MAP 311 - Aléatoire

Leçon 3

2016-2017

Rappel : Espérance

Proposition

Soient $X : \Omega \to \mathbb{R}$ *une v.a. et* $g : \mathbb{R} \to \mathbb{R}$ *(mesurable) avec* $g \geq 0$ *ou* $g(X)$ *intégrable. Alors*

$$\mathbb{E}(g(X)) = \int_{\Omega} g(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}} g(x)P_X(dx)$$

Pour une v.a. discrète X à valeurs dans $E = \{x_i, i \in I\} \subset \mathbb{R}$ de loi $\{p_i, i \in I\}$:

$$\mathbb{E}(g(X)) = \sum_{i \in I} p_i g(x_i)$$

Pour une v.a. X à densité f :

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx$$

Calcul de lois

La proposition suivante est très utile. Elle va nous permettre de trouver la loi d'une variable aléatoire.

Proposition

Soit X une variable aléatoire réelle. S'il existe une probabilité μ sur $\mathbb{R}$ telle que $\forall h : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\mathbb{E}\big(h(X)\big) = \int_{\mathbb{R}} h(x)\mu(dx),$$

alors la loi de X est égale à μ.

En particulier, si nous avons pour tout $h : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\mathbb{E}\big(h(X)\big) = \int_{\mathbb{R}} h(x)f(x)dx,$$

avec f densité de probabilité, alors la loi de X a pour densité f (méthode de la fonction muette).

Calcul de lois

Idée de preuve : Les fonctions continues bornées peuvent approcher les fonctions de la forme $h_y = \mathbf{1}_{]-\infty,y]}$. Or pour une telle fonction,

$$\mathbb{E}\big(h_y(X)\big) = \mathbb{P}(X \leq y) = P_X(]-\infty, y])$$

fonction de répartition qui caractérise P_X. Nous en déduisons donc la fonction de répartition de X, qui caractérise sa loi.

La plupart du temps, il faut utiliser cette proposition pour trouver la loi d'une variable aléatoire.
Son usage est souvent plus facile que l'usage de la fonction de répartition.

Exemple : Si Y est une variable aléatoire normale de loi $\mathcal{N}(m, \sigma^2)$, alors

$$X = \frac{Y - m}{\sigma}$$

est de loi $\mathcal{N}(0, 1)$.
Par changement de variable $x = \frac{y-m}{\sigma}$, on a :

$$\begin{aligned}
\mathbb{E}(h(X)) &= \mathbb{E}\Big(h(\frac{Y-m}{\sigma})\Big) \\
&= \int_{\mathbb{R}} h(\frac{y-m}{\sigma}) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(y-m)^2}{2\sigma^2}\Big) dy \\
&= \int_{\mathbb{R}} h(x) \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big) dx
\end{aligned}$$

Morale : L'ensemble des variables aléatoires de loi normale est stable par transformation affine.

Question fondamentale : Soit X une variable aléatoire réelle de densité f_X. $Y = g(X)$ admet-elle une densité et comment la calculer ?
Remarque : Cette densité n'existe pas toujours !
Si $g(x) \equiv a$, alors la variable aléatoire $Y = g(X)$ est la constante a. Dans ce cas

$$\mathbb{E}\big(h(Y)\big) = h(a) = \int_{\mathbb{R}} h(y)\delta_a(dy),$$

où δ_a est la mesure de Dirac en a. Cette probabilité est définie pour tout borélien A par

$$\delta_a(A) = 1 \text{ si } a \in A \text{ et } \delta_a(A) = 0 \text{ si } a \notin A.$$

Nous en déduisons que $P_Y = \delta_a$.
Ce n'est pas une loi à densité ! Elle ne charge qu'un point : le point a. Or une loi à densité ne charge pas les points.

Dans le cas général, pour étudier la loi de Y, on écrit

$$\mathbb{E}\big(h(Y)\big) = \mathbb{E}\big(h \circ g(X)\big) = \int_{\mathbb{R}} h \circ g(x) f_X(x) dx \stackrel{?}{=} \int_{\mathbb{R}} h(y) f_Y(y) dy$$

Si on montre la dernière égalité, on peut affirmer que Y est à densité f_Y.
Méthode par changement de variable : on pose $y = g(x)$.
Attention: souvent, on est amené à subdiviser l'espace pour avoir g bijective et dérivable sur des intervalles.

