Lesson n°1: Generalities on random variables

UCE, verano 2013

Contents

1 Random variables

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

1.1 First definitions

Definition 1 (Random variable).
A measurable application $X : (\Omega, \mathcal{F}) \longmapsto (E, \mathcal{E})$ is called a random variable. Its law $\mathbb{P}_X$ is defined by

$$\begin{array}{rcl} \mathbb{P}_X : \mathcal{E} & \longmapsto & [0,1] \\ A & \longmapsto & \mathbb{P}(X^{-1}(A)) \end{array}$$

Theorem 2 (Transport theorem).
Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longmapsto (E, \mathcal{E})$ be a random variable, and $\varphi : (E, \mathcal{E}) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a measurable function. If $\varphi(X)$ is $\mathbb{P}$ integrable:

$$\mathbb{E}[\varphi(X)] = \int_\Omega \varphi(X(\omega))\mathbb{P}(d\omega) = \int_E \varphi(x)\mathbb{P}_X(dx)$$

The two main examples of random variables are the discrete and absolute continuous cases.

Definition 3.

i) *A random variable X is called discrete if there exists a finite or countable set $\mathcal{S}$ such that $\mathbb{P}(X \in \mathcal{S}) = 1$. Assume that $\mathcal{S} = \{x_i,\ i \in I\}$ with $x_i \neq x_j$ for $i \neq j$. Then, the law of X is given by:*

$$\mathbb{P}_X = \sum_{i \in I} p_i \delta_{x_i}$$

where δ_{x_i} denotes Dirac measure at x_i and $p_i = \mathbb{P}(X = x_i)$.

ii) *A random variable X taking values in $\mathbb{R}^d$ is said to be absolutely continuous with respect to the Lebesgue measure if there exists a measurable function $f : \mathbb{R}^d \longrightarrow [0, +\infty]$ such that:*

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \qquad \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f(x)dx.$$

f is called the probability density function of X.

1.2 Characterization of laws in $\mathbb{R}^d$

Definition 4 (Characteristic function)**.**
Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longmapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a random variable. The characteristic function of X is defined by:

$$\Phi_X(t) = \mathbb{E}\left[e^{i\langle t,X\rangle}\right] = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} \mathbb{P}_X(dx), \qquad \forall t \in \mathbb{R}^d.$$

Definition 5 (Cumulative distribution function)**.**
Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longmapsto (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a random variable. The cumulative distribution function of $X = (X_1, \ldots, X_d)$ is defined by:

$$F_X(t_1, \ldots, t_d) = \mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d) = \mathbb{P}_X\left(\prod_{i=1}^{d}]-\infty, t_i]\right), \qquad \forall t = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$

These functions characterize the law of X in the following sense.

Theorem 6. *Let X and Y be two $\mathbb{R}^d$-valued random variables. The following assertions are equivalent:*

i) *X and Y have the same law,*

ii) *$\Phi_X = \Phi_Y$.*

iii) *$F_X = F_Y$,*

Be careful that the equality in law $X \overset{\text{(law)}}{=} Y$ does not mean that X and Y are a.s. equal. Indeed, if X follows a uniform law on $[0, 1]$, then $Y = 1 - X$ also follows a uniform law on $[0, 1]$ so their characteristic functions and cumulative distribution functions are equal, but of course, X is not equal to Y a.s.

1.3 Independence

Definition 7. *Let $n \in \mathbb{N}^*$. The $\mathbb{R}^d$-valued random variables $X_1, \ldots, X_n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if*

$$\forall A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^d), \qquad \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i).$$

This may be written shortly:

$$\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}.$$

The independence between random variables may be directly seen on the characteristic functions.

Theorem 8. *Let $X_1, \dots, X_n$ be n random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that X_i is $\mathbb{R}^{d_i}$-valued. Then the random variables $X_1, \dots, X_n$ are mutually independent if and only if*

$$\Phi_{(X_1,\dots,X_n)}(t_1, \dots, t_n) = \prod_{i=1}^{n} \Phi_{X_i}(t_i), \qquad \forall (t_1, \dots, t_n) \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_n}.$$

Proof. By definition, there is the equivalence :

$$X_1, \dots, X_n \text{ are mutually independent} \quad \Longleftrightarrow \quad \mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

Since Fourier's transform is injective, this is also equivalent to :

$$X_1, \dots, X_n \text{ are mutually independent} \quad \Longleftrightarrow \quad \mathcal{F}\left(\mathbb{P}_{(X_1,\dots,X_n)}\right) = \mathcal{F}\left(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}\right),$$

hence the result follows from:

$$\Phi_{(X_1,\dots,X_n)} = \mathcal{F}\left(\mathbb{P}_{(X_1,\dots,X_n)}\right) = \mathcal{F}\left(\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}\right) = \prod_{i=1}^{n} \mathcal{F}\left(\mathbb{P}_{X_i}\right) = \prod_{i=1}^{n} \Phi_{X_i}.$$

