined.

By definition, $\forall Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$,

$$\mathbb{E}[XZ] = \mathbb{E}[\tilde{X}Z].$$

In particular, this is true if $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$, whence the existence.

Proof (L^1 variables). Given $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ there exists a sequence of L^2 variables $(X_n)_{n\in\mathbb{N}}$ which converges L^1 to X.

Each X_n admits a conditional expectation $\tilde{X}_n$ knowing the σ-subalgebra $\mathcal{A}$.

Let us then remark that

$$\mathbb{E}[\,|\tilde{X}_n - \tilde{X}_m|\,] = \mathbb{E}[(\tilde{X}_n - \tilde{X}_m)\underbrace{\operatorname{sgn}(\tilde{X}_n - \tilde{X}_m)}_{\in L^\infty(\Omega, \mathcal{A}, \mathbb{P}),\ =Z}]$$

$$= \mathbb{E}[(X_n - X_m)\underbrace{Z}_{=\pm 1}] \leq \mathbb{E}[\,|X_n - X_m|\,].$$

So $(\tilde{X}_n)_{n\in\mathbb{N}}$ is Cauchy in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ and converges towards some $\tilde{X}$.

We then have, for any $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$:

$$\left|\mathbb{E}[XZ] - \mathbb{E}[\tilde{X}Z]\right| \leq \left|\mathbb{E}[(X - X_n)Z]\right| + \left|\mathbb{E}[(\tilde{X}_n - \tilde{X})Z]\right|$$

$$\leq \|Z\|_\infty \left(\mathbb{E}[\,|X - X_n|\,] + \mathbb{E}[\,|\tilde{X}_n - \tilde{X}|\,]\right) \xrightarrow[n\to+\infty]{} 0.$$

$\square$.

<u>Properties</u> of $\mathbb{E}[\cdot\,|\,\mathcal{A}] : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to L^1(\Omega, \mathcal{A}, \mathbb{P})$:

① linearity.

② positivity: if $X \geq 0$ a.s., then $\mathbb{E}[X\,|\,\mathcal{A}] \geq 0$ a.s.
(take $Z = \mathbb{1}_{\tilde{X}<0}$ in $*$).

③ triangular inequality: $|\mathbb{E}[X\,|\,\mathcal{A}]| \leq \mathbb{E}[\,|X|\,|\,\mathcal{A}]$.
(positivity of $|X| - X$ and $|X| + X$).

④ if X is already $\mathcal{A}$-measurable, then $X = \mathbb{E}[X\,|\,\mathcal{A}]$ a.s.

⑤ if X is independent from $\mathcal{A}$, then $\forall Z \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}[XZ] = \mathbb{E}[X]\,\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[X]Z]$$

so $\mathbb{E}[X\,|\,\mathcal{A}] = \mathbb{E}[X]$.

⑥ successive conditioning: if $\mathcal{A} \subset \mathcal{B} \subset \mathcal{F}$, $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{B}]\,|\,\mathcal{A}] = \mathbb{E}[X\,|\,\mathcal{A}]$

⑦ expectation by conditioning: $\mathbb{E}[\mathbb{E}[X\,|\,\mathcal{A}]] = \mathbb{E}[X]$.

remarks: ① Since $\mathbb{E}[X \mid \mathcal{A}] \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ it is defined up to a $\mathbb{P}$-negligible subset. In particular, one is allowed to modify its value on any subset with $\mathbb{P}$ probability 0.

② generalisation of ④ and ⑤: one can extract $\mathcal{A}$-measurable variables
integrate $\mathcal{A}$-independent

2. Computations and conditional distributions.

① discrete conditioning

if $\mathcal{A} = \sigma(A) = \{\phi, A, A^c, \Omega\}$, then any $\mathbb{E}[X \mid \mathcal{A}]$ must be a linear combination of $\mathbb{1}_A$ and $\mathbb{1}_{A^c}$.
$\downarrow$
some event with $\mathbb{P}[A] > 0$

More generally, suppose that $\Omega = \bigsqcup_{i \in I} A_i$ is a decomposition of Ω into disjoint $\mathcal{F}$-measurable subsets with $\left| \begin{array}{l} I \text{ countable} \\ \mathbb{P}[A_i] > 0 \quad \forall i \in I. \end{array} \right.$

$\mathcal{A} = \sigma(A_i, i \in I) = \left\{ \bigsqcup_{j \in J} A_j, \; J \subset I \right\}$ discrete σ-subalgebra of $\mathcal{F}$.

Lemma: Any $\mathcal{A}$-measurable function is a linear combination of functions $\mathbb{1}_{A_i}$.

Indeed, the sense $\Leftarrow$ is obvious, and conversely, if f is $\mathcal{A}$-measurable, then for any $x, y \in A_i$ for some $i \in I$,

$f^{-1}(f(x))$ is some $\bigsqcup_{j \in J} A_j$ containing x, hence A_i.

so $f(x) = f(y) \Rightarrow f$ is constant on any $A_i \Rightarrow f$ is a linear combination of functions $\mathbb{1}_{A_i}$.

