Université Paris-Dauphine **2021-2022**
Cours de "Intégrale de Lebesgue et probabilités" Juin 2019

Chapitres 8 - Vecteurs Gaussiens et Conditionnement

Table des matières

1 Vecteurs Gaussiens **1**

2 Conditionnement $\mathcal{L}^1$ **4**

3 Calculs d'espérance conditionnelle **8**

Dans ce chapitre on désigne par $(\Omega, \mathcal{A}, P)$ un espace probabilisé, c'est-à-dire, un espace mesurable $(\Omega, \mathcal{A})$ muni d'une probabilité P. Sont écrites en rouge les parties hors programme, en violet les parties traitées en TD (résultats à connaître pour sa culture).

1 Vecteurs Gaussiens

Définition 1.1. *On dit qu'un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est* gaussien *si, pour tout $\eta \in \mathbb{R}^d$, la variable aléatoire réelle $\eta \cdot X = \sum_i \eta_i X_i$ suit une loi gaussienne.*

Remarque 1.2.

(i) En choisissant $\eta := e_i$, où $(e_1, \ldots, e_d)$ est la base canonique de $\mathbb{R}^d$, on voit que les composantes d'un vecteur gaussien sont des var gaussiennes. En général, la réciproque est fausse !

(ii) Si X est une var gaussienne et $a \in \mathbb{R}$, alors $X + a$ est une var gaussienne, ce que l'on voit à partir de la caractérisation des lois gaussiennes et des propriétés élementaires de la transformation de Fourier.

(iii) Plus généralement, si X est un vecteur gaussien, tout vecteur de la forme $Y = MX + a$ où $M \in M_{d,k}(\mathbb{R})$ et $a \in \mathbb{R}^k$ est aussi un vecteur gaussien, puisqu'alors $\eta \cdot Y = ({}^t M\eta) \cdot X + \eta \cdot a$ est une var gaussienne pour tout $\eta \in \mathbb{R}^k$.

(iv) Si X est un vecteur aléatoire, $a \in \mathbb{R}$ et $M \in M_d(\mathbb{R})$, alors

$$\varphi_{MX+a}(\xi) = e^{-ia\cdot\xi}\varphi_X({}^t M\xi). \tag{1}$$

(v) On rappelle que si $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma \geq 0$, alors

$$\varphi_X(\xi) = \exp\left(i\mu\xi - \frac{\sigma^2}{2}\xi^2\right).$$

Exercice 1.3. *Donner un exemple de vecteur aléatoire (X, Y) tel que X et Y sont gaussien mais pas $X + Y$.*

Définition 1.4. *Soient $\mu \in \mathbb{R}^d$ et $K \in M_d(\mathbb{R})$ symétrique positive. La* loi gaussienne $\mathcal{N}_d(\mu, K)$ *est l'unique loi sur $\mathbb{R}^d$ dont la transformée de Fourier est la fonction*

$$g_{\mu,K}(\xi) := \exp\left(i\mu \cdot \xi - \frac{1}{2}{}^t\xi K\xi\right).$$

Remarque 1.5. *On commence par observer que la mesure de probabilité gaussienne centrée*

$$\gamma(x)dx, \quad \gamma(x) := (2\pi)^{-d/2}\exp(-|x|^2/2),$$

notée plus simplement γ, satisfait $\hat{\gamma} = g_{0,I}$. Cela signifie que $\gamma = \mathcal{N}_d(0, I)$. On note alors Z un vecteur aléatoire de $\mathbb{R}^d$ de loi $\mathcal{N}_d(0, I)$. On observe également qu'il existe des matrices R orthogonale et Δ diagonale positive telles que $K = {}^tR\Delta R$. Comme dans la Remarque 1.2, on montre que le vecteur aléatoire $X := {}^tR\Delta^{1/2}Z$ a pour fonction caractéristique $\varphi_X = g_{0,K}$. Par ailleurs, en introduisant l'application $T : \mathbb{R}^d \to \mathbb{R}^d$, $T(x) := {}^tR\Delta^{1/2}x$, on observe que X a pour loi $\nu = T\sharp\gamma$. Cela permet bien d'exhiber une mesure de probabilité ν telle que $\hat{\nu} = g_{0,K}$. Par unicité de la transformation de Fourier, on conclut que $\nu = \mathcal{N}_d(0, K)$.

Théorème 1.6. *Un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est gaussien si, et seulement si, sa loi est gaussienne, c'est-à-dire, s'il existe un vecteur $\mu \in \mathbb{R}^d$ et une matrice symétrique $K \in M_d(\mathbb{R})$ tels que*

$$\varphi_X = g_{\mu,K} \tag{2}$$

De plus, μ et K sont l'espérance et la matrice de covariance de X.