■

Example 9. Assume for instance that X and Y are two independent $\mathbb{R}$-valued random variables with respective probability density functions f_X and f_Y. Then, for any $\lambda \in \mathbb{R}$:

$$\begin{aligned}\mathbb{E}\left[e^{i\lambda(X+Y)}\right] = \mathbb{E}\left[e^{i\lambda X}\right]\mathbb{E}\left[e^{i\lambda Y}\right] &= \int_{\mathbb{R}} e^{i\lambda t} f_X(t)dt \int_{\mathbb{R}} e^{i\lambda t} f_Y(t)dt \\ &= \int_{\mathbb{R}} e^{i\lambda t} \left(\int_{\mathbb{R}} f_X(t-s) f_Y(s) ds\right) dt\end{aligned}$$

which proves that the random variable $X + Y$ is also an absolutely continuous random variable and that its probability density function is given by:

$$f_{X+Y}(t) = \int_{\mathbb{R}} f_X(t-s) f_Y(s) ds.$$

2 Gaussian variables

2.1 Real-valued Gaussian random variables

Definition 10 (Gaussian random variables)**.** *A random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is Gaussian with mean m and variance $\sigma^2 > 0$ if its probability law admits the density function:*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

We shall write $X \sim \mathcal{N}(m, \sigma^2)$.

Remark 11. A random variable G which follows the law $\mathcal{N}(0, 1)$ is called a standard Gaussian random variable, and, for any $m \in \mathbb{R}$ and $\sigma > 0$,

$$m + \sigma G \sim \mathcal{N}(m, \sigma^2).$$

Therefore, in most situations, it is enough to make the computations with a standard Gaussian random variable, and the general case follows from this relation.

Proposition 12 (Characteristic function)**.** *If $X \sim \mathcal{N}(m, \sigma^2)$, its characteristic function is given by:*

$$\mathbb{E}\left[e^{itX}\right] = \exp\left(imt - \frac{\sigma^2 t^2}{2}\right).$$

Proof. Assume first that $G \sim \mathcal{N}(0,1)$. We want to compute :

$$\mathbb{E}\left[e^{itG}\right] = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \cos(tx) dx + i \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \sin(tx) dx$$

Observe first the imaginary part of this expression is null, as the integral of an odd function on an interval symmetric with respect to 0. Next, we set:

$$\Phi(t) = \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \cos(tx) dx$$

Since

$$\left| e^{-\frac{x^2}{2}} x \sin(tx) \right| \leq |x| e^{-\frac{x^2}{2}}$$

which is integrable, we may apply Leibniz integral rule (differentiation under the integral sign) to obtain:

$$\Phi'(t) = -\int_{\mathbb{R}} e^{-\frac{x^2}{2}} x \sin(tx) dx,$$

and integrating by part this last expression:

$$\Phi'(t) = -t \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \cos(tx) dx = -t\Phi(t).$$

Therefore, there exists a constant $k \in \mathbb{R}$ such that:

$$\mathbb{E}\left[e^{itG}\right] = \frac{k}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

and taking $t = 0$, we deduce that $k = \sqrt{2\pi}$. Finally, if $X \sim \mathcal{N}(m, \sigma^2)$, the general expression follows easily from:

$$\mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[e^{it(m+\sigma G)}\right] = e^{itm} \mathbb{E}\left[e^{it\sigma G}\right] = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

■

Remark 13. In particular, thanks to the Taylor series of the exponential function, it is easily seen that the moments of a standard Gaussian random variable G are given by:

$$\begin{cases} \mathbb{E}\left[G^{2n+1}\right] = 0 \\ \mathbb{E}\left[G^{2n}\right] = \dfrac{(2n)!}{2^n n!} \end{cases}$$

Proposition 14 (Sum of independent Gaussian r.v.'s)**.** *Let $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ be two independent Gaussian random variables. Then, $X_1 + X_2$ is a Gaussian random variable with law $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Proof. Since X_1 and X_2 are independent, we may write, for every $t \in \mathbb{R}$:

$$\begin{aligned} \mathbb{E}\left[e^{it(X_1+X_2)}\right] &= \mathbb{E}\left[e^{itX_1}\right] \mathbb{E}\left[e^{itX_2}\right] \\ &= \exp\left(itm_1 - \frac{t^2\sigma_1^2}{2}\right) \exp\left(itm_2 - \frac{t^2\sigma_2^2}{2}\right) \\ &= \exp\left(it(m_1 + m_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right). \end{aligned}$$

■

2.2 Gaussian random vectors

Definition 15 (Gaussian random vectors)**.** *A random vector $X = (X_1, \ldots, X_n)$ taking values in $\mathbb{R}^n$ is said to be Gaussian if, for any $\lambda \in \mathbb{R}^n$, the random variable*

$$< \lambda, X >= \sum_{i=1}^{n} \lambda_i X_i \qquad \textit{is a Gaussian random variable.}$$

Remark 16.

a) It is clear that if X is a Gaussian vector, then each of its components X_i is a Gaussian random variable.

b) However, the converse is not true ! Take for instance:

$$X = \begin{pmatrix} G \\ \varepsilon G \end{pmatrix}$$

where G is a standard Gaussian r.v. and ε is an independent Rademacher variable, i.e.

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}.$$

Both components of X are Gaussian random variables, but

$$X_1 + X_2 = G + \varepsilon G = (1+\varepsilon)G$$

is not a Gaussian random variable since $\mathbb{P}((1+\varepsilon)G = 0) = \frac{1}{2}$.

c) Of course, if $X_1, \dots, X_n$ are independent Gaussian random variables, from Proposition 14, $X = (X_1, \dots, X_n)$ is a Gaussian random vector.

