Leçon 9

Espérance, variance, théorème de transport

La théorie de l'intégration d'une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ par rapport à la probabilité $\mathbb{P}$ est la même que la théorie de l'intégration d'une fonction mesurable sur un espace mesuré quelconque (de masse totale finie). À titre d'observation utile, le fait de travailler avec une mesure de probabilité autorise l'application de l'inégalité de Jensen.

Dans toute cette leçon, les variables aléatoires sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et qualifiées simplement de *variables aléatoires réelles*.

1 Espérance, variance

Définition 1 (Espérance)**.** *Si X est une variable aléatoire réelle intégrable, son espérance, ou moyenne, est*

$$\mathbb{E}(X) = \int_\Omega X \, d\,\mathbb{P}.$$

L'espérance d'une variable aléatoire (intégrable) X est donc tout simplement son intégrale (par rapport à $\mathbb{P}$). Elle existe dès que

$$\mathbb{E}\big(|X|\big) = \int_\Omega |X| d\,\mathbb{P} < \infty,$$

et a un sens, éventuellement infini, si $X \geq 0$ (presque sûrement).

Une variable aléatoire X sera dite *centrée* si elle est intégrable et $\mathbb{E}(X) = 0$.

Toutes les propriétés et inégalités développées pour l'intégrale (et rappelées dans les Leçons 3-5) s'appliquent donc de la même façon pour les espérances de variables aléatoires. En particulier, l'espérance est linéaire.

La définition introduit néanmoins dans le même temps la notation $\mathbb{E}(X)$, la lettre $\mathbb{E}$ résonnant avec $\mathbb{P}$. La force et la souplesse de cette notation seront amplement manifestées dans la suite.

Il est utile de rappeler également que puisque $\mathbb{P}$ est une probabilité, si X est une variable aléatoire constante, $\mathbb{P}(X = c) = 1$ où $c \in \mathbb{R}$, alors $\mathbb{E}(X) = c$. Pour la même raison, toute variable aléatoire X presque sûrement bornée, i.e. telle que $\mathbb{P}(|X| \leq c) = 1$ pour un $c \geq 0$, est intégrable (car $\mathbb{E}(|X|) \leq c$).

Bien entendu, la définition précédente couvre l'intégration de toute variable aléatoire, en particulier des transformations (mesurables) d'une variable aléatoire donnée. Ainsi, si X est une variable aléatoire réelle, et $\phi : \mathbb{R} \to \mathbb{R}$ est une fonction borélienne, alors $\mathbb{E}(\phi(X))$ existe dès que $\phi(X)$ est intégrable. Par exemple, étant donné une variable aléatoire réelle X, l'intégrabilité et les intégrales des variables aléatoires X^3, $\cos(X)$, e^X etc. peuvent être envisagées.

Cette observation conduit notamment à la définition des moments et moments absolus d'une variable aléatoire réelle X. Si $r > 0$, le *moment absolu d'ordre r de X* est $\mathbb{E}(|X|^r)$ (supposé fini) ; pour tout entier $r \geq 1$, $\mathbb{E}(X^r)$ est le *moment d'ordre r de X* (sous réserve que $\mathbb{E}(|X|^r) < \infty$).

Les moments sont susceptibles de l'application de l'inégalité de Jensen (Leçon 4) rappelée ici dans la notation probabiliste.

Proposition 2 (Inégalité de Jensen)**.** *Soit $\phi : I \to \mathbb{R}$ une fonction convexe définie sur un intervalle I de $\mathbb{R}$, et soit X une variable aléatoire intégrable à valeurs dans I telle que en outre $\phi(X)$ soit intégrable ; alors*

$$\phi\big(\mathbb{E}(X)\big) \leq \mathbb{E}\big(\phi(X)\big).$$

À titre donc de conséquence, si $0 < r \leq r'$,

$$\big(\mathbb{E}\big(|X|^r\big)\big)^{\frac{r'}{r}} \leq \mathbb{E}\big(|X|^{r'}\big).$$

(Comme déjà évoqué dans la Leçon 4, cette inégalité peut également se déduire de celle de Hölder pour les exposants conjugués $(\frac{r'}{r}, \frac{r'}{r'-r})$ sur l'écriture

$$\mathbb{E}\big(|X|^r\big) = \mathbb{E}\big(|X|^r \cdot 1\big)$$

en rappelant que $\mathbb{E}(1) = 1$.)

En particulier, si X^2 est intégrable (il sera dit que « X est de carré intégrable »), alors X est intégrable. Cette dernière remarque est pertinente dans la définition de la variance d'une variable aléatoire.