Exemple : Considérons une variable aléatoire X de loi à densité f_X. Alors, la variable aléatoire $X^2 = Y$ a une loi à densité et sa densité vaut

$$f_Y(y) = \frac{1}{2\sqrt{y}}\Big(f_X(-\sqrt{y}) + f_X(\sqrt{y})\Big)\mathbf{1}_{]0,+\infty[}(y)$$

Preuve :

$$\begin{aligned}
\mathbb{E}\big(h(Y)\big) &= \int_{-\infty}^{0} h(x^2) f_X(x) dx + \int_{0}^{\infty} h(x^2) f_X(x) dx \\
&= \int_{0}^{\infty} h(y) f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} dy + \int_{0}^{\infty} h(y) f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} dy
\end{aligned}$$

Appliquons cela au calcul de la loi de la variable aléatoire X^2, où X est une variable aléatoire gaussienne centrée réduite. $Y = X^2$ admet la densité

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} \exp\Big(-\frac{y}{2}\Big) \mathbf{1}_{]0,+\infty[}(y)$$

Cette loi s'appelle la loi du chi-deux à un degré de liberté, notée $\chi^2(1)$.

Exercice : Loi de Cauchy
La variable aléatoire X suit la loi de Cauchy de paramètre $a > 0$ si sa loi possède la densité

$$f(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}$$

1) Vérifier que c'est bien une densité.
2) Que dire de l'espérance de X ?
3) Quelle est la loi de $Y = 1/X$?

Nous avons vu des exemples de lois discrètes et des exemples de lois à densité.
Il existe des variables aléatoires qui ne sont ni discrètes, ni à densité !

Exemple : Le cycle d'un feu de circulation est le suivant: le feu est vert sur l'intervalle $[0, v]$ et orange ou rouge sur l'intervalle $[v, v+r]$ avec $v, r > 0$. L'instant d'arrivée U d'un automobiliste est supposé uniformément réparti sur l'intervalle $[0, r+v]$. Soit T le temps d'attente d'un automobiliste au feu (dans le cas où aucun véhicule n'est arrêté devant lui). Ce temps d'attente T vérifie:

$$T = 0 \text{ si } U \leq v; \quad T = v + r - U \text{ si } U \in]v, v+r]$$

Pour h continue bornée

$$\begin{aligned}
\mathbb{E}\big(h(T)\big) &= \mathbb{E}\big(h(0)\mathbf{1}_{U\in[0,v]}\big) + \mathbb{E}\big(h(v+r-U)\mathbf{1}_{U\in]v,v+r]}\big) \\
&= h(0)\frac{v}{r+v} + \frac{1}{r+v}\int_v^{r+v} h(v+r-u)du \\
&= h(0)\frac{v}{r+v} + \frac{1}{r+v}\int_0^r h(s)ds
\end{aligned}$$

Ainsi $\mathbb{P}(T=0) = \frac{v}{r+v}$ et pour $t \neq 0$, $\mathbb{P}(T=t) = 0$.
La loi de T est la mesure de probabilité

$$\frac{v}{r+v}\delta_0 + \frac{1}{r+v}\mathbf{1}_{[0,r]}(t)dt$$

Inégalité de Bienaymé-Chebyshev

Proposition

(i) Inégalité de Markov : Si $X \in L^1$ et $a > 0$, alors

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

(ii) Inégalité de Bienaymé-Chebyshev (Contrôle de l'écart de X à sa moyenne) : Si $X \in L^2$, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}$$

Exemple : $\mathbb{P}\big(|X - \mathbb{E}(X)| \geq 10\sigma_X\big) \leq 1\%$ (Universel mais peu précis).

Preuve :

$$\mathbb{E}(|X|) \geq \mathbb{E}\big(|X|\mathbf{1}_{|X|\geq a}\big) \geq \mathbb{E}\big(a\mathbf{1}_{|X|\geq a}\big) = a\mathbb{P}(|X| \geq a)$$

Application : une inégalité de concentration

Résultat : Soit X une v.a. à valeurs dans $\{x_1, \dots, x_n\}$.
On pose $\mathrm{Osc}(X) = \max_i x_i - \min_i x_i$.
Alors, pour tout $r > 0$, on a

$$\mathbb{P}(X - \mathbb{E}(X) \geq r) \leq \exp\left(-\frac{2r^2}{\mathrm{Osc}(X)^2}\right)$$

Exemple : $X \sim \mathcal{B}(n, p)$. Alors $\mathbb{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$ et $\mathrm{Osc}(X) = n$.
Bienaymé-Chebyshev :

$$\mathbb{P}\big(|X - np| \geq 10\sqrt{np(1-p)}\big) \leq \frac{np(1-p)}{100np(1-p)} = 0{,}01.$$

Concentration :

$$\mathbb{P}\big(X - np \geq 10\sqrt{np(1-p)}\big) \leq \exp\left(-\frac{200p(1-p)}{n}\right) = e^{-10} = 4{,}5\,10^{-5},$$

si $n = 5$ et $p = 1/2$.