Definition 17. *The covariance matrix of a $\mathbb{R}^n$-valued random vector X is the matrix :*

$$K = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^*\right],$$

where $*$ denotes transposition. In particular, the components $(K_{i,j})_{1 \le i,j \le n}$ are given by :

$$K_{i,j} = \operatorname{cov}(X_i, X_j) = \mathbb{E}\left[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right].$$

This matrix is symmetric and positive.

Proof. The fact that K is symmetric is obvious from the definition. To show that K is positive, observe that for any vector $u \in \mathbb{R}^n$,

$$u^* K u = \mathbb{E}\left[u^*(X - \mathbb{E}[X])(X - \mathbb{E}[X])^* u\right] = \mathbb{E}\left[\left(u^*(X - \mathbb{E}[X])\right)^2\right] \ge 0.$$

■

Proposition 18. *Let K be the covariance matrix of a Gaussian random vector. Then, for every $u \in \mathbb{R}^n$,*

$$\mathbb{E}\left[\exp\left(i\langle u, X\rangle\right)\right] = \exp\left(i\langle u, \mathbb{E}[X]\rangle - \frac{1}{2}u^* K u\right)$$

Proof. By definition, the random variable $Z = \langle u, X\rangle = u^* X$ is Gaussian, with expectation $\mathbb{E}[Z] = \langle u, \mathbb{E}[X]\rangle = u^*\mathbb{E}[X]$ and variance $\operatorname{Var}(Z) = u^* K u$, hence the result is a direct consequence of Proposition 12.

■

In particular, to characterize a Gaussian random vector, we only need its expectation and covariance matrix. The converse is also true thanks to the following result.

Theorem 19. *Let $m \in \mathbb{R}^n$ and Γ be a symmetric and positive matrix of order n. Then, their exists a Gaussian random vector with expectation m and covariance matrix Γ.*

Proof. Observe first that there exists a matrix A such that

$$\Gamma = AA^*.$$

Indeed, since Γ is a symmetric and real matrix, it may be diagonalized in an orthonormal basis, i.e. there exists an orthogonal matrix P such that $D = P^*\Gamma P$ is a diagonal matrix. Since Γ is positive, the terms $\lambda_1, \dots, \lambda_n$ on the diagonal of D are all positive, so we may consider the diagonal matrix Δ whose terms on the diagonal are $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$, and A is finally given

by $A = P\Delta$.
Now, let $G_1, \dots, G_n$ be n independent standard Gaussian random variables, and define

$$X = m + AG \qquad \text{with} \quad G = \begin{pmatrix} G_1 \\ \vdots \\ G_n \end{pmatrix}.$$

Since the random variables $(G_i)_{1 \le i \le n}$ are independent, X is a Gaussian random vector. Its expectation is given by

$$\mathbb{E}[X] = m + A\mathbb{E}[G] = m$$

and its covariance matrix by:

$$K = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^*\right] = \mathbb{E}\left[(AG)(AG)^*\right] = A\mathbb{E}[GG^*]A^* = AI_nA^* = \Gamma.$$

■

Theorem 20. *Let $X = (X_1, \dots, X_n)^*$ be a Gaussian random vector. Then, the random variables $X_1, \dots, X_n$ are independent if and only if the covariance matrix is diagonal.*

Proof. It is clear that if $X_1, \dots, X_n$ are independent, then $\text{cov}(X_i, X_j) = 0$ when $i \neq j$, hence K is diagonal. Assume now that K is diagonal. In particular, for any $u \in \mathbb{R}^n$,

$$u^*Ku = \sum_{j=1}^n u_j^2 K_{j,j} = \sum_{j=1}^n u_j^2 \text{Var}(X_j),$$

so the characteristic function of X reads:

$$\begin{aligned}
\mathbb{E}\left[e^{i\langle u, X\rangle}\right] &= \exp\left(i\langle u, \mathbb{E}[X]\rangle - \frac{1}{2}u^*Ku\right) \\
&= \exp\left(i\sum_{j=1}^n u_j\mathbb{E}[X_j] - \frac{1}{2}\sum_{j=1}^n u_j^2\text{Var}(X_j)\right) \\
&= \prod_{j=1}^n \exp\left(iu_j\mathbb{E}[X_j] - \frac{1}{2}u_j^2\text{Var}(X_j)\right) \\
&= \prod_{j=1}^n \mathbb{E}\left[e^{iu_jX_j}\right]
\end{aligned}$$

which implies, from Theorem 8 that the random variables $X_1, \dots, X_n$ are independent.

■

Remark 21. In particular, if (X, Y) is a Gaussian vector, then X and Y are independent if and only if their covariance matrix is null :

$$K = \mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])^*] = 0.$$

We must insist that this is no longer the case if X and Y are only Gaussian random variables. Indeed, if we take back the example $X = G$ and $Y = \varepsilon G$ with G a standard Gaussian r.v. and ε an independent Rademacher variable, then

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] = \mathbb{E}[\varepsilon G^2] = \mathbb{E}[\varepsilon]\mathbb{E}[G^2] = 0$$

but X and Y are obviously not independent since $|X| = |Y|$.