Preuve du Théorème 1.6. Etape 1. Montrons que X gaussien implique (2). Soit X un vecteur gaussien. Notons $\tilde{\mu}$ son espérance et $\tilde{K}$ sa matrice de covariance. Pour $\eta \in \mathbb{R}^d$ fixé, nous définissons $Y = \eta \cdot X$ et nous calculons

$$\mathbf{E}(Y) = \eta \cdot \tilde{\mu}, \quad \mathrm{Var}(Y) = \mathbf{E}(\sum_{ij} \eta_i (X_i - \tilde{\mu}_i)\, \eta_j (X_j - \tilde{\mu}_j)] = {}^t\eta\tilde{K}\eta.$$

Comme par ailleur, on sait que Y suit une loi gaussienne, on a nécessairement $Y \sim \mathcal{N}(\eta \cdot \tilde{\mu}, {}^t\eta\tilde{K}\eta)$. D'après le calcul de la fonction caractéristique rappelé en remarque, on a donc

$$\varphi_Y(\xi) = \exp\left(i(\eta \cdot \tilde{\mu})\xi - \frac{1}{2}({}^t\eta\tilde{K}\eta)\xi^2\right).$$

On en déduit

$$\varphi_X(\eta) = \mathbf{E}\left(e^{iX\cdot\eta}\right) = \varphi_{\eta\cdot X}(1) = \varphi_Y(1),$$

ce qui donne bien l'expression annoncée.

Etape 2. Montrons que (2) implique X gaussien. Soient $\eta \in \mathbb{R}^d$ et $Y := \eta \cdot X$. D'après la formule (1), pour tout $\xi \in \mathbb{R}$, on a

$$\begin{aligned} \varphi_Y(\xi) &= \varphi_{{}^t\eta X}(\xi) = \varphi_X(\xi\eta) \\ &= \exp\left(-i(\eta \cdot \mu)\,\xi - \frac{{}^t\eta K\eta}{2}\xi^2\right). \end{aligned}$$

On reconnaît la transformée de Fourier de la loi gaussienne $\mathcal{N}(\eta \cdot \mu, {}^t\eta K\eta)$. Par injectivité de la transformation de Fourier, Y est gaussienne, de loi $\mathcal{N}(\eta \cdot \mu, {}^t\eta K\eta)$. Donc X est gaussien.

D'après l'étape 1, on sait également que φ_X vérifie l'identité (2) associée à $\tilde{\mu}$ l'espérance de X et $\tilde{K}$ la matrice de covariance de X. On en déduit imméditament que $D\varphi_X(0) = \mu = \tilde{\mu}$ et $D^2\varphi_X(0) = K = \tilde{K}$. Cela prouve bien que dans (2), μ et K correspondent à l'espérance et la matrice de covariance de X. □

Proposition 1.7. *Si X est un vecteur gaussien, ses composantes sont indépendantes si et seulement si sa matrice de covariance K_X est diagonale.*

Preuve de la Proposition 1.7. Lorsque $X_1, ..., X_d$ sont indépendantes, c'est un fait général que les covariances

$$\mathrm{cov}(X_j, X_k) = \mathbf{E}(X_j X_k) - \mathbf{E}(X_j)\mathbf{E}(X_k) = 0$$

sont nulles si $j \neq k$.
Réciproquement, si K est diagonale :

$$K = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_d^2 \end{pmatrix},$$

la fonction caractéristique de X vérifie

$$\varphi_X(\xi) = \prod_j \varphi_{X_j}(\xi_j).$$

Donc les X_j sont indépendantes. □

Exercice 1.8. *Donner un exemple de vecteur aléatoire (X, Y) non indépendant, non gaussien et tel que la matrice de covariance associée est diagonale.*

Théorème 1.9. *Soit X un vecteur gaussien de loi $\mathcal{N}_d(\mu, K)$. Il existe des réels $\lambda_1, ..., \lambda_d \geq 0$, des variables gaussiennes $Y_1, ..., Y_d$ indépendantes de lois respectives $\mathcal{N}_1(0, \lambda_1)$, ..., $\mathcal{N}_1(0, \lambda_d)$ et une matrice orthogonale $A \in O_d(\mathbb{R})$ tels que*

$$X = AY + \mu.$$

Preuve du Théorème 1.9. La matrice K étant une matrice de covariance, elle est symétrique positive (rappelons que ceci se voit par exemple en remarquant que ${}^t\xi K_X \xi = K_{{}^t\xi X} = \mathrm{var}\,({}^t\xi \cdot X) \geq 0$). Donc, il existe une matrice orthogonale A et une matrice diagonale Λ telles que

$$K = A\Lambda{}^tA.$$

Le vecteur

$$Y = {}^tA(X - \mu)$$

est gaussien, centré, et de covariance

$$K_Y = {}^tAKA = A^{-1}KA = \Lambda$$

(rappelons en effet que $K_{MX+a} = MK_X{}^tM$). Comme Λ est diagonale, les Y_j sont indépendantes. De plus, $X = AY + \mu$. □

