

Rappels sur les intégrales multiples

Théorème 1 (Fubini-Tonelli, cas $n=2$) Soit $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction positive, alors

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$$

Où les trois termes sont ou bien nuls et égaux ou bien simultanément non nuls. Si f est de signe quelconque mais intégrable au sens que $\int_{\mathbb{R}^2} |f(x, y)| dx dy < +\infty$ alors l'égalité des trois intégrales reste vraie.

Exemple 2. $f(x, y) = x e^{xy} \mathbf{1}_{x > 0} \mathbf{1}_{1 < y < 2}$. D'une part

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) dx dy &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x e^{xy} \mathbf{1}_{x > 0} \mathbf{1}_{1 < y < 2} dx \right) dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y^2 x e^{xy} \mathbf{1}_{x > 0} dx \right) \frac{\mathbf{1}_{1 < y < 2}}{y^2} dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x e^{xy} \mathbf{1}_{x > 0} dx \right) \frac{\mathbf{1}_{1 < y < 2}}{y^2} dy \\ &= \int_1^2 \frac{dy}{y^2} = \frac{1}{2}. \end{aligned}$$

D'autre part

$$\begin{aligned} \int_{\mathbb{R}^2} f(x, y) dx dy &= \int_0^{\infty} x \left(\int_1^2 e^{xy} dy \right) dx \\ &= \int_0^{\infty} (e^{2x} - e^x) dx = \frac{1}{2}. \end{aligned}$$

Exemple 3. Voyons un contre-exemple à l'utilisation de Fubini dans un cas où l'intégrale double n'est pas définie. Soit

$$f(x, y) = \frac{x^2 y^2}{(x^2 + y^2)^2}$$

alors $I = \int_0^1 \int_0^1 f(x, y) dx dy$ n'est pas bien défini car

$$\begin{aligned} \int_0^1 \int_0^1 f(x, y) dx dy &= \int_0^1 \left(\int_0^y \frac{y^2 x^2}{(x^2 + y^2)^2} dx \right) dy + \int_0^1 \left(\int_y^1 \frac{x^2 y^2}{(x^2 + y^2)^2} dy \right) dx \\ &= \int_0^1 \frac{1}{y} \frac{1}{1 + y^2} dy + \int_0^1 \frac{1}{y} \frac{1}{1 + y^2} dy = +\infty \end{aligned}$$

Or, les intégrales

$$I_1 = \int_0^1 \int_0^1 \frac{x^2 y^2}{(x^2 + y^2)^2} dx dy, \quad I_2 = \int_0^1 \int_0^1 \frac{x^2 y^2}{(x^2 + y^2)^2} dy dx$$

sont bien définies et il satisfait $I_1 = I_2$. En effet:

$$\int_0^1 \frac{x^2 y^2}{(x^2 + y^2)^2} dy = \int_0^1 \frac{1}{(x^2 + y^2)} dy = \int_0^1 \frac{y^2}{(x^2 + y^2)^2} dy = \frac{1}{1 + x^2}$$

et alors

$$I_2 = \int_0^1 \frac{1}{1+x^2} dx = [\arctan(x)]_0^1 = \frac{\pi}{4} \neq 0$$

ce qui est en contradiction avec une application naïve de Fubini (car dans ce cas $I_2 = I = 0$).

Vecteurs aléatoires à densité

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Un vecteur aléatoire X de dimension n (ou dans \mathbb{R}^n) est une application $X: \Omega \rightarrow \mathbb{R}^n$ telle que tous les ensembles de la forme $\{X \in B\} = \{\omega : X(\omega) \in B\}$ pour B borélien de \mathbb{R}^n appartiennent à la tribu \mathcal{A} . En particulier on peut calculer la probabilité $P(X \in B)$ de l'événement $\{X \in B\}$ (car $P(A)$ est définie seulement pour $A \in \mathcal{A}$). La loi de X est la loi d'application $f_X: B(\mathbb{R}^n) \rightarrow [0, 1]$ qui à tout borélien B de \mathbb{R}^n associe $P(X \in B)$. On appelle fonction de répartition de X la fonction $F_X: \mathbb{R}^n \rightarrow [0, 1]$ définie par

$$F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

où $(X_i)_{i=1, \dots, n}$ sont les composantes de $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ (donc des v.a. réelles). La fonction de répartition caractérise la loi de X , i.e. n'importe quel événement B peut être calculé à l'aide de

