

Chapitre 11 :

Variables aléatoires discrètes

II. Espérance et Variance

PC - ENCPB

Jean-Philippe Préaux
<https://www.jean-philippe-preaux.fr>

Espérance et variance

Espérance (VAD réelle ou complexe)

Variance et covariance (VAD réelle)

Fonction génératrice d'une VAD à valeurs dans \mathbb{N}

Résultats asymptotiques

Dans toute cette partie, les variables sont à valeurs réelles ou complexes, voire à valeurs uniquement réelles dans la seconde partie.

Afin de ne pas restreindre le champ d'application, les variables à valeurs réelles positives pourront prendre leur valeur dans $\overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{+\infty\}$ (penser par exemple au rang d'apparition du premier pile de la loi géométrique).

Définition

(Espérance d'une VAD à valeur dans \mathbb{R} ou \mathbb{C})

- Une variable aléatoire discrète X à valeurs dans \mathbb{R} ou \mathbb{C} est dite d'espérance finie lorsque la famille $(x \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ est sommable.
- En se donnant $\text{Im } X$ sous la forme $\text{Im } X = \{x_n \in \mathbb{C} \mid n \in \mathbb{N}\}$, X est d'espérance finie lorsque la série $\sum_{n \geq 0} x_n \mathbb{P}(X = x_n)$ est absolument convergente.
- Si tel est le cas, on appelle espérance de X , noté $\mathbb{E}(X)$, le réel ou complexe :

$$\mathbb{E}(X) = \sum_{n=0}^{+\infty} x_n \underbrace{\mathbb{P}(X = x_n)}_{p_n} = \sum_{n=0}^{+\infty} p_n x_n.$$

Remarques.

- L'espérance de la variable aléatoire X est donc la moyenne des valeurs prises par X pondérée par leur probabilité d'apparition.
- Lorsque $\mathbb{E}(X) = 0$, on parle de variable aléatoire centrée.
- Lorsque X est à valeurs dans \mathbb{R}_+ , il arrive que l'on note $\mathbb{E}(X) = +\infty$ lorsque la variable aléatoire n'admet pas d'espérance (et que, donc, la série diverge vers $+\infty$).

Plus précisément :

Définition

(Espérance d'une VAD à valeurs dans $\overline{\mathbb{R}_+}$)

Soit X une VAD à valeur dans $\overline{\mathbb{R}_+} = \mathbb{R}_+ \cup \{+\infty\}$. On appelle espérance de X , notée $\mathbb{E}(X)$:

$$\mathbb{E}(X) = \sum_{x \in \text{Im } X} x \times \mathbb{P}(X = x)$$

avec les conventions :

- si $\mathbb{P}(X = +\infty) = 0$, alors $x \times \mathbb{P}(X = +\infty) = 0$,
- $\mathbb{E}(X) = +\infty$ lorsque la famille $(x \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ n'est pas sommable.

Pour les variables à valeurs dans $\overline{\mathbb{N}}$, et seulement pour celles là, l'espérance admet l'écriture suivante.

Propriété

(Espérance d'une variable à valeurs dans $\overline{\mathbb{N}}$)

Soit X une VAD à valeurs dans $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$, i.e. $\text{Im } X \subset \mathbb{N} \cup \{+\infty\}$.

Alors :

$$\mathbb{E}(X) = \sum_{n \in \mathbb{N}} \mathbb{P}(X > n)$$

Remarque. La famille peut ne pas être sommable, dans quel cas $\mathbb{E}(X) = +\infty$.