Inégalité de Jensen

Rappel : $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$, d'où $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$.

Proposition

Soit $X \in L^1$ et g mesurable telle que $g(X) \in L^1$. Supposons de plus g convexe. Alors

$$\mathbb{E}\big(g(X)\big) \geq g\big(\mathbb{E}(X)\big).$$

En particulier : Si $X \in L^p$, $p \geq 1$, alors $\mathbb{E}(|X|^p) \geq \mathbb{E}(|X|)^p$.
Preuve : g est convexe, donc pour tout $a \in \mathbb{R}$, il existe $\lambda_a \in \mathbb{R}$ tel que $g(x) \geq g(a) + \lambda_a(x - a)$ pour tout $x \in \mathbb{R}$.
On l'applique à $x = X(\omega)$ et $a = \mathbb{E}(X)$. Il suffit alors de prendre l'espérance pour obtenir le résultat.

Applications :

- La moyenne arithmétique est plus grande que la moyenne géométrique.
  Si $a_1, \ldots, a_n \geq 0$, alors

$$\frac{a_1 + \cdots + a_n}{n} \geq \sqrt[n]{a_1 \cdots a_n}$$

  Preuve : Si un des a_i est nul, c'est trivial.
  Sinon, on pose $b_i = \ln(a_i)$, on considère $\Omega = \{b_1, \ldots, b_n\}$ muni de la probabilité uniforme et on applique Jensen avec $X(\omega) = \omega$ et $g = \exp$.
- Géométrie.
  Le rectangle de surface 1 qui a le périmètre minimal est le carré.
  Le parallélépipède rectangle de volume 1 qui a la surface minimale est le cube.
  Le triangle de surface 1 qui a le plus petit périmètre est le triangle équilatéral.

Vecteurs aléatoires

Description de phénomènes aléatoires qui évoluent dans $\mathbb{R}^n$.

Exemple : $\boldsymbol{Z} = (Z_1, Z_2)$: couple des deux coordonnées de l'impact d'une flèche sur une cible.

Tribu borélienne $\mathcal{B}(\mathbb{R}^n)$ de $\mathbb{R}^n$: engendrée par les $\prod_{i=1}^n]-\infty, z_i]$, avec $(z_1, \ldots, z_n) \in \mathbb{Q}^n$.

Définition

Un vecteur aléatoire $\boldsymbol{Z} : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ *est formé de n variables aléatoires :* $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$.

Loi $P_{\boldsymbol{Z}}$ de $\boldsymbol{Z}$: loi jointe du n-uplet $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ - probabilité sur $\mathbb{R}^n$.

Loi P_{Z_i} de Z_i : $i$ème loi marginale - probabilité sur $\mathbb{R}$.

Loi d'un vecteur aléatoire à densité

La loi du vecteur est caractérisée par sa fonction de répartition :

$$\forall (z_1, \dots, z_n) \in \mathbb{R}^n, \quad F_{\boldsymbol{Z}}(z_1, \dots, z_n) = \mathbb{P}\Big(\boldsymbol{Z} \in \prod_{i=1}^{n}]-\infty, z_i]\Big)$$

C'est une fonction difficile à manipuler.

Définition

La loi de $\boldsymbol{Z}$ admet la densité f positive, intégrable sur $\mathbb{R}^n$ et d'intégrale 1 *si et seulement si*

$$F_{\boldsymbol{Z}}(z_1, \dots, z_n) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} f_{\boldsymbol{Z}}(y_1, \dots, y_n) dy_1 \cdots dy_n$$

Espérance pour un vecteur aléatoire à densité

Dans ce cas, pour $g : \mathbb{R}^n \to \mathbb{R}$ tel que

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(z_1, \dots, z_n)| f_{\boldsymbol{Z}}(z_1, \dots, z_n) dz_1 \cdots dz_n < +\infty$$

on définit

$$\mathbb{E}\big(g(\boldsymbol{Z})\big) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(z_1, \dots, z_n) f_{\boldsymbol{Z}}(z_1, \dots, z_n) dz_1 \cdots dz_n$$