Exemple 4. Soit $n = 2$ et $B = [x_1, y_1] \times [x_2, y_2]$ alors il est facile de vérifier que

$$P(X \in B) = P(X_1 \in [x_1, y_1], X_2 \in [x_2, y_2]) = F_X(y_1, y_2) - F_X(x_1, y_2) - F_X(y_1, x_2) + F_X(x_1, x_2)$$

en utilisant les propriétés élémentaires des probabilités (en particulier $P(A \cap B) = P(A) + P(B) - P(A \cup B)$).

Définition 5. On dit que X admet une densité $f_X: \mathbb{R}^n \rightarrow \mathbb{R}_+$ ssi pour tout borélien B de \mathbb{R}^n (la tribu borélienne de \mathbb{R}^n) on peut exprimer la probabilité de l'événement $\{X \in B\}$ par une intégrale sur B de f_X :

$$P(X \in B) = \int_B f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

La densité, si elle existe, est unique et caractérise la loi de X . On a que

$$\int_{\mathbb{R}^n} f_X(x_1, \dots, x_n) dx_1 \dots dx_n = P(X \in \mathbb{R}^n) = 1$$

en particulier f_X est intégrable. La fonction de répartition $F_X: \mathbb{R}^n \rightarrow [0, 1]$ de X est donnée par

$$F_X(t_1, \dots, t_n) \stackrel{\text{def}}{=} P(X_1 \leq t_1, \dots, X_n \leq t_n) = \int_{\mathbb{R}^n} \mathbf{1}_{\{x_1 \leq t_1, \dots, x_n \leq t_n\}} f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

C'est la probabilité de l'événement pour $B =]-\infty, t_1] \times \dots \times]-\infty, t_n]$ de \mathbb{R}^n . On peut déterminer la densité en dérivant la fonction de répartition:

$$f_X(t_1, \dots, t_n) = \frac{\partial^n F_X(t_1, \dots, t_n)}{\partial t_1 \dots \partial t_n}$$

formule valable en tout point de continuité de $f_X(t_1, \dots, t_n) / t_1 \dots t_n$.

L'interprétation intuitive de la densité est la suivante: si $x_i \rightarrow 1$ alors la probabilité de li appartenir à l'intervalle $[x_i, x_i + x_i]$ pour $i = 1, \dots, n$ est approchable par

$$P(X_i \in [x_i, x_i + x_i] \text{ pour } i = 1, \dots, n) = \int_{x_1}^{x_1 + x_1} \dots \int_{x_n}^{x_n + x_n} f_X(t_1, \dots, t_n) dt_1 \dots dt_n \\ f_X(x_1, \dots, x_n) x_1 \dots x_n.$$

La densité est donc proportionnelle à la mesure de probabilité d'un petit voisinage du point (x_1, \dots, x_n) . Autrement dit, si $B_n(x, \epsilon) = \{y \in \mathbb{R}^n : |x - y| < \epsilon\}$ \mathbb{R}^n est la boule n -dimensionnelle de rayon ϵ centrée en \mathbb{R}^n et $V_n(\epsilon) = |B_n(x, \epsilon)|$ le volume de $B(x, \epsilon)$, i.e. $V_n(\epsilon) = \int_{B_n(x, \epsilon)} dt_1 \dots dt_n$ alors si $\epsilon \rightarrow 0$ on a l'approximation $P(X \in B(x, \epsilon)) \approx f_X(x) V_n(\epsilon)$.