Démonstration. Montrons d'abord que lorsque X n'est pas d'espérance finie, alors $\sum \mathbb{P}(X \geq n)$ diverge (vers $+\infty$). Pour tout $N \in \mathbb{N}^*$, par σ -additivité :

$$\forall n \in \llbracket 1, N \rrbracket, \mathbb{P}(X \geq n) = \left(\sum_{k=n}^{+\infty} \mathbb{P}(X = k) \right) + \mathbb{P}(X = +\infty) \geq \sum_{k=n}^N \mathbb{P}(X = k)$$

d'où :

$$\sum_{n=1}^N \mathbb{P}(X \geq n) \geq \sum_{n=1}^N \sum_{k=n}^N \mathbb{P}(X = k) = \sum_{k=1}^N \sum_{n=1}^k \mathbb{P}(X = k) = \sum_{k=1}^N k \mathbb{P}(X = k)$$

Ainsi, si X n'est pas d'espérance finie, alors $\sum_{k=1}^N k \mathbb{P}(X = k) \xrightarrow{N \rightarrow +\infty} +\infty$ et donc par comparaison de séries à termes positifs, $\sum_{n=1}^N \mathbb{P}(X \geq n) \xrightarrow{N \rightarrow +\infty} +\infty$.

Supposons désormais que X est d'espérance finie ; en particulier $\mathbb{P}(X = +\infty) = 0$. Pour $N \in \mathbb{N}^*$ on a :

$$\begin{aligned} \sum_{n=0}^N n \mathbb{P}(X = n) &= \sum_{n=0}^N n (\mathbb{P}(X \geq n) - \mathbb{P}(X \geq n+1)) \\ &= \sum_{n=0}^N n \mathbb{P}(X \geq n) - \sum_{n=0}^N n \mathbb{P}(X \geq n+1) \\ &= \sum_{n=0}^N n \mathbb{P}(X \geq n) - \sum_{n=1}^{N+1} (n-1) \mathbb{P}(X \geq n) \\ &= \sum_{n=0}^N n \mathbb{P}(X \geq n) - \sum_{n=1}^{N+1} n \mathbb{P}(X \geq n) + \sum_{n=1}^{N+1} \mathbb{P}(X \geq n) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{n=0}^N n \mathbb{P}(X \geq n) - \sum_{n=1}^{N+1} n \mathbb{P}(X \geq n) + \sum_{n=1}^{N+1} \mathbb{P}(X \geq n) \\
 &= -(N+1) \mathbb{P}(X \geq N+1) + \sum_{n=1}^{N+1} \mathbb{P}(X \geq n) \\
 \implies \sum_{n=1}^{N+1} \mathbb{P}(X \geq n) &= \underbrace{\sum_{n=0}^N n \mathbb{P}(X = n)}_{\xrightarrow{N \rightarrow +\infty} \mathbb{E}(X)} + \underbrace{(N+1) \mathbb{P}(X \geq N+1)}_{\leq \sum_{n=N+1}^{+\infty} n \mathbb{P}(X = n) \xrightarrow{N \rightarrow +\infty} 0}
 \end{aligned}$$

puisque la série $\sum_{n \geq 0} n \mathbb{P}(X = n)$ converge, et que le second terme ci-dessus est majoré par son reste d'ordre N . Ainsi :

$$\sum_{n=1}^{+\infty} \mathbb{P}(X \geq n) = \sum_{n=0}^{+\infty} n \mathbb{P}(X = n) = \mathbb{E}(X).$$

■

Exemples. À connaître.

Bien sûr toute VADR à valeurs finies admet une espérance finie.

- Si X suit une loi certaine alors son espérance est $\mathbb{E}(X) = c$ où $\text{Im } X = \{c\}$.
- Si $X \leftrightarrow \mathcal{U}(E)$ suit une loi uniforme alors son espérance est la moyenne de ses valeurs prises : en notant $E = \{x_1, \dots, x_n\}$:

$$\mathbb{E}(X) = \frac{1}{n} \sum_{k=1}^n x_k$$

- Si $X \leftrightarrow \mathcal{B}(p)$ suit la loi de Bernoulli de paramètre p alors $\mathbb{E}(X) = 1 \times p + 0 \times (1 - p) = p$.