Exemple : On lance une fléchette sur une cible circulaire de rayon 1. On note (X, Y) le point d'impact, vecteur aléatoire de loi uniforme sur la cible (disque de rayon 1). La densité de cette loi est

$$f(x, y) = \frac{1}{\pi} \mathbf{1}_{x^2+y^2 \leq 1}$$

Soit $g(x, y) = \sqrt{x^2 + y^2}$ (distance du point d'impact au centre). Alors

$$\mathbb{E}\big(g(X, Y)\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{x^2 + y^2} \frac{1}{\pi} \mathbf{1}_{x^2+y^2 \leq 1} dxdy.$$

On pose $x = r \cos \theta$ et $y = r \sin \theta$,

$$\mathbb{E}\big(g(X, Y)\big) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 r\, r\, drd\theta = \frac{2}{3}$$

Formule des marginales : pour $h \in \mathcal{C}_b(\mathbb{R})$,

$$\mathbb{E}\big(h(Z_1)\big) = \int_{-\infty}^{+\infty} h(z_1)\Big(\iint_{\mathbb{R}^{n-1}} f_{\boldsymbol{Z}}(z_1, \ldots, z_n) dz_2 \cdots dz_n\Big) dz_1$$

On en déduit que Z_1 a la loi de densité

$$f_{Z_1}(z_1) = \iint_{\mathbb{R}^{n-1}} f_{\boldsymbol{Z}}(z_1, \ldots, z_n) dz_2 \cdots dz_n$$

Exemple : fléchette.

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x,y) dy = \int_{\mathbb{R}} \frac{1}{\pi} \mathbf{1}_{x^2+y^2 \leq 1} dy = \frac{1}{\pi} \mathbf{1}_{[-1,1]}(x) \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy \\ &= \frac{2}{\pi} \sqrt{1-x^2} \mathbf{1}_{[-1,1]}(x) \end{aligned}$$

Remarque : X n'est pas une loi uniforme.
Par symétrie, Y a même loi que X.
Attention, la réciproque est fausse : Si X et Y sont deux variables réelles à densité, alors (X, Y) n'est pas forcément à densité.

Couple de variables aléatoires

$L^2 = \{$variables aléatoires de carré intégrable$\}$.
Soient X et Y deux variables aléatoires réelles de L^2, alors $XY \in L^1$.
Preuve : $|XY| \leq \frac{1}{2}(X^2 + Y^2)$.
Inégalité de Cauchy-Schwarz : Si X et Y sont deux variables aléatoires réelles de L^2, alors

$$\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

Preuve : $\forall \lambda \in \mathbb{R}$,

$$\mathbb{E}\big((\lambda|X| + |Y|)^2\big) = \lambda^2\mathbb{E}(X^2) + 2\lambda\mathbb{E}(|XY|) + \mathbb{E}(Y^2) \geq 0.$$

Polynôme du second degré en λ toujours positif $\Longrightarrow$ discriminant ≤ 0.

De plus discriminant $= 0 \Longleftrightarrow Y$ et X sont proportionnelles.
Donc il y a égalité dans Cauchy-Schwarz ssi il existe $\lambda \in \mathbb{R}$ tel que $X = \lambda Y$ ou $Y = \lambda X$ (avec probabilité 1).

Covariance de variables aléatoires

Définition

Si X et $Y \in L^2$, on appelle covariance de X et Y le nombre

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big) = \mathbb{E}\big(XY\big) - \mathbb{E}(X)\mathbb{E}(Y).$$

Cov est une forme bilinéaire symétrique et $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
Exemple de la fléchette : la densité de X

$$f_X(x) = \frac{2}{\pi}\sqrt{1 - x^2}\mathbf{1}_{[-1,1]}(x)$$

est paire donc $\mathbb{E}(X) = 0$. De même, $\mathbb{E}(Y) = 0$. Par symétrie,

$$\mathbb{E}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{\pi}xy\mathbf{1}_{x^2+y^2\leq 1}dxdy = 0$$

On en déduit que $\mathrm{Cov}(X, Y) = 0$.

Coefficient de corrélation

Définition

Soient X et Y deux variables aléatoires dans L^2, avec $\sigma_X > 0$ et $\sigma_Y > 0$.
Le coefficient de corrélation est :

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Remarque : Par l'inégalité de Cauchy-Schwarz, on a $|\rho(X, Y)| \leq 1$.
$\rho > 0$ (resp. < 0) veut dire X et Y varient dans le même sens (resp. en sens contraire).