Définition 3 (Variance)**.** *Si X est une variable aléatoire réelle de carré intégrable, sa variance est*

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

L'*écart-type* est la racine carré de la variance, souvent noté

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

La variance est positive ou nulle, et égale à 0 si et seulement si X est presque sûrement constante égale à sa moyenne (car si $\mathbb{E}([X - \mathbb{E}(X)]^2) = 0$, alors $[X - \mathbb{E}(X)]^2 = 0$ presque sûrement).

La remarque est banale, mais si X est centrée ($\mathbb{E}(X) = 0$), la variance coïncide avec le moment d'ordre 2, $\mathrm{Var}(X) = \mathbb{E}(X^2)$.

La définition de la variance contient une petite identité à vérifier. D'abord, comme vu plus haut, si $\mathbb{E}(X^2) < \infty$, X est intégrable et donc l'énoncé a un sens bien établi. Ensuite, en développant le carré, pour tout $\omega \in \Omega$,

$$\big[X(\omega) - \mathbb{E}(X)\big]^2 = X(\omega)^2 - 2X(\omega)\,\mathbb{E}(X) + \big(\mathbb{E}(X)\big)^2.$$

En intégrant par rapport à $\mathbb{P}$, par linéarité de l'espérance (intégrale) et le fait que l'espérance d'une variable aléatoire constante est égale à cette constante, il vient

$$\mathbb{E}\big(\big[X - \mathbb{E}(X)\big]^2\big) = \mathbb{E}(X^2) - 2\big(\mathbb{E}(X)\big)^2 + \big(\mathbb{E}(X)\big)^2 = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2.$$

2 Médiane

La notion de médiane complète celle de moyenne, et a toute son importance dans l'estimation et la description d'échantillons de données. Par exemple, pour les notes d'une classe, elle représente une valeur pour laquelle il y a autant de notes supérieures à cette valeur que de notes inférieures à elle.

La médiane de la loi d'une variable aléatoire réelle X sur $(\Omega, \mathcal{A}, \mathbb{P})$ peut être définie sans hypothèse d'intégrabilité. En fait, elle n'est pas unique en général et il convient ainsi de parler d'*une* médiane.

Définition 4 (Médiane)**.** *Une médiane d'une variable aléatoire réelle X est un réel M tel que*

$$\mathbb{P}(X \leq M) \geq \frac{1}{2} \quad \textit{et} \quad \mathbb{P}(X \geq M) \geq \frac{1}{2}.$$

Si X suit la loi de Bernoulli $\mathcal{B}(\frac{1}{2})$ sur $\{0, 1\}$, toute valeur $M \in [0, 1]$ est une médiane. Si la loi de X a une densité $f > 0$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, il n'y a qu'une médiane.

Une caractérisation variationnelle des médianes est proposée dans l'Exercice 7.

3 Théorème de transport

Le théorème de transport décrit l'intégration par rapport à une mesure image (Proposition 6, Leçon 3), et fournit l'outil pour calculer des espérances de fonctions d'une variable aléatoire à partir de la loi de cette dernière.

Il simplement rappelé ici dans le cadre probabiliste, avec la notation espérance $\mathbb{E}$. Il est énoncé pour une variable aléatoire à valeurs dans un espace mesurable général $(E, \mathcal{B})$ mais ne sera appliqué dans la suite que pour $E = \mathbb{R}$, $\mathbb{C}$, ou $\mathbb{R}^d$, muni de la tribu des boréliens.

Théorème 5 (Théorème de transport)**.** *Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace mesurable $(E, \mathcal{B})$, de loi $\mathbb{P}_X$; si $\phi : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable, positive ou intégrable,*

$$\mathbb{E}\big(\phi(X)\big) \,=\, \int_E \phi \, d\,\mathbb{P}_X.$$

Les espérances de fonctions boréliennes d'une variable aléatoire ne dépendent donc que de la loi de celle-ci.

Même si la théorie de l'intégration ne nécessite pas l'introduction d'une variable pour spécifier l'intégration, il pourra à l'occasion être écrit

$$\int_E \phi(x) d\,\mathbb{P}_X(x)$$

(la lettre x étant muette et sans rapport (direct) avec X !).

Il n'est pas inutile de rappeler la première étape, clef, de la démonstration du théorème de transport, à savoir que si $\phi = \mathbb{1}_B$ pour $B \in \mathcal{B}$,

$$\mathbb{E}\big(\mathbb{1}_B(X)\big) \,=\, \mathbb{E}\big(\mathbb{1}_{\{X \in B\}}\big) \,=\, \mathbb{P}(X \in B) \,=\, \mathbb{P}_X(B) \,=\, \int_E \mathbb{1}_B \, d\,\mathbb{P}_X$$

par définition de la loi (mesure image) $\mathbb{P}_X$ de X.