Corollaire 1.10. *Un vecteur gaussien de covariance K_X admet une densité si et seulement si $\det K_X \neq 0$. Sa densité est alors*

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det K_X}} \exp\left(-\frac{1}{2}{}^t(x-\mu)K_X^{-1}(x-\mu_X)\right).$$

Preuve du Corollaire 1.10. Supposons d'abord $\det K_X \neq 0$. Reprenons les notations du théorème précédent : il existe un vecteur aléatoire gaussien Y de composantes indépendantes $Y_j \sim N_1(0, \lambda_j)$ et une matrice orthogonale A tels que $X = AY + \mu$.
Comme $Y = A^{-1}(X - \mu)$, $K_Y = A^{-1}K_XA$ donc $\det K_Y = \det K_X \neq 0$ et, puisque $K_Y = \mathrm{Diag}\,(\lambda_1, ..., \lambda_d)$, $\lambda_j > 0$ pour tout j. Donc, pour tout j, la loi de Y_j est

$$P_{Y_j} = \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{y_j^2}{2\lambda_j}\right) dy_j.$$

Comme les Y_j sont indépendantes, la loi conjointe de Y est le produit tensoriel de ses lois :

$$\begin{aligned} P_Y &= \bigotimes_j P_{Y_j} = \prod_j \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{y_j^2}{2\lambda_j}\right) dy_j \\ &= \frac{1}{(2\pi)^{d/2}\sqrt{\det K_Y}} \exp\left(-{}^t y K_Y^{-1} y\right) dy = f_Y(y)\, dy, \end{aligned}$$

où l'on a utilisé $K_Y = \mathrm{Diag}\,(\lambda_1^{-1}, ..., \lambda_d^{-1})$ et $\det K_Y = \prod_j \lambda_j$. La loi de Y étant connue, celle de $X = AY + \mu$ s'en déduit par un simple changement de variable : pour toute fonction réelle borélienne positive h sur $\mathbb{R}^d$,

$$\begin{aligned} \int h(X)\, dP &= \int h(AY + \mu)\, dP \\ &= \int h(Ay + \mu)\, dP_Y(y) = \int h(Ay + \mu)\, f_Y(y)\, dy \\ &= \int h(x)\, f_Y(A^{-1}(x - \mu))\, dx, \end{aligned}$$

où on a effectué le changement de variables $y = \varphi(x) = A^{-1}(x - \mu)$ de sorte que $|J\varphi(x)| = |\det A^{-1}| = 1$. On en déduit que X possède une densité, qui vaut $f_X(x) = f_Y(A^{-1}(X - \mu))$, d'où la formule voulue.

Réciproquement, supposons $\det K_X = 0$. Il existe $a \in \mathbb{R}^d$ tel que $K_X a = 0$. La variable aléatoire $Z = a \cdot X$ a pour espérance $a \cdot \mu_X$ et variance $\sigma_Z^2 = {}^t a K_X a = 0$. Donc $Z = a \cdot \mu$ p.p. Donc $a \cdot (X - \mu) = 0$ p.p. Soit H l'hyperplan

$$H = \mu + a^\perp = \{x \in \mathbb{R}^d,\ (x - \mu) \cdot a = 0\}.$$

Presque sûrement, $X \in H$ alors que $\lambda_d(H) = 0$. Si X avait une densité, on aurait

$$1 = P(X \in H) = \int_H f_X(x)\, dx = 0,$$

ce qui est absurde. □

Dans le cas dégénéré, la démonstration précédente montre que, si l'on suppose par exemple que $\sigma_1, ..., \sigma_r > 0$ tandis que $\sigma_{r+1} = ... = \sigma_d = 0$, et si V est le sous-espace vectoriel de dimension r image par A des r premiers vecteurs de la base canonique de $\mathbb{R}^n$, la loi de X est portée par V et X possède une densité relativement à la mesure de Lebesgue associée aux coordonnées d'une base orthogonale quelconque de V.