Exemples :

- $\rho(X, X) = 1$,
- $\rho(X, -X) = -1$,
- si $X \sim \mathcal{U}(0, 1)$, alors $\rho(X, X^2) = \frac{\mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(X^2)}} = \frac{\sqrt{15}}{4} \simeq 0{,}968$,
- si $X \sim \mathcal{U}(-1, 1)$, alors $\rho(X, X^2) = 0$.

Coefficient de corrélation - Droite de régression

Soit (X, Y) un couple aléatoire.

Rappel sur l'espérance : meilleure approximation de Y par une v.a. constante :

$$\mathbb{E}(Y) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \ \mathbb{E}\big((Y-a)^2\big)$$

et alors $\min_{a\in\mathbb{R}} \mathbb{E}\big((Y-a)^2\big) = \mathrm{Var}(Y)$.

Approximation (régression) linéaire : trouver la meilleure approximation de Y par une v.a. de la forme $aX + b$.

On minimise $\mathbb{E}\big([Y-(aX+b)]^2\big)$ en (a, b). On obtient :

$$\hat{a} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}, \qquad \hat{b} = \mathbb{E}(Y) - \hat{a}\mathbb{E}(X)$$

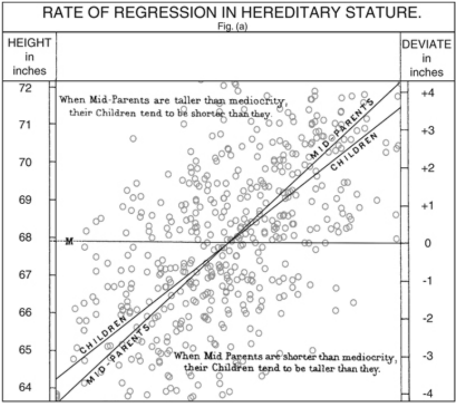
et alors

$$\min_{a,b\in\mathbb{R}} \mathbb{E}\big((Y-aX-b)^2\big) = \mathrm{Var}(Y)\big(1-\rho^2(X,Y)\big)$$

Plus $|\rho|$ est proche de 1, meilleure est l'approximation.

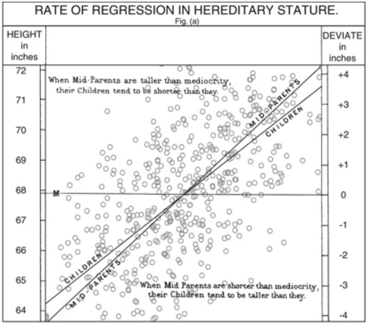
Régression vers la médiocrité, selon Galton (1886)

Terme "régression ": introduit par Galton à la suite d'une étude sur la taille des descendants de personnes de grande taille, qui diminue de générations en générations vers une taille moyenne (leur taille régresse).



En abscisse : taille du parent; en ordonnée : taille de l'enfant.

Régression vers la médiocrité, selon Galton (1886)



X : taille du parent, Y : taille de l'enfant

Supposons $\mathbb{E}(X) = \mathbb{E}(Y)$. Alors les deux droites $y = \hat{a}x + \hat{b}$ et $y = x$ passent par le même point $(\mathbb{E}(X), \mathbb{E}(Y))$.

Supposons $\sigma_X = \sigma_Y$. Alors $\hat{a} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \leq 1$.

$\hookrightarrow$ La droite de régression est sous la diagonale. La taille du descendant est plus proche de la moyenne que celle de l'ancêtre.

Galton en déduit que cela se passe ainsi pour des caractéristiques mentales et que donc, l'hérédité sans contrôle conduit à la médiocrité.

Moments du vecteur aléatoire $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$

- Espérance de $\boldsymbol{Z}$: $\boldsymbol{Z} \in L^1$ ssi $\forall i = 1, \ldots, n$, $Z_i \in L^1$ et

$$\mathbb{E}(\boldsymbol{Z}) = (\mathbb{E}(Z_1), \ldots, \mathbb{E}(Z_n)) \in \mathbb{R}^n$$