Si X est une variable aléatoire réelle de loi $\mathbb{P}_X$ discrète, par exemple de la forme $\mathbb{P}_X = \sum_{n \in \mathbb{N}} p_n \delta_n$ sur $\mathbb{N}$, l'intégration par transport fournit

$$\mathbb{E}\big(\phi(X)\big) \,=\, \sum_{n \in \mathbb{N}} \phi(n) \, p_n.$$

En fait, une démonstration avantageuse repose sur la relation de Chasles comme

$$\mathbb{E}\big(\phi(X)\big) \,=\, \sum_{n \in \mathbb{N}} \int_{\{X = n\}} \phi(X) d\,\mathbb{P} \,=\, \sum_{n \in \mathbb{N}} \phi(n) \, \mathbb{P}(X = n)$$

puisque la famille $\{X = n\}$, $n \in \mathbb{N}$, forme un système complet d'événements (pour $(\Omega, \mathcal{A}, \mathbb{P})$) et $p_n = \mathbb{P}(X = n)$, $n \in \mathbb{N}$.

Si la loi de X admet une densité f par rapport à la mesure de Lebesgue, $d\,\mathbb{P}_X = f d\lambda$,

$$\mathbb{E}\big(\phi(X)\big) \;=\; \int_{\mathbb{R}} \phi\, f\, d\lambda.$$

Si le théorème de transport est d'une grande utilité pratique pour calculer des espérances, il ne faut pas toujours en abuser, et quand un argument peut se développer directement sur $\mathbb{E}(\phi(X))$ (qui est aussi une intégrale $\int_\Omega \phi(X) d\,\mathbb{P}\,!$), il n'est pas obligatoirement nécessaire de passer par $\int_E \phi\, d\,\mathbb{P}_X$.

4 Quelques calculs d'espérance et de variance

Ce paragraphe reprend les lois de probabilités classiques présentées dans la Leçon 7, et pour chacune d'elle, exprime, avec quelques succinctes indications de calcul, l'espérance et la variance.

Comme déjà signalé, il est important d'insister, à la fois pour la compréhension, l'utilisation pratique et le langage, que ces calculs d'espérance et de variance d'une variable aléatoire ne concernent en fait, à travers le théorème de transport, que la loi de cette variable aléatoire, et qu'il pourrait être question de la même façon de la moyenne et de la variance d'une certaine mesure ou loi de probabilité (de Bernoulli, de Poisson, normale etc.). La variable aléatoire est un représentant de cette loi.

Toutes les variables aléatoires ci-desous sont réelles. C'est un préalable pour tous les exemples présentés que de vérifier les propriétés d'intégrabilité. En fait, le calcul de $\mathbb{E}(X^2)$, qui peut se faire sans hypothèse, fournira à chaque fois la conclusion nécessaire.

Loi de Dirac δ_c. C'est la loi d'une variable aléatoire constante $\mathbb{P}(X = c) = 1$, $c \in \mathbb{R}$. Clairement $\mathbb{E}(X) = c$ et $\mathrm{Var}(X) = 0$ (pour rappel, si X est une variable de carré intégrable telle que $\mathrm{Var}(X) = 0$, alors X est presque sûrement constante, égale à sa moyenne).

Loi de Bernoulli $\mathcal{B}(p)$. Si X est une variable aléatoire de loi de Bernoulli $\mathcal{B}(p)$ de paramètre de succès $p \in]0,1[$ sur $\{0,1\}$ (donc $\mathbb{P}(X=1)=p$), alors

$$\mathbb{E}(X) \;=\; 1 \cdot \mathbb{P}(X=1) + 0 \cdot \mathbb{P}(X=0) \;=\; p.$$

(La moyenne est donc exactement le paramètre de succès.) De la même façon, $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p - p^2 = p(1-p)$.

Loi uniforme (discrète) $\mathcal{U}(\{1,\ldots,n\})$. Si la variable aléatoire X est de loi uniforme sur $\{1,\ldots,n\}$, alors

$$\mathbb{E}(X) \;=\; \sum_{k=1}^{n} k\,\mathbb{P}(X=k) \;=\; \frac{1}{n}\sum_{k=1}^{n} k \;=\; \frac{1}{2}\,(n+1).$$

De la même façon $\mathbb{E}(X^2) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{6}(2n^2+3n+1)$, et donc $\mathrm{Var}(X) = \frac{1}{12}(n^2-1)$.