$X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. $\mathbb{E}[X \mid \mathcal{A}]$?

We claim that $\mathbb{E}[X \mid \mathcal{A}] = \sum_{i \in I} \dfrac{\mathbb{E}[X \mathbb{1}_{A_i}]}{\mathbb{P}[A_i]} \mathbb{1}_{A_i}$.

Indeed, if $\tilde{X}$ is the RHS, then $\forall i \in I$,

$$\mathbb{E}[\tilde{X} \mathbb{1}_{A_i}] = \mathbb{E}\left[\sum_{j \in I} \frac{\mathbb{E}[X \mathbb{1}_{A_j}]}{\mathbb{P}[A_j]} \mathbb{1}_{A_j} \mathbb{1}_{A_i}\right]$$

$$= \mathbb{E}\left[\frac{\mathbb{E}[X \mathbb{1}_{A_i}]}{\mathbb{P}[A_i]} \mathbb{1}_{A_i}\right] = \mathbb{E}[X \mathbb{1}_{A_i}]$$

so $\tilde{X}$ has the characteristic property. $\square$

In particular, if $X = \mathbb{1}_B$, then $\mathbb{E}[\mathbb{1}_B | \mathcal{A}] = \sum_{i \in I} \mathbb{P}[B | A_i] \mathbb{1}_{A_i}$

$\rightarrow$ one recovers the notion of conditional probability.

By taking the expectation, one obtains the formula:

$$\mathbb{P}[B] = \sum_{i \in I} \mathbb{P}[B | A_i] \mathbb{P}[A_i].$$

2) conditioning by a random variable

The most important case of conditioning is when $\mathcal{A} = \sigma(Y)$ with $Y : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ measurable.

Lemma: any $\sigma(Y)$ measurable function $Z : (\Omega, \sigma(Y)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ can be written as $Z = f(Y)$ with $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable.

proof: again, the $\Leftarrow$ sense is obvious.

For $\Rightarrow$, we can assume that Y is surjective: replace E by $Y(\Omega)$

$\mathcal{E}$ by $\{A \cap Y(\Omega), A \in \mathcal{E}\}$.

For any $[a,b)$, $Z^{-1}([a,b))$ is in $\sigma(Y)$, so it can be written as $Y^{-1}(M_{[a,b)})$ with $M_{[a,b)} \in \mathcal{E}$. Then:

$$Z(\omega) \in [a,b) \iff Y(\omega) \in M_{[a,b)}.$$

Since Y is surjective, $M_{[a,b)} = Y(Z^{-1}([a,b)))$. This implies that:

- if $[c,d) \subset [a,b)$, then $M_{[c,d)} \subset M_{[a,b)}$.
- if $([a_i, b_i))_{i \in I}$ is a family of disjoint intervals, then so are the $M_{[a_i,b_i)}$.

Set then $F_n = \sum_{k\in\mathbb{Z}} \frac{k}{2^n} \mathbb{1}_{M_{[\frac{k}{2^n}, \frac{k+1}{2^n})}}$.

- for any n, $\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right)_{k\in\mathbb{Z}}$ is a measurable set partition of $\mathbb{R}$, so

$\left(M_{[\frac{k}{2^n}, \frac{k+1}{2^n})}\right)_{k\in\mathbb{Z}}$ ——————————— E.

- if $Y(\omega) \in M_{[\frac{k}{2^n}, \frac{k+1}{2^n})}$, then $\frac{k}{2^n} = F_n(Y(\omega)) \leq Z(\omega) < \frac{k+1}{2^n}$,

hence, $|Z - F_n \circ Y| \leq \frac{1}{2^n}$ everywhere.

- since $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right) \sqcup \left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right)$, the same holds for the corresponding $M_{[a,b)}$, so $(F_n)_{n\in\mathbb{N}}$ is increasing with n.

$|F_{n+1} - F_n| \leq \frac{1}{2^{n+1}}$.

We conclude that $F_n \longrightarrow f$ for some measurable function f on E with $Z = f \circ Y$. $\square$

If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, we denote $\mathbb{E}[X \mid \sigma(Y)] = \mathbb{E}[X \mid Y]$.
This is a measurable function $g(Y)$, with $g \in L^1(E, \mathcal{E}, \mathbb{P}_Y)$.

- case of discrete variables X, Y with values in $\mathbb{Z}$

It is then easy to check that $\mathbb{E}[X \mid Y] = \sum_{\substack{k\in\mathbb{Z} \\ \mathbb{P}[Y=k]>0}} \frac{\mathbb{E}[X \mathbb{1}_{Y=k}]}{\mathbb{P}[Y=k]} \mathbb{1}_{Y=k}$

example: $X \sim \text{Poisson}(\lambda)$, $\tilde{X} \sim \text{Poisson}(\mu)$ independent. $Y = X + \tilde{X} \sim \text{Poisson}(\lambda+\mu)$.

