Quelques rappels sur la notion de conditionnement

1 Espérance conditionnelle

1.1 Un exemple

Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilité et soient $X : \Omega \to \{x_1, \ldots, x_m\}$, et $Y : \Omega \to \{y_1, \ldots, y_n\}$ deux variables aléatoires discrètes. La formule :

$$\begin{cases} \mathscr{F} & \to & [0,1] \\ A & \mapsto & \mathbb{P}(A \mid Y = y_j) = \frac{\mathbb{P}(A \cap \{Y = y_j\})}{\mathbb{P}(Y = y_j)} \end{cases}$$

définit une nouvelle probabilité sur $(\Omega, \mathscr{F})$, notée $\mathbb{P}^{Y=y_j}$.

Définition 1. *L'espérance conditionnelle de X sachant que $Y = y_j$ est l'espérance de X sous la loi $\mathbb{P}^{Y=y_j}$.*

$$\mathbb{E}(X \mid Y = y_j) = \sum_{i=0}^{m} x_i \mathbb{P}(X = x_i | Y = y_j) =: u_j.$$

Définition 2. *L'espérance conditionnelle de X sachant Y est la variable aléatoire :*

$$U(\omega) \mapsto u_j \text{ si } Y(\omega) = y_j.$$

Exemple 1. $\Omega = [\![1, 6]\!]$, $\mathscr{F} = \mathcal{P}(\Omega)$ *et* $\mathbb{P}$ *uniforme.*

Si on prend $X = Id$ et $Y = 1_{\{2,4,6\}}$, on a :

$$U(\omega) = \mathbb{E}(X \mid Y)(\omega) = \begin{cases} 4 & \text{si } \omega \text{ est pair} \\ 3 & \text{sinon.} \end{cases}$$

Notons $\mathcal{G} := \sigma(Y)$. Elle est formée des unions d'atomes $G_j = \{Y = y_j\}$. La variable U est constante sur chacun des G_j, donc $\mathcal{G}$-mesurable. Par ailleurs,

$$\begin{aligned} \mathbb{E}(U 1_{G_j}) = u_j \mathbb{P}(G_j) &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \mid Y = y_j) \mathbb{P}(Y = y_j) \\ &= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i \text{ et } Y = y_j) = \mathbb{E}(X 1_{G_j}) \end{aligned}$$

Pour tout $G \in \mathcal{G}$, $\mathbb{E}(U 1_G) = \mathbb{E}(X 1_G)$.

1.2 Définition de l'espérance conditionnelle

Par la suite, on fixe un espace de probabilité $(\Omega, \mathscr{F}, \mathbb{P})$.

Théorème 1. *Soient $\mathcal{G}$ une sous-tribu de $\mathscr{F}$ et $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$. Il existe une unique variable $\mathbb{E}(X \mid \mathcal{G}) \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ qui vérifie :*

$$\forall G \in \mathcal{G}, \quad \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_G) = \mathbb{E}(X1_G).$$

Remarque 1.
- *Cette dernière propriété est équivalente à : pour toute variable aléatoire Z $\mathcal{G}$-mesurable et bornée, $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})Z) = \mathbb{E}(XZ)$.*
- *L'unicité est dans L^1, i.e. aux négligeables près.*
- *Par définition, $\mathbb{E}(X \mid \mathcal{G})$ est $\mathcal{G}$-mesurable !*

Démonstration.
! : Soient Y et Y' deux versions de $\mathbb{E}(X \mid \mathcal{G})$.
Alors Y et Y' sont $\mathcal{G}$-mesurables donc $G = \{Y > Y'\} \in \mathcal{G}$.
On a aussi $\mathbb{E}(Y1_G) = \mathbb{E}(Y'1_G)$, donc $\mathbb{E}((Y - Y')1_{Y-Y'>0}) = 0$ et $Y \leqslant Y'$ p.s.
Par symétrie, $Y = Y'$ p.s.
$\exists$: Dans le cas où $X \in L^2(\Omega, \mathscr{F}, \mathbb{P})$.
Cet espace est un espace de Hilbert dont $L^2(\Omega, \mathcal{G}, \mathbb{P})$ est un sev complet.
Soit Y la projection orthogonale de X. On a :

$$\begin{cases} \mathbb{E}((Y-X)^2) = \inf\limits_{Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}((X-Z)^2) \\ \forall Z \in L^2(\Omega, \}, \mathbb{P}), \mathbb{E}((X-Y)Z) = 0. \end{cases}$$

La variable aléatoire Y convient.
Passage de L^2 à L^1 :
Lemme 1. *Soit U une variable aléatoire positive bornée. Alors $\mathbb{E}(U \mid \mathcal{G}) \geqslant 0$ p.s.*

Démonstration. Soit W une version de $\mathbb{E}(U \mid \mathcal{G})$.
Par l'absurde, si $\mathbb{P}(W < 0) > 0$ alors $\exists n \in \mathbb{N}$ tel que $\mathbb{P}\underbrace{\left(W \leqslant -\frac{1}{n}\right)}_{G} > 0$. On a alors

$$0 \leqslant \mathbb{E}(U1_G) = \mathbb{E}(W1_G) \leqslant -\frac{1}{n}\mathbb{P}(G) < 0.$$