- Si $X \hookrightarrow \mathcal{B}(n, p)$ suit la loi binomiale de paramètres n, p alors $\mathbb{E}(X) = np$; en effet :

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{k=0}^n k \times \mathbb{P}(X = k) \\
 &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} && \text{car } \binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1} \text{ si } 1 \leq k \leq n \\
 &= n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} \\
 &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
 &= np \times (p + (1-p))^{n-1} && \text{formule du binôme} \\
 &= np
 \end{aligned}$$

- Si $X \hookrightarrow \mathcal{G}(p)$ suit la loi géométrique de paramètre p , alors X admet une espérance finie et $\mathbb{E}(X) = \frac{1}{p}$; en effet (sous réserve de convergence) :

$$\mathbb{E}(X) = \sum_{n=0}^{+\infty} \mathbb{P}(X > n) \quad \text{Proposition 28}$$

$$= \sum_{n=0}^{+\infty} (1-p)^n \quad \text{Proposition 25}$$

$$= \frac{1}{1 - (1-p)} \quad \text{série géométrique convergente}$$

$$= \frac{1}{p}$$

Espérance d'une loi géométrique :

$$X \hookrightarrow \mathcal{G}(p) \text{ admet une espérance finie : } \mathbb{E}(X) = \frac{1}{p}.$$

- Si $X \hookrightarrow \mathcal{P}(\lambda)$ suit la loi de Poisson de paramètre λ , alors X admet une espérance finie et $\mathbb{E}(X) = \lambda$; en effet (sous réserve de convergence) :

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{n=0}^{+\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \\
 &= \sum_{n=1}^{+\infty} e^{-\lambda} \frac{\lambda^n}{(n-1)!} \\
 &= e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^{n+1}}{n!} \\
 &= \lambda e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \\
 &= \lambda e^{-\lambda} e^{\lambda} && \text{série exponentielle} \\
 &= \lambda
 \end{aligned}$$

Espérance d'une loi de Poisson :

$X \hookrightarrow \mathcal{P}(\lambda)$ admet une espérance finie : $\mathbb{E}(X) = \lambda$.

Formule de transfert

On rappelle que si X est une variable aléatoire discrète, alors pour toute fonction f définie sur $\text{Im } X$, $f(X)$ est aussi une variable aléatoire discrète. On dispose dans ce cas de la formule de transfert qui permet de donner (lorsqu'elle a un sens) son espérance sans connaître la loi de $f(X)$, mais juste celle de X :

Théorème

Soit X une V.A.D. à valeurs réelles, et soit f une fonction définie sur $\text{Im } X$ et à valeurs réelles ou complexes. Alors :

- la variable $f(X)$ est d'espérance finie si et seulement si la famille $(f(x) \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ est sommable et dans ce cas :

$$\mathbb{E}(f(X)) = \sum_{x \in \text{Im } X} f(x) \times \mathbb{P}(X = x)$$

- En particulier si $\text{Im } X = \{x_n \mid n \in \mathbb{N}\}$, $f(X)$ est d'espérance finie si et seulement si la série $\sum f(x_n) \times \mathbb{P}(X = x_n)$ est absolument convergente, et dans ce cas :

$$\mathbb{E}(f(X)) = \sum_{n=0}^{+\infty} f(x_n) \times \mathbb{P}(X = x_n)$$

Démonstration. Par définition $f(X)$ admet une espérance finie si et seulement si la famille $(y \times \mathbb{P}(f(X) = y))_{y \in \text{Im } f(X)}$ est sommable, et dans ce cas $\mathbb{E}(f(X)) = \sum_{y \in \text{Im } f(X)} y \times \mathbb{P}(f(X) = y)$. Pour tout $y \in f(\text{Im } X)$,

$$(f(X) = y) = \bigcup_{x \in f^{-1}(\{y\})} (X = x)$$

c'est une réunion dénombrable, puisque X est discrète, d'événements deux à deux incompatibles, et donc par σ -additivité :