Loi binomiale $\mathcal{B}(n,p)$. Si X suit la loi binomiale $\mathcal{B}(n,p)$, $n \geq 1$, $p \in]0,1[$,

$$\begin{aligned}
\mathbb{E}(X) \;&=\; \sum_{k=0}^{n} k\,\mathbb{P}(X=k) \\
&=\; \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k} \\
&=\; \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}\,p^k(1-p)^{n-k} \\
&=\; np\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} \;=\; np.
\end{aligned}$$

Un calcul similaire indique que $\mathbb{E}(X^2) = n(n-1)p^2 + np$, et donc $\mathrm{Var}(X) = np(1-p)$.

Loi de Poisson $\mathcal{P}(\theta)$. Si X est une variable aléatoire de loi de Poisson $\mathcal{P}(\theta)$ de paramètre $\theta > 0$,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{n=0}^{\infty} n\,\mathbb{P}(X=n)\\ &= \sum_{n=0}^{\infty} n\,e^{-\theta}\,\frac{\theta^n}{n!}\\ &= \theta \sum_{n=1}^{\infty} e^{-\theta}\,\frac{\theta^{n-1}}{(n-1)!} \;=\; \theta.\end{aligned}$$

Un calcul analogue indique que $\mathbb{E}(X^2) = \theta^2 + \theta$, et donc $\mathrm{Var}(X) = \theta$.

Loi géométrique $\mathcal{G}(p)$. Si X est de loi $\mathcal{G}(p)$, $p \in]0,1[$,

$$\mathbb{E}(X) \;=\; \sum_{n=1}^{\infty} n\,\mathbb{P}(X=n) \;=\; \sum_{n=1}^{\infty} n\,p(1-p)^{n-1}.$$

Des méthodes variées sont à disposition pour le calcul de cette série, comme par exemple les propriétes des séries normalement convergentes, la plupart s'appuyant sur l'observation que $n(1-p)^{n-1}$ est la dérivée, en la variable p, de p^n. Le plus simple est de commencer par une somme finie : si $N \geq 0$, la fonction en $p \in]0,1[$,

$$\sum_{n=0}^{N} (1-p)^n \;=\; \frac{1-(1-p)^{N+1}}{p}$$

est dérivable, de dérivée

$$-\sum_{n=1}^{N} n(1-p)^{n-1} \;=\; \frac{N(1-p)^N p - 1 + (1-p)^{N+1}}{p^2}\,.$$

Quand $N \to \infty$, le membre de droite tend vers $-\frac{1}{p^2}$ et donc $\mathbb{E}(X) = \frac{1}{p}$.

Mais comme déjà présenté en Leçon 3, la théorie de l'intégration de Lebesgue, et le théorème de dérivabilité, offrent un point de vue très efficace. À titre de rappel, si μ est la mesure de comptage sur $\mathbb{N} \cup \{0\}$, et $f_p : \mathbb{N} \cup \{0\} \to \mathbb{R}$ la famille de fonctions paramétrée par $p \in]0, 1[$, $f_p(n) = (1-p)^n$, $n \in \mathbb{N} \cup \{0\}$, alors

$$\int_{\mathbb{N}\cup\{0\}} f_p \, d\mu \, = \, \sum_{n=0}^{\infty} f_p(n) \, = \, \frac{1}{p}$$

(série géométrique). Le théorème de dérivabilité (Théorème 9, Leçon 3) permet de dériver en p sous le signe intégrale, à savoir

$$\int_{\mathbb{N}\cup\{0\}} \frac{d}{dp} f_p \, d\mu \, = \, \sum_{n=0}^{\infty} \frac{d}{dp} f_p(n) \, = \, \frac{d}{dp}\Big(\frac{1}{p}\Big),$$

autrement dit $-\sum_{n=1}^{\infty} n(1-p)^{n-1} = -\frac{1}{p^2}$, ce qui permet de retrouver le résultat. Maintenant, il convient de s'assurer que l'hypothèse de domination dans le théorème de dérivabilité est satisfaite, à savoir que $\big|\frac{d}{dp} f_p\big|$ est majorée par une fonction intégrable, uniformément en p au moins sur un petit intervalle $]p_1, p_2[$, $0 < p_1 < p < p_2 < 1$, autour de p. C'est bien entendu le cas puisque $n(1-p)^{n-1} \leq n(1-p_1)^{n-1}$, $n \in \mathbb{N}$, et la fonction $n \to n(1-p_1)^{n-1}$ est intégrable par rapport à μ.