2 Conditionnement $\mathcal{L}^1$

Dans cette section, nous définissons la notion d'espérance conditionnelle dans un cadre $\mathcal{L}^1$. Une extension possible est de définir cette notion pour des va positives (sans condition d'intégrabilité). Etant donnée une sous-tribu $\mathcal{C} \subset \mathcal{A}$, on note $L^p(\mathcal{C}) := L^p(\Omega, \mathcal{C}, P)$.

Théorème 2.1. *Soient $\mathcal{B} \subset \mathcal{A}$ une sous-tribu et X une var $\mathcal{L}^1(\mathcal{A})$. Il existe une var $Y \in L^1(\mathcal{B})$ unique, notée $Y = \mathbf{E}(X|\mathcal{B})$ et appellée espérance conditionnelle de X par rapport à $\mathcal{B}$, telle que*

$$\mathbf{E}(Y\, Z) = \mathbf{E}(X\, Z) \qquad \forall\, Z = \mathbf{1}_B,\ B \in \mathcal{B},\ (\text{et donc } \ \forall\, Z \in L^\infty(\mathcal{B})). \tag{3}$$

Si $X \in \mathcal{L}^2(\mathcal{A})$ alors $Y \in L^2(\mathcal{B})$ est la solution du problème de minimisation

$$\mathbf{E}((Y - X)^2) = \min_{Z \in L^2(\mathcal{B})} \mathbf{E}((Z - X)^2). \tag{4}$$

Preuve du Théorème 2.1. Commeçons par l'unicité. Supposons qu'il existe deux var Y_1 et Y_2 de $L^1(\Omega, \mathcal{B}, P)$ telles que

$$\forall B \in \mathcal{B}, \quad \mathbf{E}(Y_1 \mathbf{1}_B) = \mathbf{E}(Y_2 \mathbf{1}_B) = \mathbf{E}(X \mathbf{1}_B).$$

En prenant $B := \{Y_1 - Y_2 > 0\} \in \mathcal{B}$, on trouve

$$\forall B \in \mathcal{B}, \quad \mathbf{E}((Y_1 - Y_2) \mathbf{1}_{\{Y_1 - Y_2 > 0\}}) = 0,$$

ce qui prouve $Y_1 - Y_2 \leq 0$ p.s.. Et on montre de la même manière $Y_1 - Y_2 \geq 0$ p.s..

Pour l'existence, nous commeçons par supposer $X \geq 0$. Soit alors Q la mesure définie sur $(\Omega, \mathcal{B})$ par

$$\forall B \in \mathcal{B}, \quad Q(B) := \int_B X dP.$$

On voit clairement que $Q << P$ sur $(\Omega, \mathcal{B})$. Le Théorème de Radon-Nikodym assure l'existence d'une variable aléatoire positive Y sur $(\Omega, \mathcal{B})$ telle que

$$\forall B \in \mathcal{B}, \quad Q(B) = \int_B Y dP,$$

ce qui est bien le résultat annoncé. Si de plus, $X \in L^1(\mathcal{A})$ alors $Y \in L^1(\mathcal{B})$, ce que l'on voit en prenant $B = \Omega$.

Si maintenant, on suppose juste $X \in \mathcal{L}^1(\mathcal{A})$, on écrit $X = X_+ - X_-$ avec $0 \leq X_\pm \in \mathcal{L}^1(\mathcal{A})$ et l'étape précédente nous dit qu'il existe $0 \leq Y_\pm \in L^1(\mathcal{B})$ telles que $\mathbf{E}(Y_\pm \mathbf{1}_B) = \mathbf{E}(X_\pm \mathbf{1}_B)$. On en déduit que $Y := Y_+ - Y_-$ convient.

On suppose enfin $X \in \mathcal{L}^2(\mathcal{A})$. En prenant $Z := Y \mathbf{1}_{|Y| \leq n} \in L^\infty(\mathcal{B})$ dans (3), on obtient

$$\mathbf{E}(Y^2 \mathbf{1}_{|Y| \leq n}) = \mathbf{E}(Y (Y \mathbf{1}_{|Y| \leq n})) = \mathbf{E}(X Y \mathbf{1}_{|Y| \leq n}) \leq \mathbf{E}(X^2)^{1/2} \mathbf{E}(Y^2 \mathbf{1}_{|Y| \leq n})^{1/2},$$

et donc

$$\mathbf{E}(Y^2) = \lim_{n \to \infty} \mathbf{E}(Y^2 \mathbf{1}_{|Y| \leq n}) \leq \mathbf{E}(X^2).$$