Contradiction. Donc $\mathbb{P}(W \geqslant 0) = 1$. □

Si $X \in L^1$, quitte à considérer les parties positives et négatives, il est loisible de supposer $X \geqslant 0$ p.s. et il existe alors une suite croissante de variables aléatoires X_n bornées (donc dans L^2) qui converge vers X dans L^1. Alors si on pose $Y_n = \mathbb{E}(X_n \mid \mathcal{G})$, on a Y_n croissante positive (lemme 1).
On pose $Y := \limsup_{n \to +\infty} Y_n$. La variable Y ainsi définie est $\mathcal{G}$-mesurable ; comme pour tout $G \in \mathcal{G}$, on a $\mathbb{E}(X_n 1_G) = \mathbb{E}(Y_n 1_G)$, par convergence monotone, en passant à la limite, $\mathbb{E}(X1_G) = \mathbb{E}(Y1_G)$. En particulier, la variable Y est intégrable. □

Remarque 2. *Que ce soit d'un point de vue théorique ou dans les applications, il faut garder à l'esprit le cas $\mathbb{L}^2$ où l'espérance conditionnelle n'est autre qu'une projection orthogonale. L'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ est la variable aléatoire $\mathcal{G}-$mesurable la plus proche de X pour la distance $\mathbb{L}^2$. Un parallèle éclairant est celui de l'approximation des fonctions continues par les polynômes.*

Exemple 2. *Soit* $(\Omega, \mathscr{F}, \mathbb{P}) = (]0,1[, \mathscr{B}(]0,1[), \lambda)$.
Soit $f \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ *et* $\mathcal{G} = \sigma(]\frac{i-1}{n}, \frac{i}{n}], i \in [\![1, n]\!])$.

$$\mathbb{E}(f \mid \mathcal{G}) = \sum_{i=1}^{n} f_i 1_{]\frac{i-1}{n}, \frac{i}{n}]}$$

avec $f_i = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} f \, \mathrm{d}\lambda$.

$\mathbb{E}(f \mid \mathcal{G})$ *est bien* $\mathcal{G}$*-mesurable et par ailleurs,*

$$\begin{aligned} \mathbb{E}(f 1_{]\frac{j-1}{n}, \frac{j}{n}]}) &= E\left(\sum_{i=1}^{n} f_i 1_{]\frac{i-1}{n}, \frac{i}{n}]} 1_{]\frac{j-1}{n}, \frac{j}{n}]}\right) \\ &= \mathbb{E}(f_j 1_{]\frac{j-1}{n}, \frac{j}{n}]}). \end{aligned}$$

Théorème 2. *Si X est un variable aléatoire positive alors la formule* $\mathbb{E}(X \mid \mathcal{G}) = \lim_{n \to +\infty} \mathbb{E}(X \wedge n \mid \mathcal{G})$ *définit une variable $\mathcal{G}$-mesurable caractérisée par :*

$$\forall Z \geqslant 0 \quad \mathcal{G}\text{-mesurable}, \ \mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X \mid \mathcal{G})Z).$$

Définition 3. *Si X et Y sont des variables aléatoires définies sur un espace de probabilités* $(\Omega, \mathscr{F}, \mathbb{P})$*, on pose*

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

Lemme 2. *Soient U et V des variables aléatoires réélles définies sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbb{P})$. Si U est $\sigma(V)-$mesurable, alors il existe une fonction mesurable φ telle que $U = \varphi(V)$.*

Démonstration. Supposons tout d'abord que U est étagée, i.e. ne prend qu'un nombre fini de valeurs

$$U(\omega) = \sum_{i=1}^{N} u_i 1_{A_i}(\omega).$$

Comme U est $\sigma(V)-$mesurable, on a alors $A_i = \{U = u_i\} \in \sigma(V)$ donc $A_i = V^{-1}(B_i)$ avec B_i un borélien. Dès lors

$$U(\omega) = \sum_{i=1}^{N} u_i 1_{A_i}(\omega) = \sum_{i=1}^{N} u_i 1_{B_i}(V(\omega)) = \underbrace{\left(\sum_{i=1}^{N} u_i 1_{B_i}\right)}_{\varphi} \circ V(\omega).$$

Dans le cas général, U peut s'écrire $U = \lim_{n\to+\infty} U_n$ avec U_n étagée, et d'après ci-dessus $U_n = \varphi_n(V)$ pour tout $n \geq 1$. On pose alors

$$C = \{x \in \mathbb{R}, \varphi_n(x) \text{ converge}\} = \{x \in \mathbb{R}, -\infty < \liminf_{n\to+\infty} \varphi_n(x) = \limsup_{n\to+\infty} \varphi_n(x) < +\infty\}.$$

Alors C est un borélien et comme par définition $U = \lim_{n\to+\infty} U_n = \lim_{n\to+\infty} \varphi_n(V)$, on a $V(\Omega) \subset C$. Si on pose $\varphi(x) := \lim_{n\to+\infty} \varphi_n(x)$ si $x \in C$ et $\varphi(x) = 0$ sinon, alors φ est borélienne et $U = \varphi(V)$. $\square$

Corollaire 1. *Soient X et Y des variables aléatoires réelles définies sur un espace de probabilités $(\Omega, \mathscr{F}, \mathbb{P})$. Il existe une fonction mesurable φ telle que*

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

Exercice 1. *Exemples de conditionnements discrets*

1. Soient X_1 et X_2 deux variables aléatoires indépendantes de lois binomiales de paramètres respectifs (n_1, p) et (n_2, p). Déterminer la loi de la variable X_1 sachant $X_1 + X_2 = n$. Déterminer $\mathbb{E}(X_1|X_1 + X_2)$.
2. Soient $X_1, \ldots, X_p$ des variables indépendantes de loi de Poisson de paramètres respectifs $\lambda_1, \ldots, \lambda_p$. Déterminer la loi du vecteur aléatoire $(X_1, \ldots, X_p)$ sachant $X_1 + X_2 + \ldots + X_p = n$

Exercice 2. *Espérance conditionnelle par rapport à la somme*
Soient $X_1, ..., X_n$ des variables aléatoires indépendantes identiquement distribuées intégrables. On définit $S_n := X_1 + ... + X_n$.