We have $\mathbb{E}[X \mathbb{1}_{Y=k}] = \sum_{\ell=0}^{k} \mathbb{P}[X = \ell, \tilde{X} = k-\ell]\, \ell$

$$= \sum_{\ell=0}^{k} e^{-\lambda-\mu} \frac{\lambda^\ell \mu^{k-\ell}}{\ell!\,(k-\ell)!}\, \ell$$

$$= \lambda \sum_{\ell=1}^{k} e^{-\lambda-\mu} \frac{\lambda^{\ell-1} \mu^{k-\ell}}{\ell-1!\,(k-\ell)!} = \lambda e^{-\lambda-\mu} \frac{(\lambda+\mu)^{k-1}}{(k-1)!}$$

So, $\mathbb{E}[X|Y] = \frac{\lambda}{\lambda+\mu} \sum_{k=0}^{\infty} k \, \mathbb{1}_{Y=k} = \frac{\lambda Y}{\lambda+\mu}$.

important remark: the computation above involves the conditional probability

$$\mathbb{P}_{X|Y=k}[\ell] = \frac{\mathbb{P}[X=\ell \text{ et } Y=k]}{\mathbb{P}[Y=k]} = \binom{k}{\ell} \left(\frac{\lambda}{\lambda+\mu}\right)^{\ell} \left(\frac{\mu}{\lambda+\mu}\right)^{k-\ell}$$

$=$ binomial distribution.

We denote $\mathbb{P}_{X|Y}[\ell]$ = the $\sigma(Y)$-measurable random variable equal to $\mathbb{P}_{X|Y=k}[\ell]$ when $Y=k$.

$$\text{Then, } \mathbb{E}[X|Y] = \sum_{\ell \in \mathbb{N}} \mathbb{P}_{X|Y}[\ell]\, \ell.$$

- case of continuous random variables X, Y with values in $\mathbb{R}$ and a joint density f.

Then, by copying the discrete case, we define the conditional density:

$$f_{X|Y=y}(x) = \begin{cases} \dfrac{f(x,y)}{\int f(x,y)\,dx} & \text{if } f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx > 0. \\ \text{some density, e.g. } \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & \text{if } f_Y(y) = 0. \end{cases}$$

As before, we denote $f_{X|Y}(x)$ the $\sigma(Y)$-measurable random variable obtained by substitution of y by Y in the formula above.

$$\text{Then, } \mathbb{E}[X|Y] = \int_{\mathbb{R}} x \, f_{X|Y}(x)\,dx.$$

Indeed, for any bounded random variable $Z = g(Y)$, if $h(Y)$ is the right-hand side, then:

$$\mathbb{E}[h(Y)Z] = \int_{\mathbb{R}} h(y)\, g(y)\, f_Y(y)\,dy$$

$$= \iint_{\mathbb{R}^2} x\, g(y)\, f(x,y)\, dx\, dy = \mathbb{E}[X g(Y)] = \mathbb{E}[XZ].$$

<u>example</u>: $X \sim \text{Exp}(\lambda)$, $\tilde{X} \sim \text{Exp}(\mu)$ independent. $Y = X + \tilde{X}$.

The joint law of (X,Y) can be computed as follows: for any $g: \mathbb{R}^2 \to \mathbb{R}$ bounded measurable,

$$\mathbb{E}[g(X,Y)] = \iint_{\mathbb{R}_+^2} g(x, x+\tilde{x})\, \lambda\mu\, e^{-\lambda x - \mu \tilde{x}}\, dx\, dy$$

$$= \iint_{\mathbb{R}_+^2} g(x,y) \underbrace{\mathbb{1}_{0 \le x \le y}\, \lambda\mu\, e^{-(\lambda-\mu)x - \mu y}}_{\text{the joint density of } (X,Y).}\, dx\, dy$$

In particular, if $\lambda = \mu$, then the joint density is $f(x,y) = \mathbb{1}_{0 \le x \le y}\, \lambda^2 e^{-\lambda y}$.

Then, the conditional density is:

$$f_{X|Y=y}(x) = \frac{\mathbb{1}_{0 \le x \le y}\, \lambda^2 e^{-\lambda y}}{\int_{x=0}^{y} \lambda^2 e^{-\lambda y}\, dx} = \frac{\mathbb{1}_{0 \le x \le y}}{y}.$$

Therefore, $\mathbb{E}[X|Y] = \int_0^Y x\, f_{X|Y}(x)\, dx = \frac{1}{Y}\int_0^Y x\, dx = \frac{Y}{2}$.

3 <u>conditional distributions</u>

In the examples above:

- $X \sim \text{Poi}(\lambda)$, $\tilde{X} \sim \text{Poi}(\mu)$, $Y = X + \tilde{X}$: conditional to Y, it looks like $X \sim \text{Bin}(Y, \frac{\lambda}{\lambda+\mu})$.
- $X \sim \text{Exp}(\lambda)$, $\tilde{X} \sim \text{Exp}(\lambda)$, $Y = X + \tilde{X}$: conditional to Y, it looks like $X \sim \text{Unif}([0,Y])$.

random $\sigma(Y)$-measurable law $\mathbb{P}_{X|Y}$?

