Rappels sur les intégrales multiples

L'outil principal pour calculer des intégrales en plusieurs variables est le théorème de Fubini-Tonelli.

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

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

D'autre part

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

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

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

alors $I=\int_0^1\int_0^1 f(x,y)\mathrm{d}x\mathrm{d}y$ n'est pas bien défini car

$$\int_0^1\int_0^1|f(x,y)|\mathrm{d}x\mathrm{d}y=\int_0^1\left(\int_0^y\frac{y^2-x^2}{(x^2+y^2)^2}\mathrm{d}x\right)\mathrm{d}y+\int_0^1\left(\int_0^x\frac{x^2-y^2}{(x^2+y^2)^2}\mathrm{d}y\right)\mathrm{d}x$$

$$=\int_0^1\left(\frac{1}{y}-\frac{1}{1+y^2}\right)\mathrm{d}y=+\infty$$

Or, les intégrales

$$I_1 = \int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \mathrm{d}x \right) \mathrm{d}y, \qquad I_2 = \int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \mathrm{d}y \right) \mathrm{d}x$$

sont bien défini et il satisfont $I_1 = -I_2$. En effet:

$$\int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \mathrm{d}y = \int_0^1 \frac{1}{(x^2 + y^2)} \mathrm{d}y - 2 \int_0^1 \frac{y^2}{(x^2 + y^2)^2} \mathrm{d}y = \frac{1}{1 + x^2}$$

et alors

$$I_2 = \int_0^1 \frac{1}{1 + x^2} \mathrm{d}x = [\arctan(x)]_0^1 = \frac{\pi}{4} = -I_1$$

ce qui est in contradiction avec une application naïve de Fubini (car dans ce cas $I_1 = I_2 = I = 0$).

Théorème 4. *(Changement des variables) Soit h: $\mathbb{R}^n \to \mathbb{R}^n$. On note $h_i(x)$ ses composantes dans la base canonique: $h(x) = (h_1(x), \ldots, h_n(x))$ et Jh son Jacobien:*

$$Jh(x) = \det \left(\frac{\partial h_i(x)}{\partial x_j} \right)_{i,j=1,\ldots,n}.$$

Supposons que

i. les dérivées partielles de $h_i(x)$ sont continues sur $\mathbb{R}^n$ pour $i = 1, \ldots, n$.

ii. h est une bijection.

iii. $Jh(x) \neq 0$ pour tout $x \in \mathbb{R}^n$.

Alors pour toute fonction $f: \mathbb{R}^n \to \mathbb{R}$ et tout ensemble Borelién $K \subseteq \mathbb{R}^n$ tels que $\int_K |f(x)| \mathrm{d}x < \infty$

$$\int_K f(x) \mathrm{d}x = \int_{h^{-1}(K)} f(h(y)) |Jh(y)| \mathrm{d}y.$$

Remarque 5. Par le Théorème de la fonction inverse, sous le condition du Théorème 4 la fonction inverse $g(y) = h^{-1}(y)$ existe partout dans $\mathbb{R}^n$ et

$$Jg(y) = \frac{1}{Jh(g(y))}.$$

On voit donc que h vérifie les conditions i,ii et iii du Theoreme 4 si et seulement si g aussi vérifie ces conditions.

Vecteurs aléatoires

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'événement $\{X \in B\}$ (car $\mathbb{P}(A)$ est définie seulement pour $A \in \mathcal{A}$). La loi de X est la l'application μ_X: $\mathcal{B}(\mathbb{R}^n) \to [0, 1]$ 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 $X(\omega) = (X_1(\omega), \ldots, X_n(\omega))$ (donc des v.a. réelles). La fonction de répartition caractérise la loi de X, i.e. n'importe quel événement $\{X \in B\}$ peut être calculé à l'aide de F_X.