Exemple 6. Soit $Z = (X, Y) : \mathbb{R}^2$ un couple aléatoire dont la fonction de répartition est donnée par

$$F_Z(x, y) = q(x) q(y)$$

où $q(s) = \max(0, \min(s, 1))$. Alors

$$\frac{\partial^2}{\partial x \partial y} F_Z(x, y) = \begin{cases} \text{non définie} & \text{si } x = 0, 1 \text{ ou } y = 0, 1 \\ 1 & \text{si } (x, y) \in]0, 1[^2 \\ 0 & \text{autrement} \end{cases}$$

et on peut vérifier que

$$F_Z(x, y) = \int_0^x \int_0^y I_{[0,1]}(z_1) I_{[0,1]}(z_2) dz_1 dz_2.$$

Donc $f_Z(z_1, z_2) = I_{[0,1]}(z_1) I_{[0,1]}(z_2)$ est la densité de Z .

Définition 7. Soit $D \subset \mathbb{R}^n$ un ensemble mesurable et tel que $\text{Vol}(D) = \int_D dx > 0$ (son volume est positif). On dit que $X : \mathbb{R}^n$ a une loi uniforme sur D si X admet densité

$$f_X(x) = \frac{I_D(x)}{\text{Vol}(D)}.$$

Densités marginales

Définition 8. Si Z est un vecteur aléatoire dans \mathbb{R}^n admettant une densité f_Z alors tout sous-vecteurs Y de Z de dimension $k < n$ admettent une densité qui s'obtient en intégrant par rapport aux composantes qui ne figurent pas dans Y . On appelle cette densité la densité marginale de Y . Explicitement si $Y = (Z_1, \dots, Z_k)$ alors

$$P(Y \in B) = P((Z_1, \dots, Z_k) \in B) = P(Z \in B \times \mathbb{R}^{n-k}) = \int_{(z_1, \dots, z_k) \in B} f_Z(z_1, \dots, z_n) dz_1 \dots dz_n \\ = \int_B \int_{\mathbb{R}^{n-k}} f_Z(z_1, \dots, z_k, z_{k+1}, \dots, z_n) dz_{k+1} \dots dz_n dz_1 \dots dz_k$$

donc $f_Y(y_1, \dots, y_k) = \int_{\mathbb{R}^{n-k}} f_Z(y_1, \dots, y_k, z_{k+1}, \dots, z_n) dz_{k+1} \dots dz_n$.

Cas particulier ($n = 2$). Soit $Z = (X, Y)$ un vecteur aléatoire bidimensionnel de densité $f_Z(x, y)$. La densité marginale de X est $f_X(x) = \int_{\mathbb{R}} f_Z(x, y) dy$ et la densité marginale de Y est $f_Y(y) = \int_{\mathbb{R}} f_Z(x, y) dx$.

Exemple 9. Considérons le couple (X, Y) de densité

$$f_{(X,Y)}(x,y) = \frac{e^{\frac{xy}{X}}}{2\bar{X}} \mathbf{I}_{0 < x < y^2} \mathbf{I}_{y > 0}$$

Il est facile de vérifier que $\mathcal{F}_{(X,Y)}$ est un \mathbb{C} -module et que $\mathcal{F}_{(X,Y)} \neq 0$ si et seulement si X et Y sont linéairement indépendants. On définit alors la *fonction de Green* $G_{(X,Y)}$ de (X,Y) par :

• Déterminer les densités marginales f_X et f_Y .

Exercice 10 Calculer $P(X > 1)$.

Calculons

$$I = \int_{\mathbb{R}^2} f_{(X,Y)}(x,y) dx dy = \int_0^y \int_0^x \frac{dx}{2x} e^{\frac{y}{2x}} dy = \int_0^y y e^{\frac{y}{2x}} dy =$$

donc $\quad = 1$.