Theorem 22 (Density of a Gaussian vector)**.**
Let X be a $\mathbb{R}^n$-valued Gaussian vector with covariance matrix K.

1. *X admits a density if and only if K is invertible*
2. *If K is invertible, the density of X is given by:*

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(K)}} \exp\left(-\frac{1}{2}(x - m)^*K^{-1}(x - m)\right), \qquad x \in \mathbb{R}^n,$$

with $m = \mathbb{E}[X]$.

Proof.

a) Observe first that if F is a subspace of $\mathbb{R}^n$ with dimension strictly smaller than n and Z is a random vector which admits a density f, then $\mathbb{P}(Z \in F) = 0$. Indeed, if H is a hyperplane which contains F, say $H = \{(x_1, \dots, x_n) \in \mathbb{R}^n;\ x_n = 0\}$, then

$$\mathbb{P}(Z \in F) \leq \mathbb{P}(Z \in H) = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) 1_{\{x_n=0\}} dx_1 \dots dx_n = 0.$$

b) Next, we know that X may be written

$$X = m + AG$$

where $m = \mathbb{E}[X]$, G is a Gaussian vector with independent components whose laws are $\mathcal{N}(0,1)$ and $AA^* = K$. If A is not invertible, then the range of A is strictly included in $\mathbb{R}^n$, and X cannot admit a density. This proves Point 1) since $\det(K) = \det^2(A)$, i.e. A is invertible if and only if K is.

c) Assume now that K is invertible. The density of G is given by

$$\mathbb{P}(G \in dy) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} y^* y\right) dy, \qquad y \in \mathbb{R}^n,$$

and the expression for the density of X follows from the change of variable :

$$y = A^{-1}(x - m).$$

■

Theorem 23 (Central Limit Theorem)**.**
Let $(X_n, n \geq 1)$ be a sequence of random vectors in $\mathbb{R}^d$, independent and identically distributed. We assume that all these variables are in $L^2(\Omega)$, and we denote by m their expectation and by Γ the covariance matrix of X_1. Then:

$$\frac{X_1 + \ldots + X_n - nm}{\sqrt{n}} \xrightarrow[n \to +\infty]{\text{(law)}} \mathcal{N}(0, \Gamma).$$

Proof. We first translate the problem and set $Z_i = X_i - m$ in order to work with centered r.v.'s. Then

$$T_n = \frac{X_1 + \ldots + X_n - nm}{\sqrt{n}} = \frac{Z_1 + \ldots + Z_n}{\sqrt{n}},$$

so the characteristic function of T_n reads:

$$\mathbb{E}\left[e^{i\langle u, T_n \rangle}\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{i\langle \frac{u}{\sqrt{n}}, Z_j \rangle}\right] = \left(\mathbb{E}\left[e^{i\langle \frac{u}{\sqrt{n}}, Z_1 \rangle}\right]\right)^n.$$

Now, since Z_1 admits a finite moment of order 2, we may use Taylor's theorem and write

$$\mathbb{E}\left[e^{i\langle t, Z_1 \rangle}\right] \underset{t \to 0}{=} 1 - \frac{1}{2} t^* \Gamma t + \mathrm{o}\left(|t|^2\right)$$

so that, as $n \to +\infty$

$$\mathbb{E}\left[e^{i\langle u, T_n \rangle}\right] = \left(1 - \frac{1}{2}\left(\frac{u}{\sqrt{n}}\right)^* \Gamma \left(\frac{u}{\sqrt{n}}\right) + \mathrm{o}\left(\frac{|u|^2}{n}\right)\right)^n \xrightarrow[n \to +\infty]{} \exp\left(-\frac{1}{2} u^* \Gamma u\right).$$

■

3 Conditional expectation

3.1 Definition

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. In the following, we shall assume that the σ-algebras we are dealing with are complete, that is to say that they contain all the negligible sets (i.e. the sets A such that $\mathbb{P}(A) = 0$).

We first recall the following result on Hilbert space. Let H be a Hilbert space and F be a closed subspace of H. For every $x \in H$, there exists a unique $y \in F$, called the orthogonal projection of x on F, which satisfies one of the two following equivalent assertions:

i) $\forall z \in F, \qquad <x-y, z>=0,$

ii) $\forall z \in F, \qquad \|x-y\| \leq \|x-z\|.$

When applied to the Hilbert space $L^2(\Omega, \mathcal{A}, \mathbb{P})$ and the closed subspace $L^2(\Omega, \mathcal{B}, \mathbb{P})$, the precedent result gives the following characterization of conditional expectation.

Proposition 24. *For every random variable $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, there exists an a.s. unique random variable Y such that*

$$\begin{cases} Y \in L^2(\Omega, \mathcal{B}, \mathbb{P}) \\ \mathbb{E}[XZ] = \mathbb{E}[YZ], \qquad \forall Z \in L^2(\Omega, \mathcal{B}, \mathbb{P}). \end{cases} \tag{1}$$

We denote this random variable by $Y = \mathbb{E}[X|\mathcal{B}]$.