De plus, pour tout $Z \in L^2(\mathcal{B})$, on a

$$\begin{aligned} \mathbf{E}((X - Z)^2) &= \mathbf{E}((X - Y)^2) + 2\mathbf{E}((X - Y)(Y - Z)) + \mathbf{E}((Y - Z)^2) \\ &= \mathbf{E}((X - Y)^2) + \mathbf{E}((Y - Z)^2), \end{aligned}$$

puisque $Y - Z$ est $\mathcal{B}$ mesurable. On en déduit immédiatement (4). $\square$

Les principales propriétés de l'espérance conditionnelle sont résumées ci-dessous

Théorème 2.2. *Soient $\mathcal{B},\mathcal{C} \subset \mathcal{A}$ deux sous-tribus et X, Y deux var $\mathcal{L}^1(\mathcal{A})$. Alors*

- $\mathbf{E}(\lambda\, X + Y|\mathcal{B}) = \lambda\, \mathbf{E}(X|\mathcal{B}) + \mathbf{E}(Y|\mathcal{B})$; (5)
- $X \le Y \quad \Rightarrow \quad \mathbf{E}(X|\mathcal{B}) \le \mathbf{E}(Y|\mathcal{B})$; (6)

- $\mathbf{E}(\mathbf{E}(X|\mathcal{B})) = \mathbf{E}(X), \quad \mathbf{E}(1|\mathcal{B}) = 1$; (7)
- $\mathbf{E}(X|\mathcal{B}) = \mathbf{E}(X) \quad$ si $\quad X \perp\!\!\!\perp \mathcal{B}$; (8)
- $\mathbf{E}(X\,Y|\mathcal{B}) = Y\,\mathbf{E}(X|\mathcal{B}),\ \mathbf{E}(Y|\mathcal{B}) = Y \quad$ si $\quad Y$ est $\mathcal{B}$-*mesurable*; (9)
- $\mathbf{E}(\mathbf{E}(X|\mathcal{C})|\mathcal{B}) = \mathbf{E}(X|\mathcal{C}) \quad$ si $\quad \mathcal{C} \subset \mathcal{B}$; (10)

- $\varphi(\mathbf{E}(X|\mathcal{B})) \le \mathbf{E}(\varphi(X)|\mathcal{B}) \quad$ si φ est une fonction convexe, (11)

ce qui implique $|\mathbf{E}(X|\mathcal{B})|^p \le \mathbf{E}(|X|^p|\mathcal{B})$ *et* $\mathbf{E}(X|\mathcal{B}) \in L^p$ *si* $X \in L^p$; (12)
- $\mathbf{E}(X_n|\mathcal{B}) \to \mathbf{E}(X|\mathcal{B})$ en norme $L^p \quad$ si $\quad X_n \to X$ en norme L^p; (13)

- $\mathbf{E}(X_n|\mathcal{B}) \nearrow \mathbf{E}(X|\mathcal{B}) \quad$ si $\quad X_n \nearrow X\;$ p.s.; (14)
- $\mathbf{E}(\liminf X_n|\mathcal{B}) \le \liminf \mathbf{E}(X_n|\mathcal{B}) \quad$ si $\quad X_n \ge 0$; (15)
- $\mathbf{E}(X_n|\mathcal{B}) \to \mathbf{E}(X|\mathcal{B})\;$ p.s. et $L^1 \quad$ si $\quad X_n \to X\;$ p.s. et $|X_n| \le Z \in L^1$. (16)

Preuve du Théorème 2.2. (5) En notant $X' = \mathbf{E}(X|\mathcal{B})$ et $Y' = \mathbf{E}(Y|\mathcal{B})$, pour tout $Z \in L^\infty(\mathcal{B})$, on a

$$\mathbf{E}((\lambda X' + Y')Z) = \lambda \mathbf{E}(X'Z) + \mathbf{E}(Y'Z) = \lambda \mathbf{E}(XZ) + \mathbf{E}(YZ) = \mathbf{E}((\lambda X + Y)Z).$$

(6) Si $X \ge 0$, en notant $Y := \mathbf{E}(X|\mathcal{B})$ et $B := \{Y \le 0\}$, on a

$$0 \le \mathbf{E}(X\mathbf{1}_B) = \mathbf{E}(Y\mathbf{1}_B) \le 0 \text{ et } Y\mathbf{1}_B \le 0,$$

