Rappels sur les intégrales multiples

Théorème 1. *[Fubini-Tonelli, cas $n=2$] Soit $f:\mathbb{R}^2\to\mathbb{R}$ une fonction positive, alors*

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

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

Exemple 2. $f(x,y)=xe^{-xy}\mathbb{I}_{x\geqslant 0}\mathbb{I}_{1<y<2}$. D'un part

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y)\mathrm{d}x\mathrm{d}y &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} xe^{-xy}\mathbb{I}_{x\geqslant 0}\mathrm{d}x\right)\mathbb{I}_{1<y<2}\mathrm{d}y\\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} y^2xe^{-xy}\mathbb{I}_{x\geqslant 0}\mathrm{d}x\right)\frac{\mathbb{I}_{1<y<2}}{y^2}\mathrm{d}y\\ &= \int_{\mathbb{R}}\left(\int_{\mathbb{R}} xe^{-x}\mathbb{I}_{x\geqslant 0}\mathrm{d}x\right)\frac{\mathbb{I}_{1<y<2}}{y^2}\mathrm{d}y\\ &= \int_1^2\frac{\mathrm{d}y}{y^2}=\frac{1}{2}.\end{aligned}$$

D'autre part

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

Vecteurs aléatoires à densité

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

$$F_X(x_1,\ldots,x_n)=\mathbb{P}(X_1\leqslant x_1,\ldots,X_n\leqslant x_n)$$

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

Exemple 3. Soit $n=2$ et $B=[x_1,y_1]\times[x_2,y_2]$ alors il est facile de vérifier que

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

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

Définition 4. *On dit que X admet une densité $f_X\colon \mathbb{R}^n \to \mathbb{R}_+$ ssi pour tout Borélien $B \in \mathcal{B}(\mathbb{R}^n)$ (la tribu Borélienne de $\mathbb{R}^n$) on peut exprimer la probabilité de l'évènement $\{X \in B\}$ par une intégrale sur B de f_X:*

$$\mathbb{P}(X \in B) = \int_B f_X(x_1,\ldots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\,.$$

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

$$\int_{\mathbb{R}^n} f_X(x_1,\ldots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n = \mathbb{P}(X \in \mathbb{R}^n) = 1$$

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

$$F_X(t_1,\ldots,t_n) \overset{\mathrm{def}}{=} \mathbb{P}(X_1 \leqslant t_1,\ldots,X_n \leqslant t_n) = \int_{-\infty}^{t_1}\cdots\int_{-\infty}^{t_n} f_X(x_1,\ldots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\,.$$

c'est la probabilité de l'évènement $\{X \in B\}$ pour $B =]-\infty, t_1] \times \cdots \times]-\infty, t_n] \subseteq \mathbb{R}^n$. On peut déterminer la densité en dérivant la fonction de répartition:

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

formule valable en tout point de continuité de $\partial^n F_X(t_1,\ldots,t_n)/\partial t_1\cdots\partial t_n$.

L'interprétation intuitive de la densité f_X est la suivante: si $\Delta x_i \ll 1$ alors la probabilité de l'évènement $\{X_i \in [x_i, x_i + \Delta x_i] \text{ pour } i = 1,\ldots,n\}$ est approchable par

$$\begin{aligned}\mathbb{P}(X_i \in [x_i, x_i + \Delta x_i] \text{ pour } i = 1,\ldots,n) &= \int_{x_1}^{x_1+\Delta x_1}\ldots\int_{x_n}^{x_n+\Delta x_n} f_X(t_1,\ldots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n \\ &\simeq f_X(x_1,\ldots,x_n)\Delta x_1\cdots\Delta x_n\,.\end{aligned}$$

La densité est donc proportionnelle à la mesure de probabilité d'un petit voisinage du point $(x_1,\ldots,x_n)$. Autrement dit, si $B_n(x,\delta) \subseteq \mathbb{R}^n$ est la boule n-dimensionnelle de rayon δ centré en $x \in \mathbb{R}^n$ et $V_n(\delta) = |B_n(x,\delta)|$ le volume de $B(x,\delta)$, i.e.

$$V_n(\delta) = \int_{B_n(x,\delta)} \mathrm{d}t_1\cdots\mathrm{d}t_n$$

alors si $\delta \ll 1$ on a l'approximation $\mathbb{P}(X \in B(x,\delta)) \simeq f_X(x)V_n(\delta)$.

Exemple 5. Soit $Z = (X, Y)\colon \Omega \to \mathbb{R}^2$ 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

$$f_Z(x,y) = \frac{\partial^2}{\partial x \partial y} F_Z(x,y) = \begin{cases} 1 & \text{si } x \in]0,1[\text{ et } y \in]0,1[; \\ 0 & \text{sinon.}\end{cases}$$

Donc f est la densité uniforme sur le carré $[0,1]^2$.