Example 25. Take $\mathcal{B} = \{\emptyset, \Omega\}$. The random variables which are $\mathcal{B}$-measurable are a.s. constant, hence, $\mathbb{E}[X|\mathcal{B}] = a$. If Z is $\mathcal{B}$-measurable, then $Z = z$, and Equation (1) yields

$$\mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[za] = za = \mathbb{E}[zX] = z\mathbb{E}[X], \qquad \forall z \in \mathbb{R}$$

hence $a = \mathbb{E}[X]$. So the conditional expectation of a random variable X with respect to the trivial σ-algebra is simply its classical expectation.

Remark 26. If X is a positive and bounded random variable, then $\mathbb{E}[X|\mathcal{B}] \geq 0$ a.s. Indeed, set $Y = \mathbb{E}[X|\mathcal{B}]$ and assume that $\mathbb{P}(Y < 0) > 0$. In particular, for n large enough, the set $A = \{Y < -\frac{1}{n}\}$ has a strictly positive probability. Since 1_A is a bounded $\mathcal{B}$-measurable random variable, it holds

$$0 \leq \mathbb{E}[X1_A] = \mathbb{E}[Y1_A] \leq -\frac{1}{n}\mathbb{P}(A) < 0,$$

which contradicts the assumption that $\mathbb{P}(Y < 0) > 0$.

We now extend the previous construction to any random variable $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Theorem 27. *Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. There exists an a.s. unique and integrable random variable $\mathbb{E}[X|\mathcal{B}]$ such that:*

$$\begin{cases} \mathbb{E}[X|\mathcal{B}] \in L^1(\Omega, \mathcal{B}, \mathbb{P}) \\ \mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X|\mathcal{B}]], \quad \textit{for every } \mathcal{B}\textit{-measurable and bounded r.v. } Z. \end{cases}$$

Proof.

a) The basic idea of the proof is to work by truncation. We define, for any real $a \in \mathbb{R}$:

$$a^+ = \sup(a, 0) \qquad \text{and} \qquad a^- = \sup(0, -a)$$

Observe first that, by splitting X as $X = X^+ - X^-$, we may reduce our study to positive random variables. So assume now that X is $\mathbb{R}^+$-valued, and define $X_n = X \wedge n$. Since each X_n belongs to L^2 (as a bounded r.v.), we can choose a version of the conditional expectation $Y_n = \mathbb{E}[X_n|\mathcal{B}]$. Furthermore, as the sequence $(X_n, n \in \mathbb{N})$ is positive and increasing, from Remark 26 so is the sequence $(Y_n, n \in \mathbb{N})$, and we set:

$$Y(\omega) := \limsup_{n \to +\infty} Y_n(\omega).$$

Since Y_n is $\mathcal{B}$-measurable for every n, Y is also $\mathcal{B}$-measurable. Take a positive $\mathcal{B}$-measurable r.v. Z. By the monotone convergence theorem, passing to the limit in the equality $\mathbb{E}[ZX_n] = \mathbb{E}[ZY_n]$ we deduce that $\mathbb{E}[ZX] = \mathbb{E}[ZY]$. Taking $Z = 1$, we finally conclude that Y is indeed integrable.

b) To prove the uniqueness, assume that Y and $\widetilde{Y}$ are two versions of $\mathbb{E}[X|\mathcal{B}]$ such that $\mathbb{P}(Y > \widetilde{Y}) > 0$. In particular, for n large enough, the set $A = \{Y - \widetilde{Y} > \frac{1}{n}\}$ has a strictly positive probability. Since 1_A is a bounded $\mathcal{B}$-measurable random variable, it holds

$$0 = \mathbb{E}\left[(Y - \widetilde{Y})1_A\right] \geq \frac{1}{n}\mathbb{P}(A) > 0$$

which is a contradiction.

■

3.2 Properties

We list below the main properties of conditional expectation.

Theorem 28. *Let X and Y be two integrable random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Then:*

i) (Linearity) $\mathbb{E}[aX + bY|\mathcal{B}] = a\mathbb{E}[X|\mathcal{B}] + b\mathbb{E}[Y|\mathcal{B}]$

ii) (Positivity) if $X \geq 0$ *a.s., then* $\mathbb{E}[X|\mathcal{B}] \geq 0$ *a.s.*

iii) If X *is* $\mathcal{B}$*-measurable:* $\mathbb{E}[X|\mathcal{B}] = X$ *a.s.*

iv) More generally, if Y *is* $\mathcal{B}$*-measurable and such that* XY *is integrable, then:* $\mathbb{E}[XY|\mathcal{B}] = Y\mathbb{E}[X|\mathcal{B}]$.

v) (Tower property) If $\mathcal{C}$ *is a sub-*σ*-algebra of* $\mathcal{B}$*, then* $\mathbb{E}\big[\mathbb{E}[X|\mathcal{B}]|\mathcal{C}\big] = \mathbb{E}[X|\mathcal{C}]$.

vi) Fatou's lemma, the monotone convergence theorem, the dominated convergence theorem and Jensen inequality hold with conditional expectation.

In practice, it is generally difficult to compute a conditional expectation given a σ-algebra $\mathcal{B}$. One situation in which this task is easier is when the σ-algebra $\mathcal{B}$ is generated by a r.v. $T : (\Omega, \mathcal{A}) \longrightarrow (E, \mathcal{E})$. The σ-algebra generated by T is denoted by $\sigma(T)$ and defined by:

$$\sigma(T) = \{A \in \mathcal{A},\ \exists C \in \mathcal{E}, A = T^{-1}(C)\}.$$

A real-valued random variables $X : (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be $\sigma(T)$-measurable if for every $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \in \sigma(T)$. Such an application is characterized by:

X is $\sigma(T)$-measurable $\iff$ There exists a measurable function $f : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $X = f(T)$.