de sorte que $Y\mathbf{1}_B = 0$ et donc $Y \ge 0$.
(7) On a $\mathbf{E}(\mathbf{E}(X|\mathcal{B})) = \mathbf{E}(X)$ en prenant $Z = 1$ dans la définition (3) de l'espérance conditionnelle. L'identité $\mathbf{E}(1|\mathcal{B}) = 1$ est immédiate.
(8) Si $X \perp\!\!\!\perp \mathcal{B}$, on a $\mathbf{E}(XZ) = \mathbf{E}(X)\mathbf{E}(Z) = \mathbf{E}(\mathbf{E}(X)Z)$ pour tout $Z \in L^\infty(\mathcal{B})$, ce qui prouve bien $\mathbf{E}(X|\mathcal{B}) = \mathbf{E}(X)$.
(9) Si $Y \in L^\infty(\mathcal{B})$, pour $Z \in L^\infty(\mathcal{B})$, on a

$$\mathbf{E}(\mathbf{E}(X|\mathcal{B})YZ) = \mathbf{E}(XYZ) = \mathbf{E}(\mathbf{E}(XY|\mathcal{B})Z),$$

ce qui prouve $\mathbf{E}(X\,Y|\mathcal{B}) = Y\,\mathbf{E}(X|\mathcal{B})$. En particulier, en prenant $X = 1$, on a $\mathbf{E}(Y|\mathcal{B}) = Y$.
(10) Si $\mathcal{C} \subset \mathcal{B}$, alors $\mathbf{E}(X|\mathcal{C})$ est $\mathcal{B}$-mesurable, et donc $\mathbf{E}(\mathbf{E}(X|\mathcal{C})|\mathcal{B}) = \mathbf{E}(X|\mathcal{C})$ d'après (9).
(11) Notons

$$A_\varphi := \{(a,b) \in \mathbb{R}^2;\, \forall\, x \in \mathbb{R},\ \varphi(x) \ge ax + b\}.$$

Il est clair que

$$\forall\, x \in \mathbb{R}, \quad \varphi(x) = \sup_{(a,b)\in A_\varphi} (ax+b) = \sup_{(a,b)\in A_\varphi\cap\mathbb{Q}^2} (ax+b).$$

Du fait que $A_\varphi \cap \mathbb{Q}^2$ est dénombrable, on en déduit que p.s.

$$\mathbf{E}[\varphi(X)|\ \mathcal{B}] = \mathbf{E}\big[\sup_{(a,b)\in A_\varphi\cap\mathbb{Q}^2} (aX+b)|\ \mathcal{B}\big] \ge \sup_{(a,b)\in A_\varphi\cap\mathbb{Q}^2} \mathbf{E}[aX+b|\ \mathcal{B}] = \varphi(\mathbf{E}[X|\mathcal{B}]).$$

L'assertion (12) s'en déduit immédiatement.
(13) De (12), on déduit

$$\mathbf{E}(|\mathbf{E}(X_n|\mathcal{B}) - \mathbf{E}(X|\mathcal{B})|^p) = \mathbf{E}(|\mathbf{E}(X_n - X|\mathcal{B})|^p) \le \mathbf{E}(|X_n - X|^p) \to 0,$$

si $X_n \to X$ dans L^p.
(14) Si $X_n \nearrow X$ p.s. on a $Y_n := \mathbf{E}(X_n|\mathcal{B}) \nearrow Y$ p.s. avec Y $\mathcal{B}$-mesurable. D'après le théorème de convergence monotone, pour tout $B \in \mathcal{B}$, on a

$$\mathbf{E}(Y\mathbf{1}_B) = \lim \mathbf{E}(Y_n\mathbf{1}_B) = \lim \mathbf{E}(X_n\mathbf{1}_B) = \mathbf{E}(X\mathbf{1}_B),$$

ce qui dit bien que $Y = \mathbf{E}(X|\mathcal{B})$. Les assertions (15) et (16) s'en déduisent comme dans le chapitre sur les théorèmes de passage à la limite. □

Exercice 2.3. *Faire la démonstration détaillée de (15) et (16).*

On peut préciser le point (8).

Théorème 2.4. *Deux sous tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes si, et seulement si,*

$$\mathbf{E}(X|\mathcal{B}_1) = \mathbf{E}(X) \text{ pour toute var } X\ \mathcal{B}_2\text{-mesurable.}$$

Preuve du Théorème 2.4. Le sens direct n'est rien d'autre que (8). Montrons la réciproque et supposons que

$$\mathbf{E}(X|\mathcal{B}_1) = \mathbf{E}(X) \text{ pour toute var } X\ \mathcal{B}_2\text{-mesurable.}$$

Pour $Z \in L^\infty(\mathcal{B}_1)$ on a donc

$$\mathbf{E}(XZ) = \mathbf{E}(\mathbf{E}(X|\mathcal{B}_1)Z) = \mathbf{E}(\mathbf{E}(X)Z) = \mathbf{E}(X)\mathbf{E}(Z),$$

ce qui est bien dire que $\mathcal{B}_1$ et $\mathcal{B}_2$ sont indépendantes. □

Définition 2.5. *Soit X, Y deux var. On appelle espérance conditionnelle de Y sachant X, la var $\mathbf{E}(Y|X) := \mathbf{E}(Y|\sigma(X))$.*

D'après le Théorème I.4.10, il existe une fonction φ telle que $\mathbf{E}(Y|X) = \varphi(X)$. On peut donc voir l'espérance conditionnelle comme une façon de definir une "meilleur fonction" φ de sorte que $\varphi(X)$ "soit proche" de Y.