1. Montrer que $\mathbb{E}[X_1|S_n] = \mathbb{E}[X_i|S_n]$ pour tout $i \in \{1, ..., n\}$.
2. En déduire l'expression de $\mathbb{E}[X_1|S_n]$.
3. Montrer que S_n/n converge presque sûrement et dans $\mathbb{L}^1$ vers $\mathbb{E}[X_1]$.

Exercice 3. *Variables positives et conditionnement*
Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $\mathcal{G}$ une sous tribu de $\mathcal{F}$ et X une variable aléatoire réelle positive. Montrer que l'ensemble $\{\mathbb{E}(X|\mathcal{G}) > 0\}$ est le plus petit ensemble $\mathcal{G}-$mesurable (aux négligeables près) qui contient $\{X > 0\}$.

Exercice 4. *Inégalité de Cauchy–Schwarz conditionnelle*
Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $\mathcal{G}$ une sous tribu de $\mathcal{F}$ et X, Y deux variables aléatoires réelles telles que $\mathbb{E}(X^2) < +\infty$ et $\mathbb{E}(Y^2) < +\infty$. En considérant le fait que $\mathbb{E}[(X + \theta Y)^2|\mathcal{G}] \geq 0\, p.s.$ pour tout $\theta \in \mathbb{Q}$, établir l'inégalité de Cauchy–Schwarz conditionnelle $\mathbb{E}[XY|\mathcal{G}]^2 \leq \mathbb{E}[X^2|\mathcal{G}]\mathbb{E}[Y^2|\mathcal{G}]$ *p.s.*

1.3 Propriétés de l'espérance conditionnelle

1.3.1 Premières propriétés

Proposition 1. *Soit $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ et $\mathcal{G}$ une sous-tribu de $\mathscr{F}$.*

1. *Si X est $\mathcal{G}$-mesurable alors $\mathbb{E}(X \mid \mathcal{G}) = X$ p.s.*
2. *$X \mapsto \mathbb{E}(X \mid \mathcal{G})$ est linéaire*
3. *$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$*
4. *$|\mathbb{E}(X \mid \mathcal{G})| \leqslant \mathbb{E}(|X| \mid \mathcal{G})$ p.s.*
5. *Si $X \geqslant X'$ p.s., $\mathbb{E}(X \mid \mathcal{G}) \geqslant \mathbb{E}(X' \mid \mathcal{G})$ p.s.*

Démonstration.

1. Unicité
2. Si $U = \mathbb{E}(X \mid \mathcal{G})$ et $V = \mathbb{E}(Y \mid \mathcal{G})$, alors $\alpha U + \beta V$ est une version de $\mathbb{E}(\alpha X + \beta Y \mid \mathcal{G})$.
3. Définition avec $1_G = 1_\Omega = 1$.
4. $|\mathbb{E}(X \mid \mathcal{G})| = |\mathbb{E}(X^+ \mid \mathcal{G}) - \mathbb{E}(X^- \mid \mathcal{G})| \leqslant \mathbb{E}(X^+ \mid \mathcal{G}) + \mathbb{E}(X^- \mid \mathcal{G}) = \mathbb{E}(|X| \mid \mathcal{G})$.
5. Linéarité+lemme 1. □

Remarque 3.

- *$\|\mathbb{E}(X \mid \mathcal{G})\|_1 \leqslant \|X\|_1$, i.e. l'espérance conditionnelle est une contraction dans $\mathbb{L}^1$.*
- *L'opérateur $X \mapsto \mathbb{E}(X \mid \mathcal{G})$ linéaire positif de norme 1.*

1.3.2 Extensions des théorèmes usuels

Proposition 2. *Soit $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$, $\mathcal{G}$ une sous-tribu de $\mathscr{F}$.*

1. *Si $(X_n)_n$ est une suite positive croissante qui tend vers X p.s. alors $\mathbb{E}(X_n \mid \mathcal{G})$ croît vers $\mathbb{E}(X \mid \mathcal{G})$ p.s.*
2. *Si $(X_n)_n$ est une suite positive, $\mathbb{E}(\liminf X_n \mid \mathcal{G}) \leqslant \liminf \mathbb{E}(X_n \mid \mathcal{G})$ p.s.*
3. *Si $|X_n(\omega)| \leqslant V(\omega)$ avec $\mathbb{E}(V) < \infty$ et si X_n converge p.s. vers X alors $\mathbb{E}(X_n \mid \mathcal{G})$ converge p.s. vers $\mathbb{E}(X \mid \mathcal{G})$.*
4. *Si $c : \mathbb{R} \to \mathbb{R}$ est convexe et $\mathbb{E}(|c(X)|) < \infty$, alors $\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant c(\mathbb{E}(X \mid \mathcal{G}))$ p.s.*

Démonstration.