Densités marginales

Définition 6. *Si Z est un vecteur aléatoire dans $\mathbb{R}^n$ admettant une densité f_Z alors tout sous-vecteurs Y de Z de dimension $k \leqslant n$ admettent une densité qu'on obtient en intégrant f_Z par rapport aux composantes qui ne figurent pas dans Y. On appelle cette densité la densité marginale de Y. Explicitement si $Y = (Z_1, \ldots, Z_k)$ alors*

$$\mathbb{P}(Y \in B) = \mathbb{P}((Z_1, \ldots, Z_k) \in B) = \mathbb{P}(Z \in B \times \mathbb{R}^{n-k}) = \int_{(z_1, \ldots, z_k) \in B} f_Z(z_1, \ldots, z_n) \mathrm{d}z_1 \cdots \mathrm{d}z_n$$

$$= \int_B \left(\int_{\mathbb{R}^{n-k}} f_Z(z_1, \ldots, z_k, z_{k+1}, \ldots, z_n) \mathrm{d}z_{k+1} \cdots \mathrm{d}z_n \right) \mathrm{d}z_1 \cdots \mathrm{d}z_k$$

donc $f_Y(y_1, \ldots, y_k) = \int_{\mathbb{R}^{n-k}} f_Z(y_1, \ldots y_k, z_{k+1}, \ldots, z_n) \mathrm{d}z_{k+1} \cdots \mathrm{d}z_n$.

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

Exemple 7. Considérons le couple (X, Y) de densité

$$f_{(X,Y)}(x, y) = \alpha \frac{e^{-y}}{2\sqrt{x}} \mathbb{I}_{0 < x < y^2} \mathbb{I}_{y > 0}$$

- Déterminer $\alpha > 0$ t.q. $f_{(X,Y)}$ soit correctement normalisée.
- Déterminer les densités marginales f_X et f_Y.
- Calculer $\mathbb{P}(X > 1)$.

Calculons

$$\begin{aligned} I &= \int_{\mathbb{R}^2} f_{(X,Y)}(x, y) \mathrm{d}x \mathrm{d}y = \alpha \int_0^\infty \left(\int_0^{y^2} \frac{\mathrm{d}x}{2\sqrt{x}} \right) e^{-y} \mathrm{d}y \\ &= \alpha \int_0^\infty y \, e^{-y} \mathrm{d}y = \alpha \end{aligned}$$

donc $\alpha = 1$.

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) \mathrm{d}y = \frac{\mathbb{I}_{x>0}}{2\sqrt{x}} \int_{\sqrt{x}}^\infty e^{-y} \mathrm{d}y = \frac{e^{-\sqrt{x}}}{2\sqrt{x}} \mathbb{I}_{x>0}$$

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

$$\mathbb{P}(X > 1) = \int_1^\infty f_X(x) \mathrm{d}x = \int_1^\infty \frac{e^{-\sqrt{x}}}{2\sqrt{x}} \mathrm{d}x = \frac{1}{e}$$

Densité et espérance conditionnelle

Définition 8. *Soit $Z = (X, Y)$ un vecteur aléatoire dans $\mathbb{R}^m \times \mathbb{R}^n$ admettant une densité f_Z. 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*

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

Cette définition est motivée par le fait que, si $\delta \ll 1$:

$$\mathbb{P}(X \in B_m(x, \delta) | Y \in B_n(y, \delta)) = \frac{\mathbb{P}(X \in B_m(x, \delta), Y \in B_n(y, \delta))}{\mathbb{P}(Y \in B_n(y, \delta))} \simeq \frac{f_{(X,Y)}(x, y) V_m(\delta) V_n(\delta)}{f_Y(y) V_n(\delta)}$$

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

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

Exemple 9. Considérons $Z = (X, Y)$ de densité $f_Z(x, y) = 2\lambda^2 e^{-\lambda(x+y)} \mathbb{I}_{0<x<y}$. Quelle est la densité conditionnelle de X sachant $Y = y$?