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y) dy = \frac{I_{x>0}}{2} \frac{1}{\bar{x}} e^{\frac{y}{\bar{x}}} dy = \frac{e^{\frac{x}{\bar{x}}} \bar{x}}{2} \frac{1}{\bar{x}} I_{x>0}$$

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y) dx = \mathbb{I}_{y>0} e^{\frac{y^2}{2\bar{X}}} \int_0^y \frac{dx}{2\bar{X}} = y e^{\frac{y^2}{2\bar{X}}} \mathbb{I}_{y>0}$$

$$P(X > 1) = \int_1^{\infty} f_X(x) dx = \int_1^{\infty} \frac{e^{-x/2}}{2} dx = \frac{1}{e}$$

Exemple 10. Deux densités $f_{X,Y}(x,y)$ et $g_{X,Y}(x,y)$ peuvent avoir les mêmes marginales. Par exemple il est facile de montrer que les densités

$$f_{X,Y}(x,y) = \frac{1}{2} e^{\mathbb{P}_d(x^2+y^2)/2}, \quad g_{X,Y}(x,y) = \frac{1}{2} e^{\mathbb{P}_d(x^2+y^2)/2} [1 + xy \mathbb{I}_{[\mathbb{P}_d,1]}(x) \mathbb{I}_{[\mathbb{P}_d,1]}(y)]$$

ont les mêmes marginales ($f_X = g_X$ et $f_Y = g_Y$). En effet, en utilisant l'intégrale remarquable

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx = 1$$

on obtient que

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y) dy = \frac{1}{2} e^{-\frac{1}{2} x^2}$$

et

$$\begin{aligned} g_X(x) &= \int_{\mathbb{R}} g_{X,Y}(x,y) dy = \int_{\mathbb{R}} \frac{1}{2} e^{\mathbb{E}_0(x^2+y^2)/2} [1 + xy I_{[\mathbb{E}_0 d, 1]}(x) I_{[\mathbb{E}_0 d, 1]}(y)] dy \\ &= \frac{1}{2} e^{\mathbb{E}_0 x^2/2} \int_{\mathbb{R}} \frac{1}{2} e^{\mathbb{E}_0 y^2/2} [1 + xy I_{[\mathbb{E}_0 d, 1]}(x) I_{[\mathbb{E}_0 d, 1]}(y)] dy = \frac{1}{2} e^{\mathbb{E}_0 x^2/2} \end{aligned}$$

car

$$\int_{\mathbb{R}} \frac{1}{2} e^{-y^2/2} y I_{[0,1]}(y) dy = 0$$

par symétrie.

Densité et espérance conditionnelle

Définition 11. Soit $Z = (X, Y)$ un vecteur aléatoire dans $\mathbb{R}^m \times \mathbb{R}^n$ admettant une densité $f_{Z, \mathbb{R}^m \times \mathbb{R}^n}$. Soient f_X et f_Y les densités marginales des vecteurs X et Y . On appelle densité conditionnelle de X sachant $Y = y$ la densité donnée par

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} \quad \text{pour tout } y \in \mathbb{R}^n \text{ t.q. } f_Y(y) > 0.$$

Cette définition est motivée par le fait que, si $A \subset \mathbb{R}^m$ et $B \subset \mathbb{R}^n$:

$$P(X \in A | Y \in B) = \frac{P(X \in A, Y \in B)}{P(Y \in B)} = \frac{\int_A \int_B f_{(X,Y)}(x, y) V_m(\cdot) V_n(\cdot) dy dx}{\int_B f_Y(y) V_n(\cdot) dy}$$

$$\frac{f_{(X,Y)}(x, y)}{f_Y(y)} V_m(\cdot) = f_{X|Y=y}(x) V_m(\cdot)$$

donc la densité conditionnelle est proportionnelle à la probabilité conditionnelle de trouver dans une petite boule centrée en x sachant que Y est dans une petite boule centrée en y .