1. Voir passage $L^2 \to L^1$.
2. Exo
3. Exo
4. On utilise $c(x) = \sup_{n \in \mathbb{N}}(a_n x + b_n)$. Soit $n \in \mathbb{N}$.
   On a $c(X) \geqslant a_n X + b_n$ donc $\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant a_n \mathbb{E}(X \mid \mathcal{G}) + b_n$ p.s.
   Comme on a un sup sur un ensemble dénombrable, on peut passer au sup et on a :
   $$\mathbb{E}(c(X) \mid \mathcal{G}) \geqslant c(\mathbb{E}(X \mid \mathcal{G}))\text{p.s.}$$ □

Remarque 4. *On a ainsi, pour tout $p \geqslant 1$, $\|\mathbb{E}(X \mid \mathcal{G})\|_p \leqslant \|X\|_p$, de sorte que l'espérance conditionnelle est une contraction dans $\mathbb{L}^p$.*

1.3.3 Propriétés spécifiques

Proposition 3. *Soient X une variable aléatoire et Y une variable $\mathcal{G}$-mesurable.*

$$\mathbb{E}(XY \mid \mathcal{G}) = Y\mathbb{E}(X \mid \mathcal{G}) \ p.s.$$

dès lors que les espérances sont bien définie.

Démonstration. $Y\mathbb{E}(X \mid \mathcal{G})$ est $\mathcal{G}$-mesurable comme produit de deux variables $\mathcal{G}$-mesurables. Si $G \in \mathcal{G}$, on a

$$\mathbb{E}(Y\mathbb{E}(X \mid \mathcal{G})1_G) = \mathbb{E}(Y1_G\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(XY1_G).$$

Par unicité de l'espérance conditionnelle, on a le résultat. □

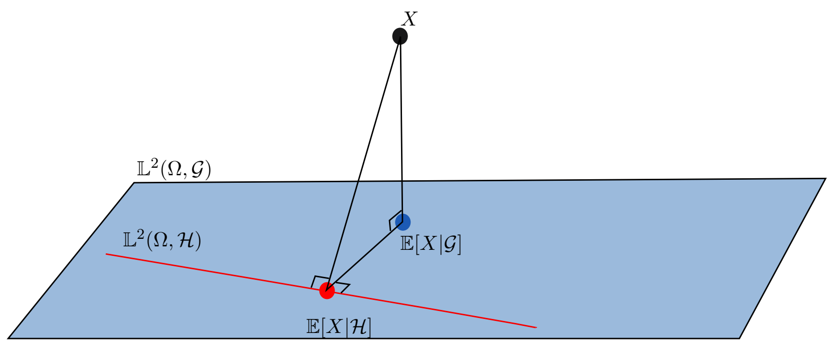
Proposition 4. *Soient $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ deux sous-tribus et $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors*

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}) \ p.s.$$

Démonstration. $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})$ est $\mathcal{H}$-mesurable. Soit $H \in \mathcal{H}$.

$$\mathbb{E}(\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H})1_H) = \mathbb{E}(\mathbb{E}(\mathbb{E}(X1_H \mid \mathcal{G}) \mid \mathcal{H})) = \mathbb{E}(X1_H).$$ □

Remarque 5. *Cette dernière propriété est évidente si on pense l'espérance conditionnelle comme une projection orthogonale. Voir le dessin ci-dessous.*



Proposition 5. *Deux tribus $\mathcal{G}$ et $\mathcal{H}$ sont indépendantes ssi pour tout événement $G \in \mathcal{G}$, $\mathbb{E}(1_G \mid \mathcal{H}) = \mathbb{E}(1_G) = \mathbb{P}(G)$ p.s.*

Remarque 6. *X et Y sont indépendantes ssi pour tout g mesurables*

$$\mathbb{E}(g(X) \mid \sigma(Y)) = \mathbb{E}(g(X)) \ p.s.$$

Démonstration.

$\Rightarrow$ Si $H \in \mathcal{H}$ et $G \in \mathcal{G}$. On a

$$\mathbb{E}(1_G1_H) = \mathbb{E}(1_G)\mathbb{E}(1_H) = \mathbb{E}(\mathbb{E}(1_G)1_H)$$

donc $\mathbb{E}(1_G) = \mathbb{E}(1_G \mid \mathcal{H})$ par unicité.

$\Leftarrow$ Si $H \in \mathcal{H}$ et $G \in \mathcal{G}$, on a

$$\mathbb{P}(H \cap G) = \mathbb{E}(1_H 1_G) = \mathbb{E}(\mathbb{E}(1_G)1_H) = \mathbb{P}(G)\mathbb{P}(H),$$

d'où $H \perp\!\!\!\perp G$. $\square$

Proposition 6. *Soient X, Y deux variables aléatoires à valeurs dans E, F, $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On suppose que $X \perp\!\!\!\perp \mathcal{G}$ et que Y est $\mathcal{G}$-mesurable.*

Alors pour tout $f : E \times F \to \mathbb{R}$ mesurable,

$$\mathbb{E}(f(X,Y) \mid \mathcal{G}) = \int_E f(x,Y)\mathbb{P}_X(\mathrm{d}x) = \Phi(Y).$$