Example 29. Let (X, Y) be a centered Gaussian random vector, with Y not degenerated. Then :

$$\mathbb{E}[X|Y] = \left(\frac{\mathbb{E}[XY]}{\mathrm{Var}(Y)}\right) Y.$$

To prove this result, we shall look for $a \in \mathbb{R}$ such that the two Gaussian random variables $X - aY$ and Y are independent. By Theorem 20, these two random variables are independent if and only if:

$$\mathrm{cov}(X - aY, Y) = 0$$

that is,

$$\mathbb{E}[(X - aY)Y] = \mathbb{E}[XY] - a\mathbb{E}[Y^2] = \mathbb{E}[XY] - a\mathrm{Var}(Y) = 0.$$

Since Y is not degenerated, $\mathrm{Var}(Y) > 0$ so a is unique and given by

$$a = \frac{\mathbb{E}[XY]}{\mathrm{Var}(Y)}.$$

But, by independence,

$$\mathbb{E}[X - aY|Y] = \mathbb{E}[X - aY] = 0$$

and by linearity,

$$\mathbb{E}[X - aY|Y] = \mathbb{E}[X|Y] - a\mathbb{E}[Y|Y] = \mathbb{E}[X|Y] - aY = 0,$$

which proves the announced result.

3.3 Conditional laws

Definition 30. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. A kernel $N : E \times F \longrightarrow \mathbb{R}^+$ is a transition probability from E to F if :*

i) for every $x \in E$*, the application* $A \in \mathcal{F} \longmapsto N(x, A)$ *is a probability on* $(F, \mathcal{F})$.

ii) for every $A \in \mathcal{F}$*, the application* $x \in E \longmapsto N(x, A)$ *is measurable from* $(E, \mathcal{E})$ *on* $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$.

In other words, a transition probability is a measurable family of probabilities $(N(x, \cdot),\ x \in E)$ on $(F, \mathcal{F})$ indexed by the set E.

Definition 31. *Let X and Y be two r.v.'s taking values respectively in $(E, \mathcal{E})$ and $(F, \mathcal{F})$. The conditional law of Y given X is a transition probability N from E to F such that, for every measurable and positive function φ:*

$$\mathbb{E}[\varphi(Y)|X] = \int_\Omega \varphi(y) N(X, dy).$$

Remark 32.

a) As for conditional expectation, the conditional law of Y given X is only determined up to a set of null $\mathbb{P}_X$-measure.

b) Heuristically, $N(x, dy)$ denotes the law of Y given that $X = x$.

c) Of course, if X and Y are independent, then $N(x, dy) = \mathbb{P}_Y(dy)$.

Proposition 33. *Let (X, Y) be a pair of $\mathbb{R}$-valued random variables, whose joint density is given by $f(x, y)$. Then the law of the r.v. Y conditionally to X is given by:*

$$N(x, dy) = \frac{f(x, y)}{\alpha(x)} 1_{\{\alpha(x)>0\}} dy$$

where

$$\alpha(x) = \int_\mathbb{R} f(x, y) dy$$

is the density of the r.v. X.

Proof. Let h and g be two measurable and bounded functions. By definition:

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

By taking $h(x) = 1_{\{\alpha(x)=0\}}$ and applying Fubini's theorem, we obtain

$$\mathbb{P}(\alpha(X) = 0) = \iint_{\mathbb{R}^2} 1_{\{\alpha(x)=0\}} f(x, y)dxdy = \int_\mathbb{R} 1_{\{\alpha(x)=0\}} \left(\int_\mathbb{R} f(x, y)dy \right) dx = \int_\mathbb{R} 1_{\{\alpha(x)=0\}} \alpha(x)dx = 0.$$

Therefore, going back to the general expression and applying Fubini's theorem again :

$$\begin{aligned}
\mathbb{E}[h(X)g(Y)] &= \mathbb{E}[h(X)g(Y)1_{\{\alpha(X)>0\}}] \\
&= \int_\mathbb{R} h(x)\alpha(x) \left(\int_\mathbb{R} g(y) \frac{f(x, y)}{\alpha(x)} 1_{\{\alpha(x)>0\}} dy \right) dx \\
&= \int_\mathbb{R} h(x)\alpha(x) \left(\int_\mathbb{R} g(y)N(x, dy) \right) dx \\
&= \mathbb{E}\left[h(X) \left(\int_\mathbb{R} g(y)N(X, dy) \right) \right]
\end{aligned}$$

hence, by definition of the conditional expectation:

$$\mathbb{E}[g(Y)|X] = \int_\mathbb{R} g(y)N(X, dy).$$

It remains to justify that N is indeed a transition probability, but this is immediate since :

$$\int_\mathbb{R} N(x, dy) = 1_{\{\alpha(x)>0\}} \overset{(\mathbb{P}_X-\text{a.s.})}{=} 1.$$