L'un quelconque des arguments précédents montre de la même façon que $\mathbb{E}(X^2) = \frac{2-p}{p^2}$, et donc $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

Loi uniforme (continue) $\mathcal{U}(a, b)$. Si X est une variable aléatoire de loi uniforme $\mathcal{U}(a, b)$ sur un intervalle (a, b), $a < b$, alors, par intégration par rapport à une mesure à densité par rapport à la mesure de Lebesgue,

$$\mathbb{E}(X) \, = \, \frac{1}{b-a} \int_{(a,b)} x \, d\lambda \, = \, \frac{1}{b-a} \int_a^b x \, dx \, = \, \frac{a+b}{2} \, .$$

De la même façon, $\mathbb{E}(X^2) = \frac{a^2+b^2+ab}{3}$, de sorte que $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$. En particulier, si X est uniforme sur $(0, 1)$, alors $\mathbb{E}(X) = \frac{1}{2}$ et $\mathrm{Var}(X) = \frac{1}{12}$.

Loi exponentielle $\mathcal{E}(\alpha)$. Si X est une variable aléatoire de loi exponentielle $\mathcal{E}(\alpha)$ de paramètre $\alpha > 0$,

$$\mathbb{E}(X) \,=\, \alpha \int_{]0,\infty[} x\, e^{-\alpha x} d\lambda \,=\, \alpha \int_0^\infty x\, e^{-\alpha x} dx \,=\, \frac{1}{\alpha}$$

après une intégration par parties. De la même façon, $\mathbb{E}(X^2) = \frac{2}{\alpha^2}$, de sorte que $\mathrm{Var}(X) = \frac{1}{\alpha^2}$. En particulier, pour la loi $\mathcal{E}(1)$, $\mathbb{E}(X) = \mathrm{Var}(X) = 1$.

Loi gaussienne ou normale $\mathcal{N}(m, \sigma^2)$. Si X est une variable aléatoire de loi normale $\mathcal{N}(0,1)$ de paramètres 0 et 1,

$$\mathbb{E}(X) \,=\, \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\, e^{-\frac{1}{2}x^2} d\lambda \,=\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty x\, e^{-\frac{1}{2}x^2} dx \,=\, 1$$

(car $xe^{-\frac{1}{2}x^2}$ a pour primitive $-e^{-\frac{1}{2}x^2}$). De la même façon, après une intégration par parties, $\mathbb{E}(X^2) = 1$, de sorte que $\mathrm{Var}(X) = 1$.

Après un changement de variable (translation et homothétie), si X est de loi $\mathcal{N}(m, \sigma^2)$ de paramètres $m \in \mathbb{R}$ et $\sigma > 0$, alors $\mathbb{E}(X) = m$ et $\mathrm{Var}(X) = \sigma^2$. En effet, pour toute fonction borélienne ϕ avec les bonnes conditions d'intégrabilité,

$$\int_{\mathbb{R}} \phi(x)\, \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-m)^2} d\lambda(x) \,=\, \int_{\mathbb{R}} \phi(\sigma x + m)\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} d\lambda(x).$$

Cette observation justifie la terminologie de *loi normale* $\mathcal{N}(m, \sigma^2)$ *de moyenne* m *et de variance* σ^2. La loi $\mathcal{N}(0,1)$ sera dite *centrée réduite*.

5 Moyenne et covariance de vecteur aléatoire

Un vecteur aléatoire $X = (X_1, \ldots, X_d)$ est donc une variable aléatoire, sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Il est naturel de s'interroger sur les notions d'espérance et de variance d'un tel vecteur.

L'espérance se définit aisément coordonnée par coordonnée : si chaque variable coordonnée $X_1, \dots, X_d$ est intégrable, l'espérance du vecteur aléatoire X est le *vecteur des espérances*

$$\mathbb{E}(X) \,=\, \big(\mathbb{E}(X_1), \dots, \mathbb{E}(X_d)\big) \,\in\, \mathbb{R}^d.$$

La notion de variance demande de passer aux matrices : si chaque variable coordonnée $X_1, \dots, X_d$ est de carré intégrable, poser

$$\Sigma_{k,\ell} \,=\, \mathbb{E}\big(\big[X_k - \mathbb{E}(X_k)\big]\big[X_\ell - \mathbb{E}(X_\ell)\big]\big)$$

pour $k, \ell = 1, \dots, d$. La matrice (carrée $d \times d$) $\Sigma = (\Sigma_{k,\ell})_{1 \leq k, \ell \leq d}$ est la *matrice de covariance* du vecteur aléatoire $X = (X_1, \dots, X_d)$.