We need to endow the set of probability measures $\mathcal{M}^1(\mathbb{R})$ with a σ-algebra:

$$\mathcal{G} = \sigma(\{\mu \mapsto \mu(A),\ A \text{ Borel subset of } \mathbb{R}\})$$
$$= \sigma(\{\mu \mapsto \int F(x)\mu(dx),\ F \text{ measurable bounded function on } \mathbb{R}\}).$$

<u>Theorem</u> Let X, Y be random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$

X with values in $\mathbb{R}$.

Y ———— in some space $(E, \mathcal{E})$.

There exists a measurable map $\mathbb{P}_{X|}$: $E \longrightarrow \mathcal{M}^1(\mathbb{R})$,
$y \longmapsto \mathbb{P}_{X|y}$

unique up to a $\mathbb{P}_Y$ negligible subset, such that for any bounded measurable function $F: \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[F(X) \mid Y] = \int_{\mathbb{R}} F(x)\, \mathbb{P}_{X|Y}(dx). \quad (**)$$

<u>remarks</u>: • for the space of values of X, we can replace $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by any measurable space $(D, \mathcal{D})$ measurably isomorphic to a measurable subset of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; in particular this works with $\mathbb{R}^n$.

• by using the DCT / MCT, one can extend the identity $**$ to more general functions, e.g. $F(x) = x^k$.

<u>lemma</u>: if $(X_n)_{n\in\mathbb{N}}$ is an increasing sequence of bounded random variables,
$$X = \lim_{n\to+\infty} X_n$$
then for any σ-subalgebra $\mathcal{A}$, $\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{A}] = \mathbb{E}[X|\mathcal{A}]$ a.s.

<u>proof</u>: since $\mathbb{E}[\cdot|\mathcal{A}]$ is a positive linear map:

- $(\mathbb{E}[X_n|\mathcal{A}])_{n\in\mathbb{N}}$ is an increasing sequence of random variables.
- $\mathbb{E}[X_n|\mathcal{A}] \leq \mathbb{E}[X|\mathcal{A}]$ a.s. for any n.

Therefore, $\mathbb{E}[X_n \mid \mathcal{A}]$ converges almost surely towards a $\mathcal{A}$-measurable random variable Y which is smaller than $\mathbb{E}[X \mid \mathcal{A}]$.

Moreover, since $\mathbb{E}[\cdot \mid \mathcal{A}]$ is 1-Lipschitz $L^1(\Omega,\mathcal{F},\mathbb{P}) \to L^1(\Omega,\mathcal{A},\mathbb{P})$

$$\mathbb{E}[|Y - \mathbb{E}[X \mid \mathcal{A}]|] \leq \mathbb{E}[|Y - \mathbb{E}[X_n \mid \mathcal{A}]|] + \mathbb{E}[|\mathbb{E}[X_n \mid \mathcal{A}] - \mathbb{E}[X \mid \mathcal{A}]|]$$
$$\leq \mathbb{E}[|Y - \mathbb{E}[X_n \mid \mathcal{A}]|] + \mathbb{E}[|X_n - X|]$$

both go to 0 by the dominated convergence theorem

so $Y = \mathbb{E}[X \mid \mathcal{A}]$ almost surely. $\square$

<u>Proof of the theorem</u> We define the conditional cumulative distribution function.
For $r \in \mathbb{Q}$, consider $\mathbb{E}[\mathbb{1}_{X \leq r} \mid Y] = F_r(Y)$ for some measurable function $F_r : E \to \mathbb{R}$.

- since $0 \leq \mathbb{1}_{X \leq r} \leq 1$, $0 \leq F_r \leq 1$ $\mathbb{P}_Y$ almost surely.
- if $r \leq s$, then $\mathbb{1}_{X \leq r} \leq \mathbb{1}_{X \leq s}$, so $F_r \leq F_s$ $\mathbb{P}_Y$ almost surely.
- since $\lim_{r \to \pm\infty} \mathbb{1}_{X \leq r} = 0$ or 1, $\lim_{r \to \pm\infty} F_r = 0$ or 1 $\mathbb{P}_Y$ a.s.

By countable intersection, there exists a subset $B \subset E$ with $\mathbb{P}_Y[B] = 1$ and all the properties above for F_r, $r \in \mathbb{Q}$.

We define for $t \in \mathbb{R}$:
$$F_t : E \to [0,1]$$
$$y \mapsto \mathbb{1}_{y \in B} \inf\{F_r(y),\ r \in \mathbb{Q},\ r \geq t\} + \mathbb{1}_{y \notin B} \mathbb{1}_{t \geq 0}.$$

For any $y \in E$, $t \mapsto F_t(y)$ is a cumulative distribution function of a real valued random variable.

Denote $\mathbb{P}_{X|y}$ the associated probability measures on $\mathbb{R}$.
$$\mathbb{P}_{X|y}[(-\infty, t]] = F_t(y).$$

1) For any $A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}[X \in A | Y] = \int_A \mathbb{P}_{X|Y}[dx]$. $\mathbb{P}_Y$ almost surely.