Calculons d'abord $f_Y(y)$:

$$f_Y(y) = \int_{\mathbb{R}} 2\lambda^2 e^{-\lambda(x+y)} \mathbb{I}_{0<x<y} \mathrm{d}x = 2\lambda e^{-\lambda y} \mathbb{I}_{y>0} \int_0^y \lambda e^{-\lambda x} \mathrm{d}x = 2\lambda e^{-\lambda y}(1 - e^{-\lambda y}) \mathbb{I}_{y>0}$$

Il vient que

$$f_{X|Y=y}(x) = \frac{2\lambda^2 e^{-\lambda(x+y)} \mathbb{I}_{0<x<y}}{2\lambda e^{-\lambda y}(1 - e^{-\lambda y}) \mathbb{I}_{y>0}} = \frac{\lambda e^{-\lambda x} \mathbb{I}_{0<x<y}}{1 - e^{-\lambda y}} \qquad \text{pour tout } y > 0.$$

Définition 10. *Une famille $(X_i)_{i=1,\ldots,n}$ de v.a. est indépendante ssi pour tout B_i, $i = 1, \ldots, n$, on a que les évènements $\{X_i \in B_i\}_{i=1,\ldots,n}$ sont indépendants, i.e.:*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

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 11. *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 là $f_{X|Y=y}(x) = f_X(x)$.*

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 dépends pas de y

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

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

Proposition 12. *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 $f_{(X,Y)}(x, y) = g(x) h(y)$:

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\mathrm{d}y = g(x) \int_{\mathbb{R}} h(y)\mathrm{d}y, \qquad f_Y(y) = h(y) \int_{\mathbb{R}} g(x)\mathrm{d}x$$

$$1 = \int f_X(x)\mathrm{d}x = \int_{\mathbb{R}} g(x)\mathrm{d}x \int_{\mathbb{R}} h(y)\mathrm{d}y$$

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

Exemple 13. Soit (X, Y) un couple de v.a. dans $\mathbb{R}^2$ admettant pour densité $f_{(X,Y)}(x, y) = 8\, x\, y\, \mathbb{I}_{0<x<y<1}$. X et Y ne sont pas indépendantes car la fonction $\mathbb{I}_{0<x<y<1}$ 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 $g: \mathbb{R}^n \to \mathbb{R}$ est une fonction positive alors on défini l'espérance $\mathbb{E}[g(X)]$ de la v.a. $g(X)$ par la formule

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

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

Définition 14. *Soit $Z = (X, Y)$ un vecteur aléatoire de $\mathbb{R}^n \times \mathbb{R}^m$. Soit $g: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ une fonction telle que $g(X, Y)$ est intégrable, c-à-d $\mathbb{E}|g(X, Y)| < +\infty$. On appellera espérance conditionnelle de $g(X, Y)$ sachant Y et on notera $\mathbb{E}[g(X, Y)|Y]$ la v.a. $\Psi(Y)$ où*

$$\Psi(y) = \int_{\mathbb{R}^n} g(x, y) f_{X|Y=y}(x)\mathrm{d}x, \qquad \forall y \in \mathbb{R}^m: f_Y(y) > 0.$$

Il est importante de remarquer que l'espérance conditionnelle $\mathbb{E}[g(X,Y)|Y]$ est une variable aléatoire.

Exemple 15. Revenons à l'exemple 9 et calculons $\mathbb{E}[X\,Y|Y]$ (il faut donc prendre $g(x,\,y) = x\,y$). Vérifions d'abord la condition d'integrabilité (qui donne sens au calcul de l'espérance conditionnelle):

$$\mathbb{E}[|X Y|] = \int_{\mathbb{R}^2} |x\,y|\,f_{(X,Y)}(x,\,y)\mathrm{d}x\mathrm{d}y = 2\lambda^2 \int\int x y e^{-\lambda(x+y)} \mathbb{I}_{0<x<y} \mathrm{d}x\mathrm{d}y$$

$$\leqslant 2\lambda^2 \int_0^\infty \int_0^\infty x y e^{-\lambda(x+y)} \mathrm{d}x\mathrm{d}y = 2\lambda^2 \left(\int_0^\infty x\,e^{-\lambda x}\mathrm{d}x \right)^2 < \infty$$

donc $X Y$ est bien intégrable.

$$\Psi(y) = \int_{\mathbb{R}} x\,y\,f_{X|Y=y}(x)\mathrm{d}x = y\,\mathbb{I}_{y>0} \int_{\mathbb{R}} x \frac{\lambda e^{-\lambda x}\mathbb{I}_{0<x<y}}{1-e^{-\lambda y}}\,\mathrm{d}x = \frac{\lambda y}{1-e^{-\lambda y}}\,\mathbb{I}_{y>0} \int_0^y x\,e^{-\lambda x}\mathrm{d}x$$

$$\Psi(y) = \frac{y}{1-e^{-\lambda y}}\,\mathbb{I}_{y>0}\,\frac{1-e^{-\lambda y} - \lambda y\,e^{-\lambda y}}{\lambda}$$

et donc

$$\mathbb{E}[XY|Y] = \frac{Y}{1-e^{-\lambda Y}}\,\frac{1-e^{-\lambda Y}-\lambda Y e^{-\lambda Y}}{\lambda}$$