La matrice de covariance Σ est symétrique, sa diagonale est composée des variances des variables marginales $\mathrm{Var}(X_1), \dots, \mathrm{Var}(X_d)$. La matrice Σ est en outre (semi)-définie positive au sens où pour tout $c = (c_1, \dots, c_d) \in \mathbb{R}^d$,

$$\langle \Sigma c, c \rangle \,=\, \sum_{k,\ell=1}^{d} c_k c_\ell \, \Sigma_{k,\ell} \,\geq\, 0$$

(extension donc de la positivité de la variance uni-dimensionnelle). Cette propriété résulte de la linéarité de l'espérance et du théorème de Fubini-Tonelli (pour des sommes) puisque

$$\begin{aligned}
\sum_{k,\ell=1}^{d} c_k c_\ell \, \Sigma_{k,\ell} \,&=\, \sum_{k,\ell=1}^{d} c_k c_\ell \, \mathbb{E}\big(\big[X_k - \mathbb{E}(X_k)\big]\big[X_\ell - \mathbb{E}(X_\ell)\big]\big) \\
&=\, \mathbb{E}\bigg(\sum_{k,\ell=1}^{d} c_k c_\ell \big[X_k - \mathbb{E}(X_k)\big]\big[X_\ell - \mathbb{E}(X_\ell)\big]\bigg) \\
&=\, \mathbb{E}\bigg(\bigg[\sum_{k=1}^{d} c_k \big[X_k - \mathbb{E}(X_k)\big]\bigg]^2\bigg) \,\geq\, 0.
\end{aligned}$$

6 Corrélation

La matrice de covariance d'un vecteur aléatoire $(X_1, \dots, X_d)$ donne lieu à la notion de corrélation, et de coefficient de corrélation.

Deux variables aléatoires (réelles) X et Y de carré intégrable sont dites *non corrélées* si

$$\mathbb{E}(XY) \,=\, \mathbb{E}(X)\,\mathbb{E}(Y).$$

(L'hypothèse de carré intégrable est nécessaire pour donner sens à $\mathbb{E}(XY)$ puisque par l'inégalité de Cauchy-Schwarz $\mathbb{E}(|XY|) \leq (\mathbb{E}(X^2))^{\frac{1}{2}}(\mathbb{E}(Y^2))^{\frac{1}{2}}$). De façon équivalente, par linéarité,

$$\mathbb{E}\big(\big[X - \mathbb{E}(X)\big]\big[Y - \mathbb{E}(Y)\big]\big) \,=\, 0.$$

Autrement dit, les variables centrées $X - \mathbb{E}(X)$ et $Y - \mathbb{E}(Y)$ sont orthogonales dans l'espace de Hilbert $\mathrm{L}^2 = \mathrm{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ pour le produit scalaire

$$\langle U, V\rangle_{\mathrm{L}^2} \,=\, \mathbb{E}(UV), \quad U, V \in \mathrm{L}^2,$$

(voir Leçon 5).

Par analogie avec la matrice de covariance, il est parfois question de la *covariance* des variables X et Y de carré intégrable sous la forme

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

Cette covariance est donc nulle si X et Y sont non corrélées. Si en outre $\sigma(X) = \sqrt{\mathrm{Var}(X)} > 0$ et $\sigma(Y) = \sqrt{\mathrm{Var}(Y)} > 0$, le *coefficient de corrélation* de X et Y est défini par le quotient

$$\frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}\,.$$

(Ce dernier est positivement homogène, invariant par le changement de X en aX, $a > 0$.)

Des variables $X_1, \dots, X_d$ de carré intégrable sont deux à deux non corrélées si et seulement si la matrice de covariance du vecteur aléatoire $(X_1, \dots, X_d)$ est diagonale.

Proposition 6 (Identité de Bienaymé[1])**.** *Si $X_1, \dots, X_n$, $n \geq 1$, de carré intégrable sont deux à deux non corrélées,*

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

Démonstration. Par définition de la variance, la variance de la somme $\sum_{k=1}^{n} X_k$ n'est rien d'autre que la norme L^2 au carré de la somme des variables centrées

$$\begin{aligned}\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) &= \mathbb{E}\left(\left[\sum_{k=1}^{n} X_k - \mathbb{E}\left(\sum_{k=1}^{n} X_k\right)\right]^2\right) \\ &= \left\| \sum_{k=1}^{n} \left[X_k - \mathbb{E}(X_k)\right] \right\|_2^2 .\end{aligned}$$

Par hypothèse de non corrélation, les variables $X_k - \mathbb{E}(X_k)$, $k = 1, \dots, n$, sont orthogonales dans l'espace de Hilbert L^2. Donc, d'après l'égalité de Parseval, l'expression précédente est égale à

$$\sum_{k=1}^{n} \left\| X_k - \mathbb{E}(X_k) \right\|_2^2 = \sum_{k=1}^{n} \mathbb{E}\left(\left[X_k - \mathbb{E}(X_k)\right]^2\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k),$$

ce qui est le résultat annoncé. □

1. Irénée-Jules Bienaymé, probabiliste et statisticien français (1796–1878).

Exercices

(*Une étoile * désignera une question de difficulté supérieure.*)

Exercice 1. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs réelles positives ou nulles. Démontrer que

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)dt.$$

(*Indication* : utiliser par exemple le théorème de Tonelli.)