Indeed, this is true if $A = (-\infty, t]$: if $Y \in B$, which happens almost surely, then $\int_A \mathbb{P}_{X|Y}[dx] = \int_{-\infty}^{t} \mathbb{P}_{X|Y}[dx] = F_t(Y)$

$$= \inf \{ F_r(Y), r \geq t, r \in \mathbb{Q} \}$$

$$= \inf \{ \mathbb{E}[\mathbb{1}_{X \leq r} | Y], r \geq t, r \in \mathbb{Q} \}$$

$$= \text{(by the lemma) } \mathbb{E}[\mathbb{1}_{X \leq t} | Y] = \mathbb{P}[X \in A | Y].$$

Then, by the monotone class lemma, this extends to any Borel subset.

2) If $F : \mathbb{R} \to \mathbb{R}$ is measurable bounded, then it is a limit of step functions:

$$F(x) = \lim_{n \to +\infty} \uparrow \underbrace{\sum_{k \in \mathbb{Z}} \mathbb{1}_{x \in F^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))} \frac{k}{2^n}}_{F_n}.$$

and $\|F - F_n\|_\infty \leq \frac{1}{2^n}$.

Then, $\mathbb{E}[F_n(X) | Y] = \sum_{k \in \mathbb{Z}} \mathbb{E}[\mathbb{1}_{X \in F^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))} | Y] \frac{k}{2^n}$

$$\underset{a.s.}{=} \sum_{k \in \mathbb{Z}} \frac{k}{2^n} \int \mathbb{1}_{F^{-1}([\frac{k}{2^n}, \frac{k+1}{2^n}))}(x) \mathbb{P}_{X|Y}[dx]$$

$$= \int F_n(x) \mathbb{P}_{X|Y}[dx]$$

and by taking the limit, we get the same idendity for F.

3) $y \mapsto \mathbb{P}_{X|y}$ is $\sigma(Y)$-measurable: the set of subsets A such that $y \mapsto \mathbb{P}_{X|y}[A]$ is ________ is a monotone class containing the $(-\infty, t]$.

so, $\forall A \in \mathcal{B}(\mathbb{R})$, $y \longmapsto \mathbb{P}_{X|y}[A]$ is $\sigma(Y)$-measurable, which means that $y \mapsto \mathbb{P}_{X|y}$ is $\sigma(Y)$-measurable.

4) unicity: suppose that $M_1, M_2 : E \to \mathcal{M}^1(\mathbb{R})$ are two $\sigma(Y)$-measurable maps such that $(**)$ holds:

$$\mathbb{E}[F(X)|Y] \underset{a.s}{=} \int_{\mathbb{R}} F(x)\, M_i(Y)[dx] \quad \forall F \text{ bounded measurable},$$

then $\exists B$ with $\mathbb{P}[Y \in B] = 1$ such that $\forall y \in B$, $\forall r \in \mathbb{Q}$,

$$\int_{-\infty}^{r} M_1(y)[dx] = \int_{-\infty}^{r} M_2(y)[dx].$$

Then, $M_1(y)$ and $M_2(y)$ have the same cumulative distribution function $\forall y \in B$, so they are the same measure $\square$.

---

3. Applications

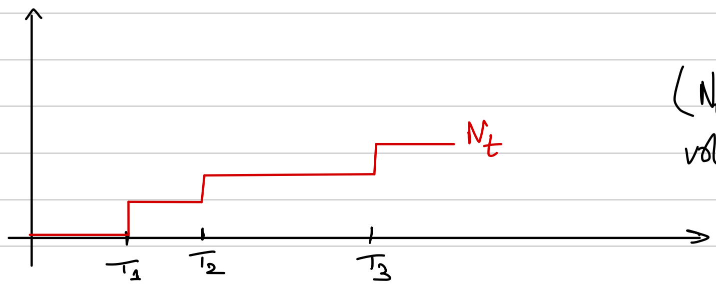
(1) Poisson process on $\mathbb{R}_+$

Consider independent exponential random variables $E_1, E_2, \dots, E_n, \dots \sim \text{Exp}(\lambda)$.

Set $T_i = E_1 + \dots + E_i$.

$$N_t = \sup\{i \geq 0 : T_i \leq t\} \quad \text{for } t \in \mathbb{R}_+.$$

Poisson process with intensity λ.



$(N_t)_{t \in \mathbb{R}_+}$ is right continuous, with values in $\mathbb{N}$. $\lim_{t \to +\infty} N_t = +\infty$ a.s.