■

3.4 The Gaussian case

We now turn our attention back to Gaussian vectors. Let X (resp. Y) be a $\mathbb{R}^n$ (resp. $\mathbb{R}^d$)-valued Gaussian random vector and assume that X admits a density. In this section, we want to compute the law of Y conditionally to X. We use the following notation:

$$\begin{cases}
K_{11} \text{ denotes the covariance matrix of } X, \text{ of order } n, \\
K_{22} \text{ denotes the covariance matrix of } Y, \text{ of order } d, \\
K_{12} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^*] \text{ is a matrix of order } n \times d \text{ and } K_{21} = K_{12}^*.
\end{cases}$$

Proposition 34. *Let N be the conditional law of Y given X. Then, $N(x, \cdot)$ is the Gaussian law $\mathcal{N}(Ax+m, \Gamma)$ where $A = K_{21}K_{11}^{-1}$, $m = \mathbb{E}[Y] - A\mathbb{E}[X]$ and $\Gamma = K_{22} - K_{21}K_{11}^{-1}K_{21}^*$.*

Proof.
Consider the Gaussian vector $Z = Y - AX$. We claim that Z is independent from X. Indeed, the covariance matrix of Z and X is given by:

$$\begin{aligned}\mathbb{E}[(Z-\mathbb{E}[Z])(X-\mathbb{E}[X])^*] &= \mathbb{E}[(Y - AX - \mathbb{E}[Y] + A\mathbb{E}[X])(X-\mathbb{E}[X])^*] \\ &= \mathbb{E}[(Y-\mathbb{E}[Y])(X-\mathbb{E}[X])^*] + A\mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^*] \\ &= K_{21} - AK_{11} = 0\end{aligned}$$

by definition of A. Now, we may write :

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)g(AX+Z)] = \mathbb{E}[\mathbb{E}[f(X)g(AX+Z)|X]] = \mathbb{E}[f(X)\mathbb{E}[g(AX+Z)|X]],$$

but, since Z is independent from X,

$$\mathbb{E}[g(AX+Z)|X] = G(X)$$

where G is given by

$$G(x) = \mathbb{E}[g(Ax+Z)] = \int_{\mathbb{R}^d} g(y)N(x,dy)$$

with $N(x, \cdot)$ the law of the random vector $Ax+Z$, i.e. the Gaussian law with expectation $\mathbb{E}[Ax+Z] = Ax + \mathbb{E}[Z] = Ax+m$ and covariance matrix

$$\mathbb{E}[(Ax+Z-(Ax+m))(Ax+Z-(Ax+m))^*] = \mathbb{E}[(Z-m)(Z-m)^*] = \mathbb{E}[(Y-\mathbb{E}[Y]-AX+A\mathbb{E}[X])(Y-\mathbb{E}[Y]-AX+A\mathbb{E}[X])^*].$$

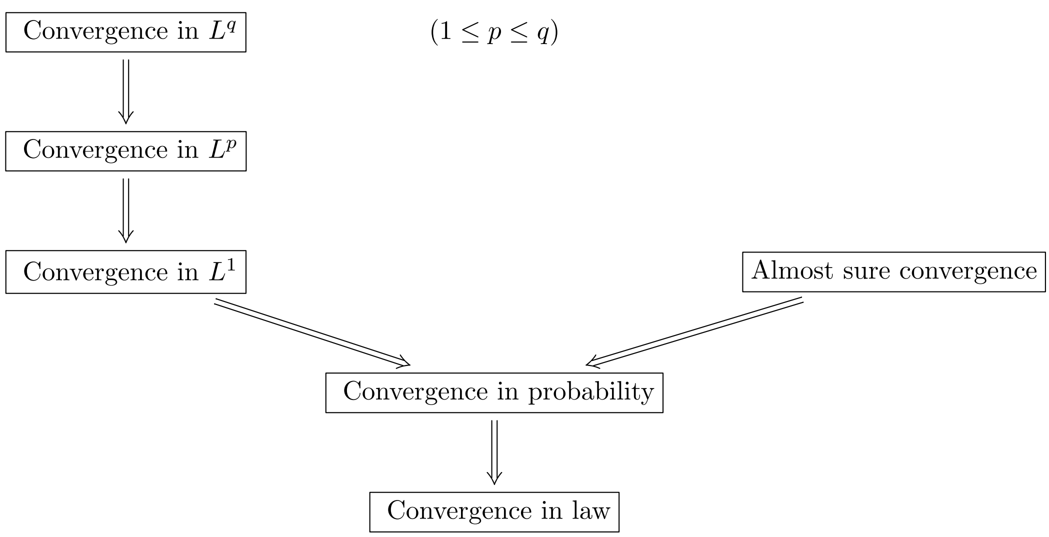
To simplify the notation, we set $X_0 = X - \mathbb{E}[X]$ and $Y_0 = Y_0 - \mathbb{E}[Y]$. Then:

$$\begin{aligned}\mathbb{E}[(Y_0 - AX_0)(Y_0-AX_0)^*] &= \mathbb{E}[(Y_0 - AX_0)(Y_0^* - X_0^*A^*)] \\ &= \mathbb{E}[Y_0Y_0^*] - \mathbb{E}[Y_0X_0^*]A^* - A\mathbb{E}[X_0Y_0^*] + A\mathbb{E}[X_0X_0^*]A^* \\ &= K_{22} - K_{21}A^* - AK_{12} + AK_{11}A^* \\ &= K_{22} - AK_{12} \qquad\qquad (\text{since } K_{21} = AK_{11}) \\ &= K_{22} - K_{21}K_{11}^{-1}K_{21}^*.\end{aligned}$$