Exemple 12. Considérons $Z = (X, Y)$ de densité $f_{Z, \mathbb{R}^2}(x, y) = 2^{-2} e^{-\frac{1}{2}(x+y)} I_{0 < x < y}$. Quelle est la densité conditionnelle de X sachant $Y = y$?

Calculons d'abord $f_Y(y)$:

$$f_Y(y) = \int_{\mathbb{R}} 2^{-2} e^{-\frac{1}{2}(x+y)} I_{0 < x < y} dx = 2^{-2} e^{-\frac{1}{2}y} I_{y > 0} \int_0^y e^{-\frac{1}{2}x} dx = 2^{-2} e^{-\frac{1}{2}y} (1 - e^{-\frac{1}{2}y}) I_{y > 0}$$

Il vient que

$$f_{X|Y=y}(x) = \frac{2^{-2} e^{-\frac{1}{2}(x+y)} I_{0 < x < y}}{2^{-2} e^{-\frac{1}{2}y} (1 - e^{-\frac{1}{2}y}) I_{y > 0}} = \frac{e^{-\frac{1}{2}x} I_{0 < x < y}}{1 - e^{-\frac{1}{2}y}} \quad \text{pour tout } y > 0.$$

Définition 13. Une famille $(X_i)_{i=1, \dots, n}$ de v.a. est indépendante ssi pour tout $B_i, i = 1, \dots, n$, on a que les événements $\{X_i \in B_i\}_{i=1, \dots, n}$ sont indépendants, i.e. :

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n).$$

Dans cette définition les v.a.s X_i peuvent être réelles ou bien des vecteurs aléatoires elles-mêmes. Les v.a. X, Y sont indépendantes ssi $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$. Pour les v.a. avec densité on a la proposition suivante.

Proposition 14. Soient X et Y 2 v.a. admettant respectivement les densités f_X et f_Y . Alors X et Y sont indépendantes ssi $f_{X|Y=y}$ ne dépend de y . Dans ce cas li $f_{X|Y=y}(x) = f_X(x)$.

Démonstration. Si X, Y sont indépendantes alors $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ et donc on a que le couple admet la densité jointe $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ car

$$\frac{\partial^2}{\partial x \partial y} f_{(X,Y)}(x, y) = \frac{\partial^2}{\partial x \partial y} F_X(x)F_Y(y) = f_X(x) f_Y(y)$$

et donc $f_{X|Y=y}(x) = f_{(X,Y)}(x, y)/f_Y(y) = f_X(x)$ qui ne dépend pas de y . Réciproquement on a $f_{(X,Y)}(x, y) = f_{X|Y=y}(x) f_Y(y)$ et si la densité conditionnelle ne dépend pas de

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

et donc $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$ qui implique l'indépendance de X et Y .

Proposition 15. Soient X et Y deux v.a. avec densité jointe $f_{(X,Y)}(x, y)$. Alors X et Y sont indépendantes ssi il existe deux applications g, h telles que $f_{(X,Y)}(x, y) = g(x)h(y)$ pour tout couple (x, y) t.q. $f_{(X,Y)}(x, y) > 0$.

Démonstration. Si X et Y sont indépendantes alors on peut prendre $g = f_X$ et $h = f_Y$. Réciproquement: supposons que $f_{(X,Y)}(x, y) = g(x)h(y)$:

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy = g(x) \int_{\mathbb{R}} h(y) dy, \quad f_Y(y) = h(y) \int_{\mathbb{R}} g(x) dx$$

$$1 = \int_{\mathbb{R}} f_X(x) dx = \int_{\mathbb{R}} g(x) dx \int_{\mathbb{R}} h(y) dy$$

et donc $f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$.

Exemple 16. Soit (X, Y) un couple de v.a. dans \mathbb{R}^2 admettant pour densité $f_{(X,Y)}(x, y) = 8xy \mathbf{1}_{0 < x < y < 1}$. X et Y ne sont pas indépendantes car la fonction $\mathbf{1}_{0 < x < y < 1}$ ne peut pas s'écrire sous la forme d'un produit.