- $\forall t$, N_t follows a Poisson law with parameter λt. Indeed:

$$\mathbb{P}[N_t = 0] = \mathbb{P}[E_1 > t] = e^{-\lambda t}.$$

and $\forall k \geq 1$, $\mathbb{P}[N_t = k] = \mathbb{P}[T_k \leq t < T_k + E_{k+1}]$

$$= \int_{(\mathbb{R}_+)^{k+1}} \lambda^{k+1} e^{-\lambda(s_1+\dots+s_{k+1})} \mathbb{1}_{s_1+\dots+s_k \le t < s_1+\dots+s_{k+1}} ds_1 \dots ds_{k+1}$$

$$= \int_{\mathbb{R}_+^k} \lambda^k e^{-\lambda(s_1+\dots+s_k)} \mathbb{1}_{s_1+\dots+s_k \le t} \underbrace{\left(\int_{s_{k+1}=t-s_1-\dots-s_k} \lambda e^{-\lambda s_{k+1}} ds_{k+1} \right)}_{e^{-\lambda t + \lambda(s_1+\dots+s_k)}} ds_1 \dots ds_k$$

$$= \lambda^k e^{-\lambda t} \underbrace{\int_{\mathbb{R}_+^k} \mathbb{1}_{s_1+\dots+s_k \le t} ds_1 \dots ds_k}_{\text{volume of the } k\text{-dimensional simplex } = \frac{t^k}{k!}}$$

$$= (\lambda t)^k \frac{e^{-\lambda t}}{k!}.$$

- conditional to $N_t = k$, $\{T_1, \dots, T_k\}$ has the law of k independent uniform points chosen in $[0,t]$.

$$\mathbb{E}[F(T_1, T_2, \dots, T_k) \mid N_t = k] = \frac{k!}{t^k} \int_{[0,t]^k} \mathbb{1}_{t_1 \le t_2 \le \dots \le t_k \le t} F(t_1, \dots, t_k) \, dt_1 \dots dt_k.$$

Indeed :

$$\mathbb{E}[F(T_1, \dots, T_k) \mathbb{1}_{N_t = k}] =$$

$$\int_{\mathbb{R}_+^k} ds_1 \dots ds_k \; F(s_1, \dots, s_1+\dots+s_k) \lambda^k e^{-\lambda(s_1+\dots+s_k)} \mathbb{1}_{s_1+\dots+s_k \le t} \, ds_1 \dots ds_k.$$

$$\underset{\text{(change of variables)}}{=} \lambda^k e^{-\lambda t} \int_{[0,t]^k} \mathbb{1}_{t_1 \le t_2 \le \dots \le t_k \le t} F(t_1, \dots, t_k) \, dt_1 \dots dt_k$$

and we get the desired formula by dividing by $\mathbb{P}[N_t = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$.

② Gaussian vectors

Definition: A Gaussian vector is a random vector $(X_1, \dots, X_n) \in \mathbb{R}^n$ such that any linear combination $\sum_{i=1}^n t_i X_i$ follows a normal distribution $\mathcal{N}(m(t), \sigma^2(t))$.

Since $\mathbb{E}[\cdot]$ is a linear form, $m(t) = \sum_{i=1}^n t_i m_i$ with $m_i = \mathbb{E}[X_i]$

$m = (m_1, \dots, m_n)$ is the mean vector

Then, $t \mapsto \sigma^2(t)$ is a quadratic form, so it writes as:

$$\sigma^2(t) = t^* \Sigma t \quad \text{with} \quad \Sigma = (\Sigma_{ij})_{1 \le i,j \le n} = \underbrace{(\mathrm{cov}(X_i, X_j))_{1 \le i,j \le n}}_{\text{covariance matrix}}$$

with $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ if $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$

$\mathrm{var}(X) = \mathrm{cov}(X, X)$.

Lemma: The distribution of $(X_1, \dots, X_n) \in \mathbb{R}^n$ is entirely determined by the pair (m, Σ).

Proof: We shall see in the next chapter that the Fourier transform $\Phi_X(\xi_1, \dots, \xi_n) = \mathbb{E}[e^{i(\xi_1 X_1 + \dots + \xi_n X_n)}]$ which is a continuous map $\mathbb{R}^n \to \mathbb{C}$ with $\Phi_X(0, \dots, 0) = 1$, determines uniquely the law of X.

However, $\Phi_X(\xi_1, \dots, \xi_n) =$ Fourier transform at $\xi = 1$ of $\mathcal{N}(m(\xi), \sigma^2(\xi))$

$$= e^{i m(\xi) - \frac{\sigma^2(\xi)}{2}}$$

$$= e^{i \sum_{j=1}^n m_j \xi_j - \frac{1}{2} \sum_{j,k=1}^n m_j m_k \Sigma_{jk}}. \qquad \square$$

We then denote $(X_1, \dots, X_n) \sim \mathcal{N}(m, \Sigma)$. By construction, $m \in \mathbb{R}^n$ and Σ is a non negative symmetric matrix.

Conversely, any such pair corresponds to a distribution of normal vector.