$$\mathbb{P}(f(X) = y) = \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x) \quad (*)$$

on partitionne $\text{Im } X$ en la famille $f^{-1}(\{y\})$ lorsque y décrit $\text{Im } f(X)$. Lorsque la famille $(f(x) \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ est sommable, d'après le théorème de sommation par paquets :

$$\begin{aligned} \sum_{x \in \text{Im } X} f(x) \times \mathbb{P}(X = x) &= \sum_{y \in \text{Im } f(X)} \sum_{x \in f^{-1}(\{y\})} f(x) \times \mathbb{P}(X = x) \\ &= \sum_{y \in \text{Im } f(X)} y \sum_{x \in f^{-1}(\{y\})} \mathbb{P}(X = x) \\ &= \sum_{y \in \text{Im } f(X)} y \times \mathbb{P}(f(X) = y) \\ &= \mathbb{E}(f(X)) \end{aligned}$$

et dans ce cas $f(X)$ admet une espérance finie.

Réciproquement supposons que la famille $(f(x) \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ n'est pas sommable, et montrons que la famille $(y \times \mathbb{P}(f(X) = y))_{y \in \text{Im } f(X)}$ ne l'est pas non plus (et donc que $f(Y)$ n'a pas une espérance finie).

Par définition :

$$\sup \left\{ \sum_{x \in A} |f(x)| \times \mathbb{P}(X = x) \mid A \subset \text{Im } X \text{ fini} \right\} = +\infty.$$

Or d'après (*), pour tout $A \subset \text{Im } X$ (fini) :

$$\sum_{x \in A} |f(x)| \times \mathbb{P}(X = x) \leq \sum_{y \in f(A)} |y| \times \mathbb{P}(f(X) = y)$$

et bien sur lorsque $A \subset \text{Im } X$ est fini, $f(A) \subset \text{Im } f(X)$ est aussi fini. Ainsi :

$$\sup \left\{ \sum_{y \in B} |y| \times \mathbb{P}(f(X) = y) \mid B \subset \text{Im } f(X) \text{ fini} \right\} = +\infty.$$

et donc par définition la famille $(y \times \mathbb{P}(f(X) = y))_{y \in \text{Im } f(X)}$ n'est pas sommable. ■

Remarque. La formule s'applique aussi aux couples ou vecteurs de variables aléatoires. Si (X, Y) est un couple de VAD et $f : \text{Im}(X, Y) \rightarrow \mathbb{C}$ alors si $f(X, Y)$ admet une espérance, on a la formule de transfert :

$$\mathbb{E}(f(X, Y)) = \sum_{(x, y) \in \text{Im}(X, Y)} f(x, y) \times \mathbb{P}(X = x, Y = y)$$

Exercice. Soit $X \leftrightarrow \mathcal{G}(p)$; montrer que la VAD $\frac{1}{X}$ est d'espérance finie et calculer son espérance.

Résolution.

D'après la formule de transfert $\frac{1}{X}$ est d'espérance finie si et seulement si la série $\sum_{n \in \mathbb{N}^*} \frac{1}{n} q^{n-1} p$ est absolument convergente, dans quel cas sa somme vaut $\mathbb{E}(1/X)$. Sous réserve de convergence :

$$\sum_{n \in \mathbb{N}^*} \frac{1}{n} q^{n-1} p = \frac{p}{q} \times \sum_{n \in \mathbb{N}^*} \frac{1}{n} q^n$$

Or $\sum_{n \in \mathbb{N}^*} \frac{1}{n} q^n$ est la série primitive de $\sum_{n \in \mathbb{N}^*} q^{n-1} = \frac{1}{1-q}$

$$= -\frac{p}{q} \ln(1-q) = -\frac{p}{q} \ln(p) = \mathbb{E}(1/X)$$

Proposition

Soient X et Y deux VAD définies sur un même espace probabilisé et à valeurs complexes. Si $|X| \leq Y$ et si Y est d'espérance finie, alors X est aussi d'espérance finie.