Démonstration. $\Phi(Y)$ est $\mathcal{G}$-mesurable. Soit Z $\mathcal{G}$-mesurable bornée, on a :

$$\begin{aligned}\mathbb{E}(f(X,Y)Z) &= \int f(x,y)z\mathbb{P}_{(X,Y,Z)}(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) \\ &= \int\int f(x,y)\mathbb{P}_X(\mathrm{d}x)z\mathbb{P}_{Y,Z}(\mathrm{d}y, \mathrm{d}z) \\ &= E(\Phi(Y)Z).\end{aligned}$$

$\square$

2 Espérance conditionnelle gaussienne

2.1 Rappels sur les vecteurs gaussiens

On dit qu'une variable aléatoire réelle X est suit la loi normale $\mathcal{N}(0,1)$, ou encore est une variable gaussienne centrée réduite, si elle admet la densité f_X suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}.$$

Sa fonction caractéristique φ_X, i.e. sa transformée de Fourier, est alors donnée par la formule

$$\varphi_X(\xi) = \mathbb{E}[e^{i\xi X}] = e^{-\frac{1}{2}\xi^2}.$$

On dit qu'une variable aléatoire réelle X est suit la loi normale $\mathcal{N}(m, \sigma^2)$, ou encore est une variable gaussienne de moyenne m et de variance σ^2, si elle admet la densité f_X suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Sa fonction caractéristique φ_X est alors donnée par la formule

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

Si $X \sim \mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$, alors $(X-m)/\sigma \sim \mathcal{N}(0,1)$. Si $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$ alors $X_n \overset{loi}{\to} X \sim \mathcal{N}(m, \sigma^2)$ lorsque $n \to +\infty$ ssi $m_n \to m$ et $\sigma_n^2 \to \sigma^2$.
On dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_d)$ de $\mathbb{R}^d$ est un vecteur gaussien si pour tout vecteur $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, la variable réelle $a \cdot X = \langle a, X \rangle = \sum_{i=1}^d a_i X_i$ est une variable gaussienne.
La loi d'un vecteur gaussien $X = (X_1, \ldots, X_d)$ est caractérisée par la seule donnée de la moyenne $m = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$ et de la matrice de covariance $\Sigma = (\Sigma_{ij})_{1 \leq i,j \leq d}$ où $\Sigma_{i,j} = \text{cov}(X_i, X_j) := \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$. On note alors $X \sim \mathcal{N}(m, \Sigma)$.
Il existe un vecteur gaussien centré de matrice de covariance Σ ssi la matrice Σ est symétrique et positive, i.e. $x^t \Sigma x \geq 0$, pour tout $x \in \mathbb{R}^d$.
Le vecteur gaussien $X = (X_1, \ldots, X_d) \sim \mathcal{N}(m, \Sigma)$ admet une densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si $\det(\Sigma) \neq 0$ auquel cas, celle-ci est donnée par la formule :

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-m)^t \Sigma^{-1}(x-m)\right), \quad x \in \mathbb{R}^d.$$

Sa fonction caractéristique est quant à elle donnée par

$$\varphi_X(\xi) = e^{i\xi \cdot m} e^{\frac{1}{2}\xi^t \Sigma \xi}.$$

Si X est gaussien $\mathcal{N}(m, \Sigma)$ et si A est une matrice, alors $AX \sim \mathcal{N}(Am, A\Sigma A^t)$. En particulier, si la matrice de covariance Σ est inversible et si $\sqrt{\Sigma^{-1}}$ est une racine carrée de son inverse, on a $\sqrt{\Sigma^{-1}}(X-m) \sim \mathcal{N}(0, \text{Id})$.

La proposition suivante est cruciale, moralement dans le monde gaussien indépendance se confond avec orthogonalité.

Proposition 7. *Soit $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ un vecteur gaussien centré. Alors les vecteurs $(X_1, \ldots, X_n)$ et $(Y_1, \ldots Y_m)$ sont indépendants ssi cov$(X_i, Y_j) = 0$.*

Démonstration. On peut le voir par exemple sur la fonction caractéristique. La matrice de covariance est alors diagonale par bloc, cela a pour conséquence que la fonction caractéristique du "couple" n'est autre que le produit des fonctions caractéristiques. □

Exercice 5. *Un tour de chauffe*
On considère une matrice $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, où ρ est un réel tel que $|\rho| < 1$.

1. Montrer qu'il existe un vecteur aléatoire gaussien $X = (X_1, X_2)^t$ centré et de matrice de covariance Σ.
2. On pose $Y_1 = \frac{1}{\sqrt{2}}(X_1 + X_2)$ et $Y_2 = \frac{1}{\sqrt{2}}(X_1 - X_2)$. Montrez que le vecteur $Y = (Y_1, Y_2)^t$ est un vecteur gaussien et précisez sa matrice de covariance.
3. Les variables Y_1 et Y_2 sont-elles indépendantes ?
4. Explicitez la densité de Y.

Exercice 6. *Invariance par les isométries*
Si X est un vecteur gaussien centré de matrice de covariance $\sigma^2 \mathrm{Id}_{\mathbb{R}^d}$ et si O est une matrice orthogonale, quelle est la loi du vecteur $Y = OX$?