Indeed, note that if $X \sim \mathcal{N}(m, \Sigma)$, then $AX \sim \mathcal{N}(Am, A\Sigma A^*)$ for $A \in \mathbb{R}^{n\times n'}$.
Choose $A \in O(n, \mathbb{R})$ such that:

$$A \Sigma A^* = \mathrm{diag}(\lambda_1 \geq \dots \geq \lambda_n \geq 0)$$
$$Am = (\mu_1, \dots \mu_n).$$

Y = vector of independent random vectors with $Y_i \sim \mathcal{N}(\mu_i, \lambda_i)$.
$X = A^{-1}Y$ works.

(with by convention $\mu_i = \mathcal{N}(\mu_i, 0)$).

Suppose Σ is invertible. Then, Y has density w.r.t. the Lebesgue measure on $\mathbb{R}^n$:

$$\frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2} \sum \frac{(y_i - \mu_i)^2}{\lambda_i}\right).$$

Therefore:

$$\mathbb{E}[f(X)] = \int f(A^{-1}y) \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle (y - Am) | (A\Sigma^{-1}A^*)(y - Am)\rangle} dy.$$

$$= \int_{\mathbb{R}^n} f(x) \underbrace{\frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2} (x-m)^t \Sigma^{-1} (x-m)\right)}_{\text{density of } X \sim \mathcal{N}(m,\Sigma) \text{ w.r.t Lebesgue measure.}} dx$$

Let us then investigate conditioning and independence of Gaussian vectors.

Independence. Let $X = (X_1, \dots, X_m, X_{m+1}, \dots, X_n)$ be a Gaussian vector.

Proposition: $(X_1, \dots, X_m)$ and $(X_{m+1}, \dots, X_n)$ are independent

$\iff$ the covariance matrix Σ is block-diagonal.

$$\Sigma = (\mathrm{cov}(X_i, X_j))_{i,j \leq n} = \begin{pmatrix} \Sigma_{[\![1,m]\!]} & 0 \\ 0 & \Sigma_{[\![m+1,n]\!]} \end{pmatrix}.$$

Proof $\Rightarrow$ If the subvectors $(X_1, \dots, X_m)$ and $(X_{m+1}, \dots, X_n)$ are independent, then $\forall i \leq m$ and $j \geq m+1$,

$$\begin{aligned} \operatorname{cov}(X_i, X_j) &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \\ &= \mathbb{E}[X_i]\mathbb{E}[X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] = 0. \end{aligned}$$

Conversely, if Σ is block-diagonal, then the construction of $X \sim \mathcal{N}(m, \Sigma)$ as above allows one to write $X_{[\![1,m]\!]} = A_1^{-1}(Y_1, \dots, Y_m)$

$X_{[\![m+1,n]\!]} = A_2^{-1}(Y_{m+1}, \dots, Y_n)$ with the Y_i independent.

Therefore, $X_{[\![1,m]\!]}$ and $X_{[\![m+1,n]\!]}$ are independent.

(this is a property of the distribution $\mathbb{P}_X$, so we are allowed to realise the variables $X_1, \dots, X_n$ as we like). $\square$

<u>remark</u> Given a centered Gaussian vector $X \sim \mathcal{N}(0, \Sigma)$, set $V = \operatorname{Span}_{\mathbb{R}}(X_1, \dots, X_n)$. We have $\dim V = \operatorname{rank} \Sigma$.

Indeed, with the same notation as above, $V = \operatorname{Span}_{\mathbb{R}}(X) = \operatorname{Span}_{\mathbb{R}}(Y_1, \dots, Y_r)$ with $Y = AX$, $\Sigma = A^* \operatorname{diag}(\lambda_1, \dots, \lambda_r, 0 - 0) A$.

Then, cov is a scalar product on V with orthogonal basis $Y_1, \dots, Y_r$.

$\Rightarrow \dim V = \operatorname{rank} Y = \operatorname{rank} \Sigma$.

<u>Conditioning</u>. Let $(X_1, \dots, X_n, Y)$ be a Gaussian vector. What is the conditional distribution of Y knowing $X_1, \dots, X_n$?

Without loss of generality, one can assume the vector centered: $\mathbb{E}[X_i] = \mathbb{E}[Y] = 0$

<u>Proposition</u> Denote $\overline{Y} = \sum_{i=1}^{n} \lambda_i X_i$ the orthogonal projection of Y on $\operatorname{Span}_{\mathbb{R}}(X_1, \dots, X_n)$

① We have $\overline{Y} = \mathbb{E}[Y \mid X_1, \dots, X_n]$

② Set $\sigma^2 = \mathbb{E}[(Y - \overline{Y})^2]$.

The conditional law of Y knowing $X_1, \dots, X_n$ is $\mathcal{N}(\overline{Y}, \sigma^2)$.