- Matrice des covariances de $\boldsymbol{Z}$: $\boldsymbol{Z} \in L^2$ ssi $\forall i = 1, \ldots, n$, $Z_i \in L^2$ et

$$\mathbf{C}_{\boldsymbol{Z}} = (\mathrm{Cov}(Z_i, Z_j))_{i,j=1}^n$$

$\mathbf{C}_{\boldsymbol{Z}}$ est une matrice symétrique et positive : pour tout vecteur $\boldsymbol{a} = (a_i)_{i=1}^n$,

$$\sum_{i,j=1}^n a_i a_j \mathrm{Cov}(Z_i, Z_j) \geq 0$$

En effet,

$$\sum_{i,j=1}^n a_i a_j \mathrm{Cov}(Z_i, Z_j) = \mathrm{Cov}\Big(\sum_{i=1}^n a_i Z_i, \sum_{j=1}^n a_j Z_j\Big) = \mathrm{Var}\Big(\sum_{i=1}^n a_i Z_i\Big) \geq 0$$

Proposition

Soit $\boldsymbol{Z}$ *un vecteur aléatoire à valeurs dans* $\mathbb{R}^n$*, de matrice de covariance* $\mathbf{C}_{\boldsymbol{Z}}$.
Soient $\mathbf{A}$ *une matrice de taille* $m \times n$ *et* $\boldsymbol{V} = \mathbf{A}\boldsymbol{Z}$.
Alors le vecteur aléatoire $\boldsymbol{V}$ *a pour espérance et matrice de covariance :*

$$\mathbb{E}(\boldsymbol{V}) = \mathbf{A}\mathbb{E}(\boldsymbol{Z}), \quad \mathbf{C}_{\boldsymbol{V}} = \mathbf{A}\mathbf{C}_{\boldsymbol{Z}}\mathbf{A}^t$$

où $\mathbf{A}^t$ *est la matrice transposée de* $\mathbf{A}$.

Preuve : pour $1 \leq i, k \leq m$:

$$\mathbb{E}(V_i) = \sum_{j=1}^{n} A_{ij}\mathbb{E}(Z_j)$$

et

$$\begin{aligned} \mathrm{Cov}(V_i, V_k) &= \mathrm{Cov}\Big(\sum_{j=1}^{n} A_{ij}Z_j, \sum_{l=1}^{n} a_{kl}Z_l\Big) = \sum_{j,l=1}^{n} A_{ij}A_{kl}\mathrm{Cov}(Z_j, Z_l) \\ &= (\mathbf{A}\mathbf{C}_{\boldsymbol{Z}}\mathbf{A}^t)_{ik} \end{aligned}$$

Lois conditionnelles, cas à densité

Soit (X, Y) un couple de v.a. de densité $f(x, y)$.
On note f_X et f_Y les densités marginales.

Proposition

Soit $x \in \mathbb{R}$ tel que $f_X(x) > 0$. La fonction $f_{Y|X=x} : \mathbb{R} \to \mathbb{R}^+$ définie par

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}$$

définit une densité, appelée densité conditionnelle de Y sachant $X = x$.

Attention : $\mathbb{P}(X = x) = 0$!
Justification heuristique :

$$\begin{aligned} f_{Y|X=x}(y)\Delta y &= \frac{f(x, y)\Delta x \Delta y}{f_X(x)\Delta x} \\ &\simeq \frac{\mathbb{P}(x \le X \le x + \Delta x,\, y \le Y \le y + \Delta y)}{\mathbb{P}(x \le X \le x + \Delta x)} \\ &= \mathbb{P}(y \le Y \le y + \Delta y \,|\, x \le X \le x + \Delta x) \end{aligned}$$

Proposition

Soit $h : \mathbb{R}^2 \to \mathbb{R}$ *mesurable positive ou bornée, alors*

$$\begin{aligned}
\mathbb{E}\big(h(X,Y)\big) &= \iint_{\mathbb{R}^2} h(x,y) f(x,y)\,dxdy \\
&= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} h(x,y) f_{Y|X=x}(y)\,dy \Big) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} h(x,y) f_{X|Y=y}(x)\,dx \Big) f_Y(y)\,dy
\end{aligned}$$

En effet, on a :

$$f(x,y) = f_{Y|X=x}(y) f_X(x) = f_{X|Y=y}(x) f_Y(y)$$

Espérance conditionnelle

Espérance conditionnelle de Y sachant $X = x$:

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)dy = \Psi(x)$$

Espérance conditionnelle de Y sachant X : v.a.

$$\mathbb{E}(Y|X) = \Psi(X), \text{ avec } \Psi(x) = \mathbb{E}(Y|X = x)$$

Proposition

Soit $h : \mathbb{R}^2 \to \mathbb{R}$ mesurable positive ou bornée, alors

$$\mathbb{E}\big(h(X, Y)\big) = \mathbb{E}\big(\mathbb{E}(h(X, Y)|X)\big) = \mathbb{E}\big(\mathbb{E}(h(X, Y)|Y)\big)$$