Espérance conditionnelle

Si X est un vecteur aléatoire dans \mathbb{R}^n admettant f_X comme densité et $g: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction positive alors on définit l'espérance $E[g(X)]$ de la v.a. $g(X)$ par la formule

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) f_X(x) dx \quad (1)$$

qui est toujours une quantité positive bien définie même si elle peut prendre la valeur $+\infty$. Si g est de signe quelconque et $E[|g(X)|] < +\infty$ alors on dit que $g(X)$ est intégrable et on peut définir l'espérance $E[g(X)]$ par la même formule (1). Si $g(X)$ n'est pas intégrable l'intégrale dans la formule (1) n'est pas bien définie.

Définition 17. Soit $Z = (X, Y)$ un vecteur aléatoire dans $\mathbb{R}^n \times \mathbb{R}^m$. Soit $g: \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction telle que $g(X, Y)$ est intégrable, c-à-d. $E[|g(X, Y)|] < +\infty$. On appellera espérance conditionnelle de $g(X, Y)$ sachant Y et on notera $E[g(X, Y)|Y]$ la v.a. (Y) où

$$(y) = \int_{\mathbb{R}^n} g(x, y) f_{X|Y=y}(x) dx, \quad y \in \mathbb{R}^m: f_Y(y) > 0.$$

Il est importante de remarquer que $E[g(X, Y)/Y]$ est une variable aléatoire.

Remarque 18. Par convenance on note $E[g(X)/Y = y] = \int_{\mathbb{R}^n} g(x, y) f_{X/Y=y}(x) dx$ l'espérance par rapport à la loi conditionnelle de X sachant $Y = y$. Cette espérance est une fonction réelle de y .

Exemple 19. Revenons à l'exemple 12 et calculons $E[XY/Y]$ (il faut donc prendre $g(x, y) = xy$). Vérifions d'abord la condition d'intégrabilité (qui donne sens au calcul de l'espérance conditionnelle):

$$E[|XY|] = \int_{\mathbb{R}^2} |xy| f_{(X,Y)}(x, y) dx dy = 2 \int_0^2 \int_0^y xye^{-\frac{1}{2}(x+y)} I_{0 < x < y} dx dy$$

$$= 2 \int_0^2 \int_0^y xye^{-\frac{1}{2}(x+y)} dx dy = 2 \int_0^2 x e^{-\frac{1}{2}x} dx = 2 < \infty$$

donc XY est bien intégrable.

$$(y) = \int_{\mathbb{R}} xy f_{X/Y=y}(x) dx = y \int_{\mathbb{R}} x \frac{e^{-\frac{1}{2}x} I_{0 < x < y}}{\int_0^y e^{-\frac{1}{2}x} dx} dx = \frac{y}{\int_0^y e^{-\frac{1}{2}x} dx} \int_0^y x e^{-\frac{1}{2}x} dx$$

$$(y) = \frac{y}{\int_0^y e^{-\frac{1}{2}x} dx} \int_0^y \frac{1}{2} e^{-\frac{1}{2}x} dx = \frac{y}{\int_0^y e^{-\frac{1}{2}x} dx} \left[-e^{-\frac{1}{2}x} \right]_0^y = \frac{y}{\int_0^y e^{-\frac{1}{2}x} dx} (1 - e^{-\frac{1}{2}y})$$

et donc

$$E[XY/Y] = \frac{Y}{\int_0^Y e^{-\frac{1}{2}x} dx} \frac{1 - e^{-\frac{1}{2}Y}}{\int_0^Y e^{-\frac{1}{2}x} dx}$$

Proposition 20. Soit h une application de $\mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ telle que $g(X, Y)h(Y)$ est intégrable. Alors

1. $E[g(X, Y)/Y]h(Y) = E[g(X, Y)h(Y)/Y]$
2. $E[E[g(X, Y)/Y]h(Y)] = E[g(X, Y)h(Y)]$