Exercice 7. *Un contre exemple à marginales gaussiennes*
Soit $X \sim \mathcal{N}(0,1)$ et $a > 0$. On définit une nouvelle variable $X^{(a)}$ de la façon suivante :

$$X^{(a)} = X 1_{|X|>a} - X 1_{|X|\leq a}.$$

1. Montrez que $X^{(a)}$ est une variable gaussienne.
2. Montrez qu'il existe un réel positif b tel que

$$\frac{1}{\sqrt{2\pi}} \int_0^b x^2 e^{-x^2/2} dx = \frac{1}{4}.$$

3. Calculer la covariance de X et $X^{(b)}$. Le vecteur $(X, X^{(b)})$ est-il gaussien ?

2.2 Espérance conditionnelle comme projection fini-dimensionnelle

Proposition 8. *Soit $(X, Y_1, \ldots, Y_n)$ un vecteur gaussien centré. Alors l'espérance conditionnelle $\mathbb{E}(X \mid \sigma(Y_1, \ldots, Y_n))$ est la projection orthogonale de X sur l'espace vectoriel engendré par les Y_i. Il existe donc des scalaires λ_i tels que*

$$\mathbb{E}(X \mid \sigma(Y_1, \ldots, Y_n)) = \sum_i \lambda_i Y_i =: \widehat{X}.$$

En fait, si Γ est la matrice de covariance du vecteur $(X, Y_1, \ldots, Y_n)$

$$\Gamma = \begin{pmatrix} \Gamma_X & \Gamma_{X,Y} \\ \Gamma_{X,Y}^t & \Gamma_Y \end{pmatrix}$$

on a la formule explicite

$$\mathbb{E}(X \mid \sigma(Y_1, \ldots, Y_n)) = \mathbb{E}[X] + \Gamma_{X,Y} \Gamma_Y^{-1} (Y - \mathbb{E}[Y]).$$

De plus, pour tout h mesurable,

$$\mathbb{E}(h(X) \mid \sigma(Y_1, \ldots, Y_n)) = \int h(x) q_{\widehat{X},\sigma^2}(x) \, \mathrm{d}x$$

où $q_{\widehat{X},\sigma^2}$ est la densité d'une gaussienne $\mathcal{N}(\widehat{X}, \sigma^2)$:

$$q_{\widehat{X},\sigma^2}(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(u-\widehat{X})^2}{2\sigma^2}}, \quad \textit{où} \quad \sigma^2 = \mathbb{E}[(X - \widehat{X})^2].$$

Démonstration. Si $X \in \sigma(Y_1, \dots, Y_n)$ le résultat est trivial :

$$\mathbb{E}(X \mid \sigma(Y_1, \dots, Y_n)) = X \text{ p.s.}$$

Sinon, soit $\widehat{X} = \sum_i \lambda_i Y_i$ la projection orthogonale sur Vect $\{Y_1, \dots, Y_n\}$. Par orthogonalité $\mathbb{E}((X - \widehat{X})Y_i) = 0$ donc en particulier, $(X - \widehat{X}, Y_1, \dots, Y_n)$ est un vecteur gaussien centré et d'après la proposition précédente, $X - \widehat{X}$ est indépendant de $(Y_1, \cdots, Y_n)$. D'où

$$\mathbb{E}(X \mid \sigma(Y_1, \dots, Y_n)) = \mathbb{E}(X - \widehat{X} \mid Y_1, \dots, Y_n) + \widehat{X} = \mathbb{E}(X - \widehat{X}) + \widehat{X} = \widehat{X}. \qquad \square$$

Exemple 3. *Soit (X, Y) un vecteur gaussien centré. On projette sur $\mathbb{R}Y$ donc il existe $\lambda \in \mathbb{R}$ tel que :*

$$\mathbb{E}(X \mid Y) = \lambda Y \ \textit{p.s.}$$

On passe à l'espérance :

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(X \mid Y)Y) = \lambda \mathbb{E}(Y^2)$$

Donc $\lambda = \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}$.

Exemple 4. *Soit $(X, Y, Z)'$ un vecteur gaussien centré de matrice de covariance*

$$\Gamma = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{pmatrix}.$$

Alors

$$\mathbb{E}[Z|X, Y] = (1/2, 1/2) \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} X \\ Y \end{pmatrix},$$

c'est-à-dire

$$\mathbb{E}[Z|X, Y] = \frac{X + Y}{3}.$$

Exercice 8. *Exemples de conditionnements gaussiens*
On considère un vecteur gaussien $[X, Y]'$ de moyenne $m = [1, -1]'$ et de matrice de covariance :

$$\Gamma = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}.$$

1. Écrire la densité du vecteur $[X, Y]'$. Quelle est la loi de X ? de Y ? de $X + Y$?
2. Déterminer l'espérance conditionnelle $\mathbb{E}[X|Y]$. Quelle est sa loi ?
3. Si (U, V) est un vecteur gaussien centré réduit, que vaut $\mathbb{E}[U|U + V]$? Retrouver le résultat géométriquement.