Démonstration. On peut supposer que $\text{Im } X$ est infini (dénombrable) sinon la conclusion est claire. Considérons le couple $Z = (X, Y)$; c'est une VAD. Notons $\text{Im } Z = \{(x_n, y_n) \mid n \in \mathbb{N}\}$ où les (x_n, y_n) sont deux à deux distincts.

Notons $f : (x, y) \mapsto x$ et $g : (x, y) \mapsto y$. Alors $Y = g(Z)$ et $X = f(Z)$. Comme Y est d'espérance finie, d'après la formule de transfert :

$$\mathbb{E}(Y) = \sum_{n \in \mathbb{N}} g(x_n, y_n) \times \mathbb{P}(Z = (x_n, y_n)) = \sum_{n \in \mathbb{N}} y_n \times \mathbb{P}(Z = (x_n, y_n))$$

Puisque par hypothèse, pour tout $n \in \mathbb{N}$, $|x_n| \leq y_n$, par comparaison, la série $\sum_{n \in \mathbb{N}} x_n \times \mathbb{P}(Z = (x_n, y_n))$ converge absolument, et d'après le théorème de transfert :

$$\sum_{n \in \mathbb{N}} x_n \times \mathbb{P}(Z = (x_n, y_n)) = \sum_{n \in \mathbb{N}} f(x_n, y_n) \times \mathbb{P}(Z = (x_n, y_n)) = \mathbb{E}(X)$$

et donc X est d'espérance finie. ■

Positivité de l'espérance

Établissons les propriétés de l'espérance.

Propriété

Si X est une VAD réelle à valeurs dans \mathbb{R}_+ et d'espérance finie, alors $\mathbb{E}(X) \geq 0$.

De plus $E(X) = 0$ si et seulement si X est nulle presque sûrement, c'est à dire si $(X = 0)$ est presque sur.

Démonstration. On a $\mathbb{E}(X) = \sum_{x \in \text{Im } X} x \mathbb{P}(X = x)$; puisque $\text{Im } X \subset \mathbb{R}_+$, c'est une somme de termes positifs, et donc, en cas de convergence, $\mathbb{E}(X) \geq 0$.

Si $E(X) = 0$ alors pour tout $x \in \text{Im } X$, $x \mathbb{P}(X = x) = 0$, et donc $x \neq 0 \implies \mathbb{P}(X = x) = 0$. Or

$$1 = \sum_{x \in \text{Im}(X)} \mathbb{P}(X = x) = \mathbb{P}(X = 0) + \underbrace{\sum_{x \in \text{Im}(X)^*} \mathbb{P}(X = x)}_{=0} \implies \mathbb{P}(X = 0) = 1.$$



Linéarité de l'espérance

Propriété

Soient X et Y deux V.A.D. sur un même espace probabilisé et à valeurs dans \mathbb{C} ; si X et Y sont d'espérance finie, alors pour tout $(a, b) \in \mathbb{C}^2$, $aX + bY$ est d'espérance finie et :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

Démonstration. On utilise le lemme suivant qui donne une expression de l'espérance qui n'utilise pas la loi de la VAD.

Lemme

X une VAD à valeurs complexes définie sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ avec Ω au plus dénombrable. Alors X admet une espérance finie si et seulement si la famille $(X(\omega) \times \mathbb{P}(\{\omega\}))_{\omega \in \Omega}$ est sommable et dans ce cas :

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\})$$

Démonstration. (Du Lemme). Sous ces hypothèses, l'application $Id_{\Omega} : \omega \mapsto \omega$ est une VAD puisque $Id_{\Omega}(\Omega) = \Omega$ est dénombrable et que pour tout $x \in \Omega$, $Id_{\Omega}^{-1}(\{x\}) = \{x\}$ appartient à la tribu $\mathcal{P}(\Omega)$. Appliquons la formule de transfert à la VAD $X \circ Id_{\Omega} = X$:

$X \circ Id_{\Omega} = X$ admet une espérance finie si et seulement si la famille $(X(\omega) \times \mathbb{P}(Id_{\Omega} = \omega))_{\omega \in \Omega}$ est sommable, c'est à dire si et seulement si la famille $(X(\omega) \times \mathbb{P}(\{\omega\}))_{\omega \in \Omega}$ est sommable, et dans ce cas :

$$\mathbb{E}(X \circ Id_{\Omega}) = \sum_{\omega \in Id_{\Omega}(\Omega)} X(\omega) \times \mathbb{P}(Id_{\Omega} = \omega)$$

$$\text{c'est à dire } \mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \times \mathbb{P}(\{\omega\})$$

Revenons à la preuve de la propriété. On considère d'abord le cas où l'univers Ω est au plus dénombrable ; notons $\Omega = \{\omega_n \mid n \in I\}$ où $I \subset \mathbb{N}$ est une partie finie de \mathbb{N} ou tout \mathbb{N} . D'après le lemme, puisque X et Y admettent une espérance finie, les familles $(X(\omega_n) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ et $(Y(\omega_n) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ sont sommables, c'est à dire que lorsque $I = \mathbb{N}$, les séries :

$$\sum_{n \in I} X(\omega_n) \times \mathbb{P}(\{\omega_n\}) \quad \text{et} \quad \sum_{n \in I} Y(\omega_n) \times \mathbb{P}(\{\omega_n\})$$

sont absolument convergentes. La famille $((aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\}))_{n \in I}$ est donc sommable, c'est à dire que la somme :

$$\sum_{n \in I} (aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\})$$

est bien définie et à valeur dans \mathbb{C} (c'est clair quand I est fini, et lorsque $I = \mathbb{N}$ la série converge absolument, comme combinaison linéaire de séries absolument convergentes). Or en appliquant le lemme (3 fois) et par linéarité de la somme, on obtient :

$$\begin{aligned} \mathbb{E}(aX + bY) &= \sum_{n \in I} (aX(\omega_n) + bY(\omega_n)) \times \mathbb{P}(\{\omega_n\}) = a \times \sum_{n \in I} X(\omega_n) \times \mathbb{P}(\{\omega_n\}) + b \times \sum_{n \in I} Y(\omega_n) \times \mathbb{P}(\{\omega_n\}) \\ &= a\mathbb{E}(X) + b\mathbb{E}(Y) \end{aligned}$$

Ce qui conclut la cas au plus dénombrable.

Il reste à prouver le cas où Ω est infini indénombrable. Nous allons voir qu'il se ramène au cas dénombrable en construisant des VAD \tilde{X} et \tilde{Y} sur un univers au plus dénombrable, de mêmes lois que X et Y . Supposons que X et Y sont d'espérance finie et considérons le couple de VAD (X, Y) ; c'est une VAD et donc son image

$$(X, Y)(\Omega) = \{(X(\omega), Y(\omega)) \mid \omega \in \Omega\} \subset \mathbb{C}^2$$

est au plus dénombrable. Posons $\tilde{\Omega} = (X, Y)(\Omega)$ son image et munissons-le de la tribu $\mathcal{P}(\tilde{\Omega})$; on définit une probabilité sur $(\tilde{\Omega}, \mathcal{P}(\tilde{\Omega}))$ en posant pour tout $(x, y) \in \tilde{\Omega}$:

$$\mathbb{P}(\{(x, y)\}) = \mathbb{P}(X = x, Y = y)$$

et \mathbb{P} est σ -additive, puisque $\sum_{(x,y) \in (X,Y)(\Omega)} \mathbb{P}(X = x, Y = y) = 1$. C'est un fait déjà remarqué (Chapitre "Espaces probabilisés"), résumé dans le lemme suivant.

Lemme

Soit $\tilde{\