Démonstration. Soit $(Y) = E[g(X, Y)/Y]$ et $(Y) = E[g(X, Y)h(Y)/Y]$ où

$$(y) = \int_{\mathbb{R}^n} g(x, y) f_{X/Y=y}(x) dx, \quad (y) = \int_{\mathbb{R}^n} g(x, y) h(y) f_{X/Y=y}(x) dx$$

alors $(y) = h(y) (y)$ qui donne la première égalité. Pour la deuxième on remarque que

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

$$= \int_{\mathbb{R}^n \times \mathbb{R}^m} g(x, y) h(y) f_{X/Y=y}(x) f_Y(y) dx dy = \int_{\mathbb{R}^n \times \mathbb{R}^m} g(x, y) h(y) f_{(X,Y)}(x, y) dx dy$$

$$= E[g(X, Y)h(Y)]$$

par la définition de la densité conditionnelle et d'espérance.

Cas particuliers:

$$\text{Pour } g(x, y) = x \text{ et } h(y) = 1: E[E[X/Y]] = E[X]$$

$$\text{Pour } g(x, y) = 1, h(y) \text{ intégrable: } E[h(Y)/Y] = h(Y).$$

Exemple 21. Soient X et Y deux v.a. indépendantes de loi exponentielle de densité $f(x) = e^{-x} \mathbf{1}_{x>0}$ avec $\lambda = 1$. Calculons la densité conditionnelle $f_{X|X+Y}$ et la variance conditionnelle $E[X|X+Y]$. Si $z < x$

$$P(X+Y < z, X < x) = P(X+Y < z, X < z) = \int_0^z \int_0^{z-u} f(u) f(v) du dv$$

et si $z \geq x$

$$P(X+Y < z, X < x) = \int_0^x \int_0^{z-u} f(u) f(v) du dv.$$

Par conséquent, pour $x \leq z$ la densité jointe du couple $(X+Y, X)$ est

$$f_{(X+Y, X)}(z, x) = \frac{d}{dx} P(X+Y < z, X < x) = f(z-x) f(x) = e^{-z} e^x$$

et $f_{(X+Y, X)}(z, x) = 0$ si $z < x$. Par ailleurs, la densité de $X+Y$ est la convolution de deux densités exponentielles, i.e. $f_{X+Y}(z) = z e^{-z} \mathbf{1}_{z>0}$. Donc la densité conditionnelle de X sachant $X+Y$ est

$$f_{X|X+Y}(x|z) = \frac{f_{(X+Y, X)}(z, x)}{f_{X+Y}(z)} = \frac{1}{z} \mathbf{1}_{0 \leq x \leq z}$$

qui est la densité uniforme sur l'intervalle $[0, z]$. Donc on calcule facilement que $E[X|X+Y = z] = \frac{z}{2}$ et finalement que

$$E[X|X+Y] = \frac{X+Y}{2}.$$

Variance, covariance et corrélation

Définition 22. La covariance $\text{Cov}(X, Y)$ du couple (X, Y) de v.a. réelles est donnée par $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$. La variance de X est $\text{Var}(X) = \text{Cov}(X, X) = E[(X - E[X])^2] \geq 0$.

Si $\text{Var}(X) = 0$ alors $X = E[X]$ est une constante. La covariance est une fonction symétrique ($\text{Cov}(X, Y) = \text{Cov}(Y, X)$) et linéaire par rapport à chacun de ses arguments ($\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$). $\text{Var}(X+c) = \text{Var}(X)$. On a que

$$\begin{aligned} \text{Var}(X+Y) &= \text{Cov}(X+Y, X+Y) = \text{Cov}(X, X) + 2\text{Cov}(X, Y) + \text{Cov}(Y, Y) \\ &= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y). \end{aligned}$$

Si X, Y sont indépendantes $\text{Cov}(X, Y) = 0$, le réciproque n'est pas vrai en général.