Démontrer que $\mathbb{E}(X) < \infty$ si et seulement si pour un, ou tout, $\varepsilon > 0$, $\sum_{n\in\mathbb{N}} \mathbb{P}(X \geq \varepsilon n) < \infty$. En déduire que si $\mathbb{E}(X) < \infty$ alors $\lim_{t\to\infty} t\,\mathbb{P}(X \geq t) = 0$.

Exercice 2. Soit X une variable aléatoire de loi uniforme sur $[-1, +1]$; poser $Y = |2X - 1|$. Calculer $\mathbb{E}(Y)$, $\mathbb{E}(X + Y)$, $\mathbb{E}((3Y + X^2)^2)$.

Exercice 3. Soit X une variable aléatoire de loi uniforme sur l'intervalle $[-4, 2]$. Déterminer la loi de la variable aléatoire $Y = \frac{1}{2}(X + 3 + |X - 3|)$. Calculer l'espérance et la variance de Y de deux façons différentes.

Exercice 4. Si x est un réel positif, $\lfloor x \rfloor$ désigne sa partie entière et $x - \lfloor x \rfloor$ sa partie décimale.

Soit X_n une variable aléatoire de loi uniforme $\mathcal{U}(0, n)$ sur l'intervalle $[0, n]$, où n est un entier ≥ 1.

a) Démontrer que $\mathbb{E}(X_n) = n\,\mathbb{E}(X_1)$ et calculer $\mathbb{E}(X_1)$.

b) Quelle est la loi de $\lfloor X_n \rfloor + 1$? de $X_n - \lfloor X_n \rfloor$? de $\lfloor X_n \rfloor + 1 - X_n$?

c) Soit Y_n une variable aléatoire de loi uniforme $\mathcal{U}(\{1, \ldots, n\})$ sur $\{1, \ldots, n\}$. Justifier que $\mathbb{E}(Y_n) = \mathbb{E}(X_n) + \mathbb{E}(X_1)$. En déduire la formule pour $\sum_{k=1}^n k$.

Exercice 5 (*Loi symétrique.*) Une variable aléatoire réelle X est dite de loi symétrique, ou plus simplement symétrique, si $-X$ a même loi que X.

a) Vérifier que la loi normale $\mathcal{N}(0, \sigma^2)$ est symétrique ; donner d'autres exemples.

b) Démontrer que si X est une variable aléatoire de carré intégrable symétrique,

$$\mathbb{E}(X^2) \;=\; 4 \int_0^\infty t\, \mathbb{P}(X \geq t) dt.$$

(*Indication* : $\mathbb{P}(|X| \geq t) = 2\,\mathbb{P}(X \geq t)$, $t > 0$.)

Exercice 6. Si X est une variable aléatoire de carré intégrable, démontrer que

$$\mathrm{Var}(X) \;=\; \inf_{a \in \mathbb{R}} \mathbb{E}\big([X - a]^2\big).$$

Soient $a_1 \leq a_2 \leq \cdots \leq a_n$ des réels, et $m = \frac{1}{n}(a_1 + \cdots + a_n)$; démontrer que

$$\frac{1}{n} \sum_{k=1}^{n} (a_k - m)^2 \;\leq\; \frac{1}{4}\,(a_n - a_1)^2.$$

(*Indication* : trouver la variable aléatoire X et le réel a !)

Exercice 7*. Soit X une variable aléatoire intégrable et M une médiane de la loi de X ; l'objectif de l'exercice est d'établir que

$$\mathbb{E}\big(|X - M|\big) \;=\; \inf_{a \in \mathbb{R}} \mathbb{E}\big(|X - a|\big).$$

a) Si Z est une variable aléatoire intégrable, démontrer que

$$\mathbb{E}\big(|Z|\big) \;=\; \int_0^\infty \mathbb{P}(Z \geq t) dt + \int_{-\infty}^0 \mathbb{P}(Z \leq t) dt.$$

b) Déduire de la question précédente que si $a < b$,

$$\mathbb{E}(|X - b|) - \mathbb{E}(|X - a|) \;=\; \int_a^b \psi(t) dt$$

où $\psi(t) = \mathbb{P}(X \leq t) - \mathbb{P}(X \geq t)$, $t \in \mathbb{R}$.

c) Vérifier que si $t > M$, $\psi(t) \geq 0$, et si $t < M$, $\psi(t) \leq 0$.