Exemple 6. 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 $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$):

$$\begin{aligned}\mathbb{P}(X \in B) &= \mathbb{P}(x_1 < X_1 \leqslant y_1, x_2 < X_2 \leqslant y_2)\\ &= \mathbb{P}(X_1 \leqslant y_1, x_2 \leqslant X_2 \leqslant y_2) - \mathbb{P}(X_1 \leqslant x_1, x_2 < X_2 \leqslant y_2)\\ &= \mathbb{P}(X_1 \leqslant y_1, X_2 \leqslant y_2) - (X_1 \leqslant y_1, X_2 \leqslant x_2)\\ &\quad -(\mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant y_2) - \mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2))\\ &= F_X(y_1, y_2) - F_X(y_1, x_2) - F_X(x_1, y_2) + F_X(x_1, x_2)\,.\end{aligned}$$

Définition 7. *Le vecteur X à valeurs dans $\mathbb{R}^n$ est discret si il peut prendre que une quantité d au plus dénombrable des valeurs distinct. Autrement dit si il existent des ensembles au plus dénombrables $\mathcal{X}_1, \ldots, \mathcal{X}_n$ tels que $\mathbb{P}(X_1 \in \mathcal{X}_1, \ldots, X_n \in \mathcal{X}_n) = 1$. Dans ce cas on appelle la quantité*

$$p_X(x) = p_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \qquad x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n$$

la densité discrete de X.

La densité discrete d'un vecteur X satisfait

a) $p_X(x) \geqslant 0$;

b) $\sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} \cdots \sum_{x_n \in \mathcal{X}_n} p_X(x_1, \ldots, x_n) = 1$.

Définition 8. *On dit que X est un vecteur continu ou que il admet une densité (continue) f_X: $\mathbb{R}^n \to \mathbb{R}_+$ 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'événement $\{X \in B\}$ 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 (continue ou discrete), 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 $F_X: \mathbb{R}^n \to [0, 1]$ de X est donné par

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

c'est la probabilité de l'événement $\{X \in B\}$ 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'événement $\{X_i \in [x_i, x_i + \Delta x_i] \text{ pour } i = 1, \ldots, n\}$ 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 $B_n(x, \delta) = \{y \in \mathbb{R}^n : |x - y| \leqslant \delta\} \subseteq \mathbb{R}^n$ est la boule n-dimensionnelle de rayon δ centrée 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 9. 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

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

et on peut vérifier que

$$F_Z(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \mathbb{I}_{]0,1[}(z_1) \, \mathbb{I}_{]0,1[}(z_2) \mathrm{d}z_1 \mathrm{d}z_2 .$$

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

Définition 10. *Soit $D \subseteq \mathbb{R}^n$ un ensemble mesurable et tel que $\mathrm{Vol}(D) = \int_D \mathrm{d}x > 0$ (son volume est positif). On dit que $X \colon \Omega \to \mathbb{R}^n$ a une loi uniforme sur D si X admet densité*

$$f_X(x) = \frac{\mathbb{I}_D(x)}{\mathrm{Vol}(D)}.$$

Théorème 11. *Soit X un vecteur aléatoire continu à valeurs dans $\mathbb{R}^n$ de densité f_X et $h \colon \mathbb{R}^n \to \mathbb{R}^n$ une transformation vérifiant les hypotheses du Theoreme 4, alors la v.a. $Y = h(X)$ admet la densité f_Y donnée par*

$$f_Y(y) = f_X(h^{-1}(y)) \frac{1}{|Jh(h^{-1}(y))|}$$

pour tout $y \in \mathbb{R}^n$.

Démonstration. Considérons la fonction de repartition de Y:

$$F_Y(t_1, \ldots, t_n) = \mathbb{P}(Y_1 \leqslant t_1, \ldots, Y_n \leqslant t_n) = \int_{A(t)} f_X(x) \mathrm{d}x$$

où $A(t) = \{x \in \mathbb{R}^n : h_1(x) \leqslant t_1, \ldots, h_n(x) \leqslant t_n\}$. Par le Théorème 4 on a que $(y = h(x))$

$$\int_{A(t)} f_X(x) \mathrm{d}x = \int_{h(A(t))} f_X(h^{-1}(y)) \frac{\mathrm{d}y}{|Jh(h^{-1}(y))|}$$

or $h(A(t)) = \{y = h(x) : x \in A(t)\} = \{y = h(x) : h(x_1) \leqslant t_1, \ldots, h(x_n) \leqslant t_n\} = \{y_1 \leqslant t_1, \ldots, y_n \leqslant t_n\}$. Donc

$$F_Y(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f_X(h^{-1}(y)) \frac{\mathrm{d}y}{|Jh(h^{-1}(y))|}.$$

Exercice 1. Montrer que si X est un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ de densité f_X et si A est une matrice $n \times n$ inversible et $b \in \mathbb{R}^n$ alors la v.a. $Y = A X + b$ a densité f_Y donnée par

$$f_Y(y) = f_X(A^{-1}(y - b)) \frac{1}{|\mathrm{Det}(A)|}.$$

□

Densités marginales

Définition 12. *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 13. 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 $\alpha = 1$.