Exemple 23. Soit $X \sim N(0, 1)$ et $Y = X^2$. Alors $\text{Cov}(X, Y) = E[XY] = E[X^3] = 0$ mais évidemment X, Y ne sont pas indépendantes: par exemple $0 = P(X > 1, Y < 1) \neq P(X > 1)P(Y < 1) = P(X > 1)P(-1 < X < 1) > 0$.

On a l'identité

$$\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$$

[Preuve: considérer le discriminant du polynôme positif $P(t) = \text{Var}(X + tY) \geq 0$.

Le coefficient de corrélation $\rho_{X,Y}$ est défini par

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \in [-1, 1]$$

Exemple 24. Si $\rho_{X,Y} = 1$ et $\text{Var}(Y) > 0$ alors existe $c \in \mathbb{R}$ tel que $\text{Var}(X - Y) = 0$ et donc $X - Y = \text{constante}$ qui donne que $\text{Cov}(X, Y) = \text{Cov}(Y, Y) = \text{Var}(Y)$, $\text{Var}(X) = \text{Var}(Y)$ et $\rho_{X,Y} = 1$. Donc on voit bien que le signe de $\rho_{X,Y}$ est celui de $\rho_{X,Y}$.

Pour le prouver, considérer le polynôme quadratique en t $P(t) = \text{Var}(X - tY) = \text{Var}(X) - 2t \text{Cov}(X, Y) + t^2 \text{Var}(Y)$. Or $P(t) \geq 0$ et donc l'équation $P(t) = 0$ admet au plus une solution et il admet une solution seulement si le discriminant est nul. Ici $\Delta = 4[\text{Cov}(X, Y)]^2 - 4\text{Var}(X)\text{Var}(Y) = 4\text{Var}(X)\text{Var}(Y)(\rho_{X,Y}^2 - 1) \leq 0$ et donc $\Delta = 0$ / $\rho_{X,Y} = 1$. Après il est clair que si $X = Y + c$ avec c constant on doit avoir

$$\rho_{X,Y} = \frac{\text{Cov}(Y + c, Y)}{\sqrt{\text{Var}(Y + c) \text{Var}(Y)}} = \frac{\text{Cov}(Y, Y)}{\sqrt{\text{Var}(Y) \text{Var}(Y)}} = \frac{\text{Var}(Y)}{\text{Var}(Y)} = 1 = \rho_{X,Y}.$$

Exercice 1. Montrer que $\rho_{aX+b, cY+d} = \rho_{X,Y}$ c-à-d que le coefficient de corrélation est invariante par des transformation affines des variables elles vérifient $\rho_{X,Y} = 1$.

Définition 25. On appelle variance conditionnelle de X sachant Y et on notera $\text{Var}(X/Y)$ la v.a.

$$\text{Var}(X/Y) = E[(X - E[X/Y])^2/Y]$$

Proposition 26. On a $\text{Var}(X/Y) = E[X^2/Y] - (E[X/Y])^2$

Démonstration.

$$\begin{aligned} \text{Var}(X/Y) &= E[X^2/Y] - (E[X/Y])^2 \\ &= E[X^2/Y] - E[X/Y]^2 \\ &= E[X^2/Y] - E[X/Y]^2 \end{aligned}$$

car $E[X E[X/Y]/Y] = E[X/Y] E[X/Y]$ et $E[(E[X/Y])^2/Y] = (E[X/Y])^2$.

Proposition 27. (Identité de la variance conditionnelle) Soient X et Y 2 v.a. sur le même espace de probabilité et $E[X^2] < +\infty$. Alors $\text{Var}(X) = E[\text{Var}(X/Y)] + \text{Var}(E[X/Y])$

Démonstration.

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 = E[E[X^2/Y]] - (E[X])^2 \\ &= E[E[X^2/Y]] - E[E[X/Y]]^2 + E[E[X/Y]]^2 - (E[X])^2 \\ &= E[\text{Var}(X/Y)] + \text{Var}(E[X/Y]) \end{aligned}$$