<u>Proof</u>: ① Set $V = \mathrm{Span}_{\mathbb{R}}(X_1, \dots, X_n)$. This is a finite dimensional (closed) vect subspace of $L^2(\Omega, \sigma(X_1, \dots, X_n), \mathbb{P})$.
So, the orthogonal projection is well defined, and:
$\forall i, \ \mathbb{E}[(Y - \overline{Y}) X_i] = 0$.
Therefore, $Z = Y - \overline{Y}$ is independent from $(X_1, \dots, X_n)$.
For any bounded observable $F(X_1, \dots, X_n)$,

$$\mathbb{E}[Y F(X_1, \dots, X_n)] = \mathbb{E}[\underbrace{\overline{Y}}_{\sigma(X_1, \dots, X_n) \text{ measurable}} F(X_1, \dots, X_n)] + \underbrace{\mathbb{E}[Z F(X_1, \dots, X_n)]}_{= \mathbb{E}[Z]\mathbb{E}[F] = 0}$$

$\Longrightarrow \overline{Y} = \mathbb{E}[Y \mid X_1, \dots, X_n]$.

② Denote $Z = Y - \overline{Y}$. Since $\overline{Y}$ is the orthogonal projection of Y onto V,
$\mathrm{cov}(Z, X_i) = 0 \ \forall i$
$\Longrightarrow Z$ is independent from $\sigma(X_1, \dots, X_n)$.
So $Y = \overline{Y} + Z$, Z) centered Gaussian variable with variance σ^2.

Then, for any bounded observables $F(X_1, \dots, X_n)$, $G(Y)$,

$$\mathbb{E}[G(Y) F(X_1, \dots, X_n)] = \mathbb{E}[G(\sum_{i=1}^{n} \lambda_i X_i + Z) F(X_1, \dots, X_n)] \quad (\star)$$

<u>1st case</u>: $\sigma = 0 \iff Z = 0$ a.s. $\iff Y \in V$, and is $\sigma(X_1, \dots, X_n)$ measurable.

Then, $Y = \overline{Y}$ and its conditional distribution is $\delta_{\overline{Y}} = \mathcal{N}(\overline{Y}, 0)$.

<u>2nd case</u>: $\sigma > 0 \iff Y \notin V$.
Without loss of generality, we can assume the family $(X_i)_{i \in [\![1,n]\!]}$ free
$\iff$ if $\Sigma_n = \mathrm{cov}(X_i, X_j)_{i,j \leq n}$, then Σ_n is invertible

Then, the full covariance matrix Σ of $(X_1, \dots, X_n, Y)$ is also invertible,
and $\Sigma = \begin{pmatrix} \Sigma_n & A \\ A^t & \lambda \end{pmatrix}$ with $A = (\mathrm{cov}(X_i, Y))_{i \leq n}$
$\lambda = \mathrm{var}(Y)$.

We then have:

$$\star = \frac{1}{(2\pi)^{\frac{n+1}{2}}\sqrt{\det\Sigma}} \int_{\mathbb{R}^{n+1}} e^{-\frac{1}{2}(x,y)^t \Sigma^{-1} \binom{x}{y}} G(y) F(x) \, dx_1 \dots dx_n \, dy$$

Schur complement formula: Set $s = \lambda - A^t \Sigma_n^{-1} A$.

We have:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_n^{-1}(I_n + s^{-1} A A^t \Sigma_n^{-1}) & -s^{-1}\Sigma_n^{-1} A \\ -s^{-1} A^t \Sigma_n^{-1} & s^{-1} \end{pmatrix}$$

Therefore, the exponent is $-\frac{1}{2}\left(x^t \Sigma_n^{-1} x + s^{-1}(y - A^t \Sigma_n^{-1} x)^2 \right)$.

On the other hand, $\overline{Y} = A^t \Sigma_n^{-1} X$.

Indeed, write $\Sigma_n = H^2$ with H symmetric positive.

If $W = H^{-1} X$, then $W_1 \dots W_n$ is an orthonormal basis of V.

Therefore,

$$\overline{Y} = \sum_{i=1}^{n} \mathbb{E}[Y W_i] W_i = \sum_{i,j,k} \mathbb{E}[Y H^{-1}_{ji} X_j] H^{-1}_{ki} X_k$$

$$= \sum_{j,k} \mathbb{E}[Y (\Sigma_n^{-1})_{jk} X_j] X_k = \sum_{j,k} (\Sigma_n^{-1})_{jk} A_j X_k$$

$$= \sum_k (A^t \Sigma_n^{-1})_k X_k = A^t \Sigma_n^{-1} X.$$

Finally,

$$\sigma^2 = \mathbb{E}[(Y - \overline{Y})^2] = \mathbb{E}[Y^2] - \mathbb{E}[\overline{Y}^2]$$
$$= \lambda - A^t \Sigma_n^{-1} A = s.$$

So, making the change of variables $z = y - A^t \Sigma_n^{-1} x$, we get:

$$\star = \frac{1}{(2\pi)^{\frac{n+1}{2}}\sqrt{\det\Sigma_n}\sqrt{s}} \int_{\mathbb{R}^n} F(x) e^{-\frac{1}{2}x^t \Sigma_n^{-1} x} \underbrace{\left(\int_{\mathbb{R}} G(A^t \Sigma_n^{-1} x + z) e^{-\frac{1}{2s} z^2} dz \right)}_{\mathbb{E}[G(\mathcal{N}(A^t \Sigma_n^{-1} x, s))]}.$$

$\square$