■

4 Stochastic convergences

We finally conclude this first lesson by a short section on the different modes of convergence we shall use in the sequel, the general pattern being as follows:



4.1 Convergence in law

Definition 35. *A sequence of random vectors $(X_n)_{n\geq 1}$ converges in law towards a random vector X if for every continuous and bounded function $\varphi \in \mathcal{C}_0(\mathbb{R}^d)$:*

$$\lim_{n\to+\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)]. \qquad (2)$$

We shall denote :

$$X_n \xrightarrow[n\to+\infty]{\text{(law)}} X$$

The convergence in law, as its name indicates, does not depend on the random variable X, but rather on its law $\mathbb{P}_X = \mu$. In other words, the convergence in law is actually a convergence of measures: if μ_n denotes the law of the random variable X_n, then (2) may be rewritten:

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f(x)\mu_n(dx) = \int_{\mathbb{R}^d} f(x)\mu(dx).$$

In practice, we may restrict our attention to the family of functions

$$\varphi_u(x) = \exp\left(i\langle u, x\rangle\right), \qquad u \in \mathbb{R}^d,$$

thanks to Theorem 6.

Theorem 36. *The sequence of random vectors $(X_n)_{n\geq 1}$ converges in law towards X if and only if the sequence of characteristic functions Φ_{X_n} converges pointwise towards the characteristic function of X:*

$$X_n \xrightarrow[n\to+\infty]{\text{(law)}} X \qquad \Longleftrightarrow \qquad \Phi_{X_n}(u) \xrightarrow[n\to+\infty]{} \Phi_X(u) \quad \textit{for every } u \in \mathbb{R}^d.$$

4.2 Convergence in probability

Definition 37. *A sequence of random vectors $(X_n)_{n\geq 1}$ converges in probability towards a random vector X if for every $\varepsilon > 0$:*

$$\lim_{n\to+\infty} \mathbb{P}(\|X_n - X\| > \varepsilon) = 0.$$

We shall denote :

$$X_n \xrightarrow[n\to+\infty]{\text{(prob)}} X.$$

The limit in probability of a sequence of random vectors is almost surely unique.

Proposition 38. *If a sequence of random vectors $(X_n)_{n\geq 1}$ converges in probability towards a random vector X and towards a random vector Y, then:*

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

Proof. For every $\varepsilon > 0$:

$$\{\|X - Y\| > \varepsilon\} \subset \left\{\|X - X_n\| > \frac{\varepsilon}{2}\right\} \cup \left\{\|X_n - Y\| > \frac{\varepsilon}{2}\right\}$$

hence,

$$\mathbb{P}\left(\|X - Y\| > \varepsilon\right) \leq \mathbb{P}\left(\|X - X_n\| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(\|X_n - Y\| > \frac{\varepsilon}{2}\right).$$

Letting n tend to $+\infty$, we deduce that

$$\forall \varepsilon > 0, \qquad \mathbb{P}\left(\|X - Y\| > \varepsilon\right) = 0$$

and, letting then ε tend to 0, the monotone convergence theorem yields

$$\mathbb{P}\left(\|X - Y\| > 0\right) = 0,$$

which means that X and Y are equal a.s.

■

4.3 Almost sure convergence

Definition 39. *A sequence of random vectors $(X_n)_{n\geq 1}$ converges almost surely towards a random vector X if there exists a negligible set N such that, for every $\omega \notin N$, the numerical sequence $X_n(\omega)$ converges towards $X(\omega)$:*

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega) \qquad \text{for every } \omega \notin N.$$

We shall denote :

$$X_n \xrightarrow[n\to+\infty]{\text{(a.s)}} X.$$

4.4 Convergence in L^p

Definition 40. *Let $p \geq 1$. A sequence of random vectors $(X_n)_{n\geq 1}$ in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ converges in L^p towards a random vector $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ if :*

$$\lim_{n\to+\infty} \mathbb{E}\left[\|X_n - X\|^p\right] = 0$$

We shall denote :

$$X_n \xrightarrow[n\to+\infty]{L^p} X.$$

4.5 The weak law of large numbers

Theorem 41. *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with finite moment of order 2. Then:*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \xrightarrow[n\to+\infty]{\text{(prob)}} \quad \mathbb{E}[X_1].$$

Proof. By independence and scaling, we have

$$\text{Var}\left(\overline{X}_n\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\text{Var}(X_1)}{n}.$$

But, by definition of the variance:

$$\text{Var}\left(\overline{X}_n\right) = \mathbb{E}\left[|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]|^2\right] = \mathbb{E}\left[|\overline{X}_n - \mathbb{E}[X_1]|^2\right] \xrightarrow[n\to+\infty]{} 0,$$

which means that

$$\overline{X}_n \quad \xrightarrow[n\to+\infty]{L^2} \quad \mathbb{E}[X_1]$$

and the result follows since convergence in L^2 implies convergence in probability.

■