Corollaire 2.6. *Deux variables aléatoires X et Y sont indépendantes si, et seulement si,*

$$\mathbf{E}(h(X)|Y) = \mathbf{E}(h(X)), \quad \forall\, h \,\text{mesurables.}$$

Théorème 2.7. *Soit $\mathcal{B}$ une sous-tribu et soient X et Y deux variables aléatoires telles que X est indépendante de $\mathcal{B}$ et Y est $\mathcal{B}$-mesurable. Pour toute fonction mesurable et bornée φ, on a*

$$\mathbf{E}(\varphi(X,Y)|\mathcal{B}) = \Phi(Y), \quad \Phi(y) := \int \varphi(x,y) P_X(dx),$$

où P_X désigne la loi de X.

Preuve du Théorème 2.7. Soit $Z \in \mathcal{L}^\infty(\mathcal{B})$ et notons $P_{(X,Y,Z)}$ la loi du triplet (X,Y,Z). Comme X est indépendante de (Y,Z), on a $P_{(X,Y,Z)} = P_X \otimes P_{(Y,Z)}$. Grâce au théorème de Fubini, on a

$$\begin{aligned}
\mathbf{E}(\varphi(X,Y)Z) &= \int \varphi(x,y) z P_X \otimes P_{(Y,Z)}(dx,dy,dz) \\
&= \int z\Big(\int \varphi(x,y)P_X(dx)\Big) P_{(Y,Z)}(dy,dz) \\
&= \int z\Phi(Y)P_{(Y,Z)}(dy,dz) = \mathbf{E}(\Phi(Y)Z).
\end{aligned}$$

Cela démontre bien le résultat annoncé. □

Exercice 2.8. *Soient $\mathcal{B}$ et $\mathcal{C}$ deux sous-tribus et X une var $\mathcal{L}^1$. On suppose $\mathcal{C} \perp\!\!\!\perp \sigma(\sigma(X),\mathcal{B})$.*
a) Montrer que $\mathbf{E}(X\mathbf{1}_{B\cap C}) = \mathbf{E}(X\mathbf{1}_B)\mathbf{P}(C)$, pour tout $B \in \mathcal{B}$ et $C \in \mathcal{C}$.
b) Montrer que $\mathcal{D} := \{B \cap C;\ B \in \mathcal{B},\ C \in \mathcal{C}\}$ est une algèbre et que $\sigma(\mathcal{D}) = \sigma(\mathcal{B},\mathcal{C})$.
c) En déduire que $\mathbf{E}(X|\sigma(\mathcal{B},\mathcal{C})) = \mathbf{E}(X|\sigma(\mathcal{B}))$.

3 Calculs d'espérance conditionnelle

• *Le cas discret.* Soit X une va à valeurs dans un espace dénombrable E et soit $Y \in \mathcal{L}^1(\mathcal{A})$. On introduit l'ensemble

$$E' := \{x \in E;\ P(X = x) > 0\},$$

la fonction

$$\varphi(x) := \frac{\mathbf{E}(Y\,\mathbf{1}_{X=x})}{P(X = x)} \text{ si } x \in E', \ := 0 \text{ sinon},$$

et enfin la var $\varphi(X)$. Pour $Z := \mathbf{1}_{X=a}$, $a \in E'$, on calcule

$$\begin{aligned}
\mathbf{E}(\varphi(X)Z) &= \mathbf{E}\big(\sum_{x \in E'} \varphi(x')\mathbf{1}_{X=x'}\,\mathbf{1}_{X=a}\big)\\
&= \varphi(a)\,\mathbf{E}\big(\mathbf{1}_{X=a}\big) = \mathbf{E}(Y\,\mathbf{1}_{X=a}) = \mathbf{E}(YZ).
\end{aligned}$$

Comme E est discret, cela suffit pour dire que cette identité est vraie pour tout Z qui est $\sigma(X)$-mesurable. On vient donc de démontrer que $E(Y|X) = \varphi(X)$.