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

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

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

Exemple 14. 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

$$f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \qquad g_{X,Y}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}[1 + x\, y\, \mathbb{I}_{[-1,1]}(x)\mathbb{I}_{[-1,1]}(y)]$$

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

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\mathrm{d}x = 1$$

on obtient que

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

et

$$g_X(x) = \int_{\mathbb{R}} g_{X,Y}(x,y)\mathrm{d}y = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-(x^2+y^2)/2}[1 + x\, y\, \mathbb{I}_{[-1,1]}(x)\mathbb{I}_{[-1,1]}(y)]\mathrm{d}y$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}[1 + x\, y\, \mathbb{I}_{[-1,1]}(x)\mathbb{I}_{[-1,1]}(y)]\mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

car

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} y\, \mathbb{I}_{[-1,1]}(y)\mathrm{d}y = 0$$

par symétrie.

Densité et espérance conditionnelle

Définition 15. *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 centrée en x sachant que Y est dans une petite boule centrée en y.

Exemple 16. Considérons $Z=(X,Y)$ de densité $f_Z(x,y)=2\lambda^2e^{-\lambda(x+y)}\mathbb{I}_{0<x<y}$. Quelle est la densité conditionnelle de X sachant $Y=y$?

Calculons d'abord $f_Y(y)$:

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

Il vient que

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

Définition 17. *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 événements $\{X_i\in B_i\}_{i=1,\ldots,n}$ 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 18. *Soient X et Y deux 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. $\square$

Proposition 19. *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 $f_{(X,Y)}(x, y) = g(x)\, h(y)$.*

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 20. 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. En effet $f_{(X,Y)}(1/2,\, 2/3) > 0$ et $f_{(X,Y)}(1/4,\, 1/3) > 0$. Si $f_{(X,Y)}(x,\, y) = q(x)\, h(y)$ alors on doit avoir que $q(1/2) > 0$ et $h(1/3) > 0$ mais alors $f_{(X,Y)}(1/2, 1/3) > 0$ ce qu'est en contradiction avec la definition de $f_{(X,Y)}$ car $f_{(X,Y)}(1/2, 1/3) = 0$.

Espérance et espérance conditionnelle

Si X est un vecteur aléatoire dans $\mathbb{R}^n$ admettant f_X comme densité et $g\colon \mathbb{R}^n \to \mathbb{R}$ 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.

Théorème 21. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et supposons que pour tout fonction mesurable bornée $g\colon \mathbb{R}^n \to \mathbb{R}$*

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

pour une certaine fonction $q\colon \mathbb{R}^n \to \mathbb{R}_+$. Alors X admet densité $f_X = q$.

Remarque 22. Ce resultat est assez pratique pour determiner la densité par changement des variables.

Définition 23. *Soit $Z = (X, Y)$ un vecteur aléatoire de $\mathbb{R}^n \times \mathbb{R}^m$. Soit $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 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.

Remarque 24. Par convenance on note $\mathbb{E}[g(X)|Y=y]=\int_{\mathbb{R}^n} g(x,\,y) f_{X|Y=y}(x)\mathrm{d}x$ l'espérance par rapport à la loi conditionnelle de X sachant $Y=y$. Cette espérance est une fonction réelle de y.

Exemple 25. Revenons à l'exemple 16 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 26. *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: $\mathbb{E}[h(Y)|Y]=h(Y)$.