Exemple : Soient X et Y de densité jointe

$$f_{X,Y}(x,y) = \frac{1}{x}\mathbf{1}_{0<y<x<1}(x,y)$$

1) Trouver les densités marginales de X et de Y.
2) Calculer $\mathbb{E}(Y)$.
3) Calculer, pour $x \in]0,1[$, la densité conditionnelle de Y sachant $X = x$.
4) En déduire $\mathbb{E}(Y|X = x)$ et $\mathbb{E}(Y|X)$.
1)

$$f_X(x) = \int_0^1 \frac{1}{x}\mathbf{1}_{0<y<x<1}(x,y)dy = \mathbf{1}_{]0,1[}(x) \qquad \to X \sim \mathcal{U}(0,1)$$

$$f_Y(y) = \int_0^1 \frac{1}{x}\mathbf{1}_{0<y<x<1}(x,y)dx = -\mathbf{1}_{]0,1[}(y)\ln y$$

2)

$$\mathbb{E}(Y) = -\int_0^1 y\ln y dy = -\left[\frac{y^2}{2}\ln y\right]_0^1 + \int_0^1 \frac{y}{2}dy = \frac{1}{4}$$

3) Pour $x \in [0,1]$:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{x}\mathbf{1}_{]0,x[}(y) \qquad \to Y|X = x \sim \mathcal{U}(0,x)$$

4)

$$\mathbb{E}(Y|X=x) = \int y f_{Y|X=x}(y)dy = \frac{x}{2}$$

$$\mathbb{E}(Y|X) = \frac{X}{2}$$

On vérifie que $\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(Y|X)\big)$.

Vecteurs aléatoires indépendants

Définition

Soient $\boldsymbol{Z}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et $\boldsymbol{V}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^m$.
$\boldsymbol{Z}$ et $\boldsymbol{V}$ sont indépendants si pour tous boréliens A et B de $\mathbb{R}^n$ et $\mathbb{R}^m$, on a

$$\mathbb{P}(\boldsymbol{Z} \in A, \boldsymbol{V} \in B) = \mathbb{P}(\boldsymbol{Z} \in A)\mathbb{P}(\boldsymbol{V} \in B).$$

Il suffit de vérifier

$$\mathbb{P}\Big(\boldsymbol{Z} \in \prod_{i=1}^{n}]-\infty, z_i], \boldsymbol{V} \in \prod_{j=1}^{m}]-\infty, v_j]\Big)$$
$$= \mathbb{P}\Big(\boldsymbol{Z} \in \prod_{i=1}^{n}]-\infty, z_i]\Big)\mathbb{P}\Big(\boldsymbol{V} \in \prod_{j=1}^{m}]-\infty, v_j]\Big)$$

car la fonction de répartition caractérise la loi.

Vecteurs aléatoires indépendants

Proposition

Si $\boldsymbol{Z}$ et $\boldsymbol{V}$ ont des densités, alors $\boldsymbol{Z}$ et $\boldsymbol{V}$ sont indépendants si et seulement si pour (presque) tous $(\boldsymbol{z}, \boldsymbol{v}) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$f_{\boldsymbol{Z},\boldsymbol{V}}(\boldsymbol{z}, \boldsymbol{v}) = f_{\boldsymbol{Z}}(\boldsymbol{z}) f_{\boldsymbol{V}}(\boldsymbol{v})$$

Preuve : Si la densité jointe se met sous forme produit, alors

$$\begin{aligned}
\mathbb{P}(\boldsymbol{Z} \in A, \boldsymbol{V} \in B) &= \iint_{A \times B} f_{\boldsymbol{Z},\boldsymbol{V}}(\boldsymbol{z}, \boldsymbol{v}) d\boldsymbol{z} d\boldsymbol{v} \\
&= \int_A f_{\boldsymbol{Z}}(\boldsymbol{z}) d\boldsymbol{z} \int_B f_{\boldsymbol{V}}(\boldsymbol{v}) d\boldsymbol{v} = \mathbb{P}(\boldsymbol{Z} \in A) \mathbb{P}(\boldsymbol{V} \in B)
\end{aligned}$$

Donc $\boldsymbol{Z}$ et $\boldsymbol{V}$ sont indépendantes.