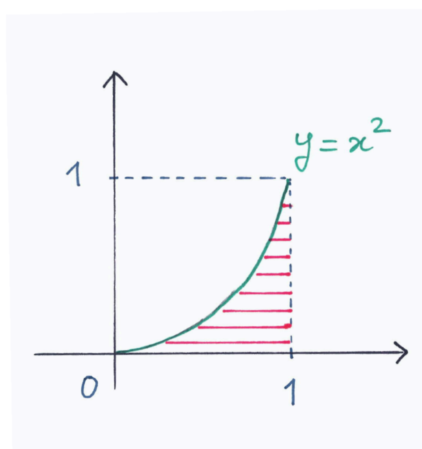
d) Conclure au résultat annoncé.

Exercice 8. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs positives, intégrable ; démontrer que

$$\big[1 + \mathbb{E}(X)\big]\, \mathbb{E}\Big(\frac{1}{1+X}\Big) \;\geq\; 1.$$

Exercice 9. Soit (X, Y) un vecteur aléatoire dans $\mathbb{R}^2$ dont la loi a pour densité par rapport à la mesure de Lebesgue $\lambda^2 = \lambda \otimes \lambda$ sur $\mathbb{R}^2$,

$$f(x, y) = \frac{4y}{x^3}\, \mathbb{1}_{\{0<x<1, 0<y<x^2\}}, \quad (x, y) \in \mathbb{R}^2.$$



Déterminer les lois de X et Y. Calculer $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, $\mathrm{Cov}(X, Y)$.

Exercice 10. Si X et Y sont deux variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de carré intégrable, démontrer que

$$\mathrm{Cov}(X, Y) \;\leq\; \sigma(X)\, \sigma(Y),$$

où il est rappelé que $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ et $\sigma(Y) = \sqrt{\mathrm{Var}(Y)}$ sont les écarts-types de X et Y respectivement (et de même pour $|\mathrm{Cov}(X,Y)|$).

Déduire de cet exercice que si A et B sont des événements quelconques de $\mathcal{A}$,

$$\big|\mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)\big| \,\leq\, \frac{1}{4}\,.$$

Donner un exemple pour lequel la valeur $\frac{1}{4}$ est atteinte.

Exercice 11 (*Entropie*). Si P est une mesure de probabilité sur $\{1,\ldots,n\}$,

$$\mathrm{H}(P) \,=\, -\sum_{k=1}^{n} p_k \ln(p_k),$$

où $p_k = P(\{k\})$ avec la convention $0\ln(0) = 0$, désigne l'entropie de P.

a) Montrer que H est à valeurs dans $\mathbb{R}_+$. Trouver P telle que $\mathrm{H}(P) = 0$. Montrer que la mesure uniforme $\mathcal{U}(\{1,\ldots,n\})$ sur $\{1,\ldots,n\}$ réalise le maximum de H.

b) Si P est une mesure de probabilité sur $\mathbb{N}$, $\mathrm{H}(P) = -\sum_{n\in\mathbb{N}} p_n \ln(p_n)$ désigne de même l'entropie de P, éventuellement infinie. Montrer que la loi géométrique $\mathcal{G}(p)$ de paramètre $p \in \,]0,1[$ réalise le maximum de l'entropie sur l'ensemble des mesures de probabilité sur $\mathbb{N}$ de moyenne inférieure ou égale à $\frac{1}{p}$. (*Indication* : écrire $p_n \ln(p_n) = p_n \ln(\frac{p_n}{(1-p)^{n-1}}) + p_n \ln((1-p)^{n-1})$.)

c) Si P est une mesure de probabilité sur les boréliens de $\mathbb{R}$, de densité f par rapport à la mesure de Lebesgue, soit l'entropie

$$\mathrm{H}(P) \,=\, -\int_{\mathbb{R}} f \ln(f) d\lambda$$

lorsque cette intégrale a un sens, $\mathrm{H}(P) = +\infty$ sinon. Calculer l'entropie de la mesure gaussienne standard $\mathcal{N}(0,1)$ et démontrer qu'elle maximise l'entropie de toute mesure de densité f vérifiant $\int_{\mathbb{R}} x^2 f(x) d\lambda \leq 1$. (*Indication* : vérifier que pour toute densité g, $\int_{\mathbb{R}} f \ln(\frac{f}{g}) d\lambda \geq 0$, et appliquer le résultat à la densité gaussienne.)