Exercice 9. *Conditionnement gaussien par un couple*
Soit $X = (X_1, X_2, X_3)'$ un vecteur gaussien centré de matrice de covariance

$$\Gamma := \begin{pmatrix} 2 & 2 & -2 \\ 2 & 5 & 1 \\ -2 & 1 & 5 \end{pmatrix}.$$

1. Le vecteur aléatoire X admet-il une densité ?
2. Déterminer $\mathbb{E}[X_1|X_2]$ et $\mathbb{E}[X_1|X_2, X_3]$.
3. Mêmes questions que ci-dessus lorsque

$$\Gamma := \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 4 \end{pmatrix}.$$

Exercice 10. *Projection itérée dans le cas gaussien*
Soit $(X, Y, Z)'$ un vecteur gaussien centré de matrice de covariance

$$\Gamma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

1. Montrer que $\mathbb{E}[X|Y, Z] = \frac{Y+Z}{3}$.
2. Montrer que $\mathbb{E}[X|Z] = Z/2$.
3. Montrer que le vecteur $(Y + Z, Z)$ est gaussien de matrice de covariance

$$\widetilde{\Gamma} = \begin{pmatrix} 6 & 3 \\ 3 & 2 \end{pmatrix}.$$

4. En déduire $\mathbb{E}[Y + Z|Z]$.
5. Conclure que l'on a bien $\mathbb{E}[\mathbb{E}[X|Y, Z]|Z] = \mathbb{E}[X|Z]$.

3 Calcul d'espérance conditionnelle

3.1 Cas des variables discrètes

Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $Y : \Omega \to \{y_1, \ldots, y_n\}$ une variable discrète.

$$\mathbb{E}(X \mid Y) = \mathbb{E}(X \mid \sigma(Y))$$

$\sigma(Y)$ est engendrée par les atomes $G_i = \{Y = y_i\}$.

Proposition 9. *On a*

$$\mathbb{E}(X \mid Y) = \sum_{i=1}^{\infty} \frac{\mathbb{E}(X 1_{G_i})}{\mathbb{P}(G_i)} 1_{G_i}$$

Démonstration. Le membre de droite est $\sigma(Y)$-mesurable. De plus, on a

$$\mathbb{E}\left(\sum_{i=1}^{\infty}\frac{\mathbb{E}(X1_{G_i})}{\mathbb{P}(G_i)}1_{G_i}1_{G_j}\right) = \mathbb{E}(X1_{G_j})$$ □

Exercice 11. *Variables géométriques et conditionnement*
On considère une variable aléatoire géométrique X telle que $\mathbb{P}(X = i) = 2/3^i$ pour $i \geq 1$. Soit Y une variable aléatoire telle que, sachant $X = i$, la loi de Y est la loi uniforme sur $\{i, i+1\}$.

1. Pour $i \in \mathbb{N}^*$, déterminer $\mathbb{E}[Y|X = i]$. En déduire $\mathbb{E}[Y|X]$, puis $\mathbb{E}[Y]$.
2. Calculer la loi jointe du couple (X, Y) et en déduire la loi de Y.
3. Pour tout $j \in \mathbb{N}^*$, déterminer $\mathbb{E}[X|Y = j]$ et en déduire $\mathbb{E}[X|Y]$.

3.2 Cas des variables à densité

Proposition 10. *Soit (X, Y) de densité $p(x, y)$ par rapport à la mesure de Lebesgue. Alors pour tout fonction mesurable h,*

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

avec

$$\varphi(u) = \begin{cases} \frac{\int h(x)p(x,u)\,\mathrm{d}x}{\int p(x,u)\,\mathrm{d}x} & \textit{si } p(x,u)\,\mathrm{d}x > 0 \\ 0 & \textit{sinon} \end{cases}$$

(densité conditionnelle de X par rapport à Y).

Démonstration. $\varphi(Y)$ est Y-mesurable.
Soit g mesurable bornée.

$$\begin{aligned} \mathbb{E}(h(x)g(y)) &= \int h(x)g(y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \frac{\int h(x)p(x,y)\,\mathrm{d}x}{q(y)} g(y)q(y)\,\mathrm{d}y \\ &= \int \varphi(y)g(y)q(y)\,\mathrm{d}y = \mathbb{E}(\phi(Y)g(Y)) \end{aligned}$$

avec $q(y) = \int p(x, y)\,\mathrm{d}y$. □

Exemple 5. *Soit (X, Y) de densité* $\mathrm{e}^{-Y}1_{0<X<Y}$.

$$\mathbb{E}(h(x) \mid Y) = \frac{\int h(x)\mathrm{e}^{-Y}1_{0<x<Y}\,\mathrm{d}x}{\int \mathrm{e}^{-Y}1_{0<x<Y}\,\mathrm{d}x} = \frac{1}{Y}\int_0^Y h(x)\,\mathrm{d}x$$

Exercice 12. *Loi uniforme dans un triangle*
Soit (U, V) un couple de variables aléatoires admettant la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ (un dessin peut aider) :

$$f(u,v) = \begin{cases} 2 & \text{si} \quad 0 \leq u \leq v \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer les lois conditionnelles de U sachant $V = v$ et de V sachant $U = u$.
2. En déduire les espérances conditionnelles $\mathbb{E}[U|V]$ puis $\mathbb{E}[V|U]$.

Exercice 13. *Calcul d'espérance conditionelle*
Calculer $\mathbb{E}[X|Y]$ lorsque la loi du couple de variables aléatoires réelles (X, Y) admet la densité f suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$:

$$\begin{aligned} (i) \quad & f(x,y) = 4y(x-y)\exp(-(x+y))1_{0\leq y\leq x}(x,y), \\ (ii) \quad & f(x,y) = \frac{1}{y}\exp\left(-\frac{x}{y} - y\right)1_{]0,+\infty[^2}(x,y). \end{aligned}$$

Exercice 14. *Somme de variables exponentielles et conditionnement*
Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de même loi exponentielle de paramètre $\lambda > 0$. On note $T := X_1 + \ldots + X_n$. Calculer $\mathbb{E}(h(X_1)|T)$ pour toute fonction h borélienne positive. Que remarque-t-on lorsque $n = 2$?