Exemple 27. Soient X et Y deux v.a. indépendantes de loi exponentielle de densité $f(x) = \lambda e^{-\lambda x}\mathbb{I}_{x>0}$ avec $\lambda > 0$. Calculons la densité conditionnelle $f_{X|X+Y}$ et l'espérance conditionnelle $\mathbb{E}[X|X+Y]$. Si $z<x$

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

et si $z \geqslant x$

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

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

$$f_{(X+Y,X)}(z,x) = \frac{\partial^2}{\partial u \partial v}\mathbb{P}(X+Y<z, X<x) = f(z-x)f(x) = \lambda^2 e^{-\lambda z}$$

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. $f_{X+Y}(z) = \lambda^2 z e^{-\lambda z}\mathbb{I}_{z>0}$. Donc la densité conditionnelle de X sachant $X+Y$ est

$$f_{X|X+Y}(x|z) = \frac{f_{(X+Y,X)}(z,x)}{f_{X+Y}(z)} = \frac{1}{z}\mathbb{I}_{0\leqslant x\leqslant z}$$

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

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

Une application de cet exemple est la suivante. Soit X l'instant où la première demande arrive à un système de service et Y l'interval de temps entre l'arrivée de la premiere et la deuxième demandes. Si on sait que la deuxième arrive au temps z (donc $X+Y=z$) alors on vient de determiner que la loi du temps d'arrivée de la premiere est donnée par la densité $1/z$ sur l'intervalle $[0, z]$. En particulier, la moyenne du temps d'arrivée de la premiere est $z/2$.

Variance, covariance et corrélation

Définition 28. *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 linéaire 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 que

$$\mathrm{Var}(X+Y) = \mathrm{Cov}(X+Y, X+Y) = \mathrm{Cov}(X, X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Y)$$

$$= \mathrm{Var}(X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Var}(Y).$$

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

Exemple 29. Soit $X \sim \mathcal{N}(0,1)$ et $Y = X^2$. Alors $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] = \mathbb{E}[X^3] = 0$ mais évidemment X, Y ne sont pas indépendantes: par exemple $0 = \mathbb{P}(X>1, Y<1) \neq \mathbb{P}(X>1)\mathbb{P}(Y<1) = \mathbb{P}(X>1)\mathbb{P}(-1<X<1) > 0$.

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

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

Exemple 30. 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 $\rho_{X,Y}=\operatorname{sign}\alpha$. Donc on voit bien que le signe de α est celui de $\rho_{X,Y}$.

Pour le prouver, considérer le polynôme quadratique en α défini par $P(\alpha)=\mathrm{Var}(X-\alpha Y)=\mathrm{Var}(X)-2\alpha\,\mathrm{Cov}(X,Y)+\alpha^2\mathrm{Var}(Y)$. Or $P(\alpha)\geqslant 0$ et donc l'équation $P(\alpha)=0$ admet au plus une solution et il admet une solution seulement si le discriminant est nul. Ici $\Delta=4\,[\mathrm{Cov}(X,Y)]^2-4\mathrm{Var}(X)\mathrm{Var}(Y)=4\,\mathrm{Var}(X)\,\mathrm{Var}(Y)\,(\rho_{X,Y}^2-1)\leqslant 0$ et donc $\Delta=0\Leftrightarrow|\rho_{X,Y}|=1$. Apres il est clair que si $X=\alpha Y+c$ avec c constant on doit avoir

$$\rho_{X,Y}=\frac{\mathrm{Cov}(\alpha Y+c,Y)}{\sqrt{\mathrm{Var}(\alpha Y+c)\,\mathrm{Var}(Y)}}=\frac{\alpha\mathrm{Cov}(Y,Y)}{\sqrt{\alpha^2\mathrm{Var}(Y)\,\mathrm{Var}(Y)}}=\frac{\alpha}{|\alpha|}=\mathrm{sgn}\,(\alpha).$$

Exercice 2. Montrer que $\rho_{X,Y}=\mathrm{sgn}(a\,d)\,\rho_{aX+b,dY+d}$, c-à-d que le coefficient de corrélation est invariante par des transformation affines des variables elles vérifient $\mathrm{sgn}(ad)=1$.

Définition 31. *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 32. *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 33. *(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.

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

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

Meilleure prévision et regression

Soit X, Y un couple aléatoire de densité jointe $f_{X,Y}$ et tel que $\mathbb{E}[Y^2] < \infty$. Le problème de la meilleure prevision est de trouver une fonction g telle que l'écart moyen quadratique de Y à $g(X)$ soit le plus petit possible:

$$\mathbb{E}[(Y - g(X))^2] = \inf_h \mathbb{E}[(Y - h(X))^2]$$

Dans le contexte de la meilleure prevision la variable X est appelée *variable explicative* ou *predicteur* et Y est appelée *variable expliquée*. L'esperance conditionnelle $\Psi(x) = \mathbb{E}[Y|X = x]$ donné l'unique solution de ce problème.