Proposition 16. *Soit h une application de $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ telle que $g(X,Y)h(Y)$ est intégrable. Alors*

1. $\mathbb{E}[g(X,Y)|Y]h(Y) = \mathbb{E}[g(X,Y)h(Y)|Y]$

2. $\mathbb{E}[\mathbb{E}[g(X,Y)|Y]h(Y)] = \mathbb{E}[g(X,Y)h(Y)]$

Démonstration. Soit $\Phi(Y) = \mathbb{E}[g(X,Y)|Y]$ et $\Psi(Y) = \mathbb{E}[g(X,Y)h(Y)|Y]$ où

$$\Phi(y) = \int_{\mathbb{R}^n} g(x,y) f_{X|Y=y}(x)\mathrm{d}x, \qquad \Psi(y) = \int_{\mathbb{R}^n} g(x,y)h(y) f_{X|Y=y}(x)\mathrm{d}x$$

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

$$\mathbb{E}[\mathbb{E}[g(X,Y)|Y]h(Y)] = \mathbb{E}[\Phi(Y)h(Y)] = \mathbb{E}[\Psi(Y)] = \int_{\mathbb{R}^m} \Psi(y) f_Y(y)\mathrm{d}y$$

$$= \int_{\mathbb{R}^n\times\mathbb{R}^m} g(x,y)h(y) f_{X|Y=y}(x) f_Y(y)\mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}^n\times\mathbb{R}^m} g(x,y)h(y) f_{(X,Y)}(x,y)\mathrm{d}x\mathrm{d}y$$

$$= \mathbb{E}[g(X,Y)h(Y)]$$

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

Cas particuliers:

- $g(x,y) = x$ et $h(y) = 1$: $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$

- $g(x,y) = 1$, $h(Y)$ intégrable: $\mathbb{E}[h(Y)|Y] = h(Y)$.

Rappels sur la variance/covariance. La covariance $\mathrm{Cov}(X, Y)$ du couple (X, Y) de v.a. réelles est donnée par $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$. La variance de X est $\mathrm{Var}(X) = \mathrm{Cov}(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \geqslant 0$. Si $\mathrm{Var}(X) = 0$ alors $X = \mathbb{E}[X]$ est une constante. La covariance est une fonction symétrique ($\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$) et lineaire par rapport à chacun de ses arguments ($\mathrm{Cov}(\alpha X + \beta Y, Z) = \alpha\mathrm{Cov}(X, Z) + \beta\mathrm{Cov}(Y, Z)$). $\mathrm{Var}(\alpha X + c) = \alpha^2\,\mathrm{Var}(X)$. On a l'inégalité

$$\mathrm{Cov}(X, Y)^2 \leqslant \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

[preuve: considérer le discriminant du polynôme positive $p(t) = \mathrm{Var}(X + tY) \geqslant 0$]. Le coefficient de corrélation $\rho_{X,Y}$ est défini par

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

Si $\rho_{X,Y} = \pm 1$ et $\mathrm{Var}(Y) > 0$ alors existe $\alpha \in \mathbb{R}$ tel que $\mathrm{Var}(X - \alpha Y) = 0$ et donc $X - \alpha Y =$ constante qui donne que $\mathrm{Cov}(X, Y) = \mathrm{Cov}(\alpha Y, Y) = \alpha\mathrm{Var}(Y)$, $\mathrm{Var}(X) = \alpha^2\,\mathrm{Var}(Y)$ et $\rho_{X,Y} = \mathrm{sign}\,\alpha$. Donc on voit bien que le signe de α est celui de $\rho_{X,Y}$.

Définition 17. *On appelle variance conditionnelle de X sachant Y et on notera* $\mathrm{Var}(X|Y)$ *la v.a.*

$$\mathrm{Var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y]$$

Proposition 18. *On a* $\mathrm{Var}(X|Y) = \mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2$

Démonstration.

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

car $\mathbb{E}[X\,\mathbb{E}[X|Y]|Y] = \mathbb{E}[X|Y]\,\mathbb{E}[X|Y]$ et $\mathbb{E}[(\mathbb{E}[X|Y])^2|Y] = (\mathbb{E}[X|Y])^2$. □

Proposition 19. *(identité de la variance conditionnelle) Soient X et Y 2 v.a. sur le même espace de probabilité et $\mathbb{E}[X^2] < +\infty$. Alors* $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathbb{E}(X|Y)]$

Démonstration.

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[\mathbb{E}[X^2|Y] - (\mathbb{E}(X|Y))^2] + \mathbb{E}[(\mathbb{E}(X|Y))^2] - (\mathbb{E}[\mathbb{E}[X|Y]])^2$$

$$= \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathbb{E}(X|Y)]$$ □