Vecteurs aléatoires indépendants

Proposition

Si $\boldsymbol{Z}$ et $\boldsymbol{V}$ ont des densités, alors $\boldsymbol{Z}$ et $\boldsymbol{V}$ sont indépendants si et seulement si pour (presque) tous $(\boldsymbol{z}, \boldsymbol{v}) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$f_{\boldsymbol{Z},\boldsymbol{V}}(\boldsymbol{z}, \boldsymbol{v}) = f_{\boldsymbol{Z}}(\boldsymbol{z}) f_{\boldsymbol{V}}(\boldsymbol{v})$$

Preuve de la réciproque (pour $n = m = 1$) : Si Z et V sont à densité et indépendants, alors en posant $f(z, v) = f_Z(z) f_V(v)$, le couple (X, Y) de densité f a pour fonction de répartition

$$\begin{aligned}
\mathbb{P}(X \leq x, Y \leq y) &= \iint_{]-\infty,x]\times]-\infty,y]} f(z, v) dz dv \\
&= \int_{-\infty}^{x} f_Z(z) dz \int_{-\infty}^{y} f_V(v) dv \\
&= \mathbb{P}(Z \leq x)\mathbb{P}(V \leq y) = \mathbb{P}(Z \leq x, V \leq y)
\end{aligned}$$

Donc (X, Y) et (Z, V) ont même loi.

Remarque : Soient Z et V deux variables aléatoires à densité indépendantes.
Soit v tel que $f_V(v) > 0$. Alors :
- la loi conditionnelle de Z sachant $V = v$ est la loi de Z :

$$f_{Z|V=v}(z) = \frac{f_{Z,V}(z,v)}{f_V(v)} = \frac{f_Z(z)f_V(v)}{f_V(v)} = f_Z(z)$$

- l'espérance conditionnelle de $h(Z)$ sachant $V = v$ est l'espérance de $h(Z)$:

$$\mathbb{E}\big(h(Z)|V = v\big) = \mathbb{E}\big(h(Z)\big)$$

pour toute fonction h positive ou bornée.

Proposition

Soient $\boldsymbol{Z} \in \mathbb{R}^n$ et $\boldsymbol{V} \in \mathbb{R}^m$ des vecteurs aléatoires indépendants. Soient g et h mesurables positives ou bornées sur $\mathbb{R}^n$ et $\mathbb{R}^m$, alors

$$\mathbb{E}\big(g(\boldsymbol{Z})h(\boldsymbol{V})\big) = \mathbb{E}\big(g(\boldsymbol{Z})\big)\mathbb{E}\big(h(\boldsymbol{V})\big)$$

Preuve :
1) Cas général: vrai pour g et h fonctions indicatrices d'événements, donc vrai pour les fonctions étagées. Vrai pour les fonctions positives par passage à la limite puis on écrit $g = g^+ - g^-$ et $h = h^+ - h^-$.
2) Si $\boldsymbol{Z}$ et $\boldsymbol{V}$ ont des lois à densité, preuve plus simple :

$$\begin{aligned}
\mathbb{E}\big(g(\boldsymbol{Z})h(\boldsymbol{V})\big) &= \iint_{\mathbb{R}^n \times \mathbb{R}^m} g(\boldsymbol{z})h(\boldsymbol{v})f(\boldsymbol{z},\boldsymbol{v})d\boldsymbol{z}d\boldsymbol{v} \\
&= \int_{\mathbb{R}^n} g(\boldsymbol{z})f_{\boldsymbol{Z}}(\boldsymbol{z})d\boldsymbol{z} \int_{\mathbb{R}^m} h(\boldsymbol{v})f_{\boldsymbol{V}}(\boldsymbol{v})d\boldsymbol{v} \\
&= \mathbb{E}\big(g(\boldsymbol{Z})\big)\mathbb{E}\big(h(\boldsymbol{V})\big) \text{ (par théorème de Fubini)}
\end{aligned}$$

Proposition

Si X et Y sont deux variables aléatoires réelles indépendantes, alors

$$\mathrm{Cov}(X, Y) = 0$$

Preuve immédiate:

$$\mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big) = \mathbb{E}\big(X - \mathbb{E}(X)\big)\mathbb{E}\big(Y - \mathbb{E}(Y)\big) = 0$$

Attention : Réciproque fausse. Voir l'exemple des fléchettes.
$\mathrm{Cov}(X, Y) = 0$ mais X et Y ne sont pas indépendantes :

$$\mathbb{P}(X^2 \geq 3/4, Y^2 \geq 3/4) = 0 < \mathbb{P}(X^2 \geq 3/4)\mathbb{P}(Y^2 \geq 3/4)$$