Théorème 34. *(*MEILLEURE PREVISION*) Soit* $\Psi(x) = \mathbb{E}[Y|X = x]$ *alors*

$$\mathbb{E}[(Y - \Psi(X))^2] \leqslant \mathbb{E}[(Y - h(X))^2]$$

pour tout $h: \mathbb{R} \to \mathbb{R}$ *mesurable et telle que* $\mathbb{E}[h(X)^2] < \infty$.

Démonstration. On a que

$$\mathbb{E}[(Y - h(X))^2] = \mathbb{E}[(Y - \Psi(X) + \Psi(X) - h(X))^2]$$

$$= \mathbb{E}[(Y - \Psi(X))^2] + \mathbb{E}[(\Psi(X) - h(X))^2] + 2\,\mathbb{E}[(Y - \Psi(X))(\Psi(X) - h(X))].$$

Or

$$\mathbb{E}[(Y - \Psi(X))(\Psi(X) - h(X))] = \mathbb{E}[\mathbb{E}[(Y - \Psi(X))|X](\Psi(X) - h(X))] = 0$$

et donc

$$\mathbb{E}[(Y - h(X))^2] = \mathbb{E}[(Y - \Psi(X))^2] + \mathbb{E}[(\Psi(X) - h(X))^2]$$

ce qu'implique que

$$\mathbb{E}[(Y - h(X))^2] \geqslant \mathbb{E}[(Y - \Psi(X))^2]$$

avec égalité si et seulement si $\mathbb{E}[(\Psi(X) - h(X))^2] = 0$ c'est à dire si $\Psi(X) = h(X)$.

□

La fonction $g(x) = \mathbb{E}[Y|X = x]$ est dite **fonction de regression de Y sur X**. Dans le cas où X, Y sont des v.a. réelles la regression est dite simple. Si X, Y sont des v.a. à valeurs vectorielles alors la regression est dite multiple ou multivariée.

Exemple 35. Soit (X, Y) un vecteur aléatoire dans $\mathbb{R}^2$ de densité

$$f_{X,Y}(x, y) = (x + y)\mathbb{I}_{0<x,y<1}.$$

Explicitions la fonction de regression de X sur Y: $g(x) = \mathbb{E}[Y|X = x]$. La densité marginale de X est donnée par

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}y = \left(x + \frac{1}{2}\right)\mathbb{I}_{0<x<1}.$$

Alors la densité conditionnelle est donnée par

$$f_{Y|X}(y|x) = \frac{x+y}{x+1/2}\mathbb{I}_{0<x,y<1}$$

et

$$g(x) = \int_{\mathbb{R}} y\, f_{Y|X}(y|x)\mathrm{d}y = \int_{\mathbb{R}} y\,\frac{x+y}{x+1/2}\mathbb{I}_{0<x,y<1}\mathrm{d}y = \frac{\frac{x}{2}+\frac{1}{3}}{x+\frac{1}{2}}$$

pour tout $x \in]0,1[$. Soulignons que, en general, g est une fonction non-lineaire de x.

Si g est la fonction de regression de X sur Y alors la v.a. $\xi = Y - g(X) = Y - \mathbb{E}[Y|X]$ represente l'erreur stochastique dans la prevision de Y par $g(X)$. On appelle ξ le residu de regression. Par definition d'esperance conditionnelle

$$\mathbb{E}[\xi|X] = 0$$

et donc aussi $\mathbb{E}[\xi] = 0$. L'erreur quadratique de l'approximation de Y par $g(X)$ est

$$\Delta = \mathbb{E}[(Y - g(X))^2] = \mathbb{E}(\xi^2) = \mathrm{Var}(\xi).$$

On appelle Δ la variance résiduelle. On a que $\Delta \leqslant \mathrm{Var}(Y)$: par le théorème de meilleur prevision, pour tout $h\colon \mathbb{R} \to \mathbb{R}$ mesurable $\Delta \leqslant \mathbb{E}[(Y - h(X))^2]$. En choisissant $h(x) = \mathbb{E}[Y]$ (constante) on a que

$$\Delta \leqslant \mathbb{E}[(Y - h(X))^2] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathrm{Var}(Y).$$

On a aussi que

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

car

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