Exemple. On prend $\Omega := \{1, \dots, 6\}$ et $P(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$. On définit

$$X(\omega) := \begin{cases} 1 & \text{si } \omega \text{ est impair,}\\ 0 & \text{si } \omega \text{ est pair,}\end{cases}$$

et $Y(\omega) = \omega$. Alors

$$\mathbf{E}(Y|X)(\omega) = \varphi(X(\omega)) = \begin{cases} 3 & \text{si } \omega \text{ est impair,}\\ 4 & \text{si } \omega \text{ est pair,}\end{cases}$$

puisque par exemple

$$\varphi(1) = \frac{\mathbf{E}(\sum_\omega \omega\,\mathbf{1}_{X(\omega)=1})}{P(X = 1)} = \frac{\sum_{\omega \text{ impair}} \omega P(\{\omega\})}{1/2} = \frac{1+3+5}{3}.$$

• *Le cas continu.* Soient X et Y deux v.a.r. telles que le couple (X, Y) a pour densité $f_{X,Y}(x, y)$. On observe en particulier que la var Y a pour densité

$$f_Y(y) := \int_{\mathbb{R}} f_{X,Y}(x, y)\,dy$$

et on pose

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x, y)}{f_Y(y)}\,\mathbf{1}_{f_Y(y)>0}.$$

Pour tout $\varphi, \psi \in \mathcal{L}^\infty(\mathbb{R})$, on calcule

$$\begin{aligned}
\mathbf{E}(\varphi(X)\psi(Y)) &= \int_{\mathbb{R}^2} \varphi(x)\psi(y)\,f_{X,Y}(x, y)\,dxdy\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x)\,f_{X|Y}(x|y)dx\,\psi(y)f_Y(y)\mathbf{1}_{f_Y(y)>0}dy\\
&= \int_{\mathbb{R}} \Phi(y)\,\psi(y)f_Y(y)\,dy = \mathbf{E}(\Phi(Y)\psi(Y)),
\end{aligned}$$

où on a posé

$$\Psi(y) := \int_{\mathbb{R}} \varphi(x)f_{X,Y}(x, y)\,dx.$$

On a ainsi démontré que $\mathbf{E}(\varphi(X)|Y) = \Phi(Y)$.

• *Le cas gaussien.* Soit $(X, Y_1, \dots, Y_n)$ un vecteur gaussien centré. Alors $\mathbf{E}(X|Y_1, \dots, Y_n)$ coïncide avec la projection orthogonale de X sur le sous-espace vectoriel engendré par $Y_1, \dots, Y_n$. Il existe donc $\lambda_1, \dots, \lambda_n$ tels que

$$\mathbf{E}[X|Y_1, \dots, Y_n] = \sum_{j=1}^{n} \lambda_j Y_j =: m_Y.$$

De plus, pour toute fonction φ mesurable et bornée, on a

$$\mathbf{E}[\varphi(X)|Y_1, \dots, Y_n] = \int \varphi(x) g_{m_Y, \sigma_Y^2}(x)\, dx$$

avec g_{m,σ^2} la densité de la loi $\mathcal{N}(m, \sigma^2)$ et

$$\sigma_Y^2 := \mathbf{E}[(X - m_Y)^2].$$

Preuve. On note $m_Y := \sum \lambda_j Y_j$ la projection orthogonale de X sur $Y_1, \dots, Y_n$. Pour tout $j \in \{1, \dots, n\}$, on a

$$\mathrm{cov}(X - m_Y, Y_j) = \mathbf{E}[(X - m_Y) Y_j] = 0,$$

par définition de la projection orthogonale. Puisque le vecteur $(Y_1, \dots, Y_n, X - m_Y)$ est gaussien centré, il en résulte que la va $X - m_Y$ est indépendante des $Y_1, \dots, Y_n$. On en déduit

$$\mathbf{E}[X|Y_1, \dots, Y_n] = \mathbf{E}[X - m_Y|Y_1, \dots, Y_n] + m_Y = \mathbf{E}[X - m_Y] + m_Y = m_Y,$$

et la première assertion. Pour la dernière assertion, ; notons $Z := X - m_Y$, de sorte que Z est indépendante de $(Y_1, \dots, Y_n)$ et suit la loi $\mathcal{N}(0, \sigma_Y^2)$. D'après le Théorème 2.7, on a alors

$$\mathbf{E}(\varphi(X)|Y_1, \dots, Y_n] = \mathbf{E}(\varphi(m_Y + Z)|Y_1, \dots, Y_n] = \int \varphi(m_Y + z) P_Z(dz).$$

On conclut grâce à un changement de variables. □