4 Notion de loi conditionnelle

4.1 Définitions

Définition 4. *Soit $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. On appelle probabilité de transition une application $\nu : E \times \mathcal{F} \to [0, 1]$ telle que pour tout $x \in E$, $A \mapsto \nu(x, A)$ est une probabilité sur $\mathcal{F}$ et pour tout $A \in \mathcal{F}$, $x \mapsto \nu(x, A)$ est mesurable.*

Exemple 6. *Si λ est une mesure σ-finie sur $\mathcal{F}$ et $f : E \times F \to \mathbb{R}$ mesurable telle que $\int f(x,y)\lambda(\mathrm{d}y) = 1$ alors $\nu(x, A) = \int_A f(x,y)\lambda(\mathrm{d}y)$ est une probabilité de transition.*

Proposition 11. *Soit ν une probabilité de transition sur $(E, \mathcal{E})$, $(F, \mathcal{F})$.*

Si h est mesurable positive (bornée) sur $(F, \mathcal{F})$ alors $\varphi(x) = \int h(y)\nu(x, \mathrm{d}y)$ est mesurable.

Si λ est une mesure de probabilité sur $(E, \mathcal{E})$, alors $\mu(A) = \int \nu(x, A)\lambda(\mathrm{d}x)$ est une mesure sur $(F, \mathcal{F})$.

Définition 5. *Soient X, Y deux variables aléatoires sur $(E, \mathcal{E})$ et $(F, \mathcal{F})$. On appelle loi conditionnelle de Y suivant X toute probabilité de transition ν sur $E \times \mathcal{F}$ telle que pour tout h mesurable positive sur $(F, \mathcal{F})$, on ait :*

$$\varphi(X) = \mathbb{E}(h(Y) \mid X) = \int h(y)\nu(X, \mathrm{d}y)$$

Remarque 7. *D'après la proposition précédente, φ est mesurable donc le membre de droite est bien mesurable par rapport à $\sigma(X)$.*

Par définition, on pose alors $\mathbb{P}(Y \in A \mid X) = \nu(X, A)$ (pour $A \in \mathcal{F}$).

4.2 Existence et unicité

Unicité : Si ν et ν' sont deux lois conditionnelles de Y sachant X alors pour tout $A \in \mathcal{F}$, $\nu(X, A) = \nu'(X, A) = \mathbb{P}(Y \in A \mid X)$ p.s.ie pour tout $A \in \mathcal{F}$, $\nu(x, A) = \nu'(x, A)$ $\mathbb{P}_X$-p.s.

Si les mesures de probabilité sur $(F, \mathcal{F})$ sont caractérisées par leurs valeurs sur un nombre dénombrables d'ensembles (c'est le cas dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ alors $\mathbb{P}_X$-p.s.pour tout $A \in \mathcal{F}$, $\nu(x, A) = \nu'(x, A)$.

Existence :

Théorème 3. *Si $(E, \mathcal{E})$ et $(F, \mathcal{F})$ sont des espaces polonais (métriques, complets, séparables) munis de leur tribu borélienne, il existe une loi conditionnelle de Y sachant X.*

4.3 En pratique

4.3.1 Cas discret

Si X est une variable discrète,

$$\nu(x, A) = \begin{cases} \mathbb{P}(Y \in A \mid X = x) & \text{si } \mathbb{P}(X = x) > 0 \\ \delta_{y_0}(A) & \text{pour } y_0 \text{ fixé dans } F \end{cases}$$

4.3.2 Cas des variables à densité

Si (X, Y) a pour densité $p(x, y)$ alors on note $q(x) = \int p(x, y)\,\mathrm{d}y$ et on a :

$$\nu(x, A) = \begin{cases} \frac{1}{q(x)} \int_A p(x, y)\,\mathrm{d}t & \text{si } q(x) > 0 \\ \delta_0(A) & \text{sinon} \end{cases}$$

Exemple 7. $p(x, y) = e^{-y} 1_{0 \leqslant x \leqslant y}$. *On cherche la loi de X sachant Y.*

$$\nu(y, A) = \frac{1}{\int p(x, y)\,\mathrm{d}x} \int_A p(x, y)\,\mathrm{d}x = \frac{\mathrm{e}^{-y}}{y\mathrm{e}^{-y}} \int_{A \cap [0, y]} 1\,\mathrm{d}x) = \frac{1}{y} \lambda(A \cap [0, y])$$

C'est donc une loi uniforme sur $[0, y]$.

4.3.3 Cas gaussien

Si $(Y, X_1, \ldots, X_n)$ est un vecteur gaussen centré, on a

$$\nu((x_1, \ldots, x_n), A) = \int_A q_{\sum \lambda_i x_i, \sigma^2}(u)\,\mathrm{d}u$$

où $q_{\sum \lambda_i x_i, \sigma^2}$ est la loi gaussienne et $\sum_i \lambda_i x_i$ le projeté orthogonal sur l'espace $\mathrm{Vect}\,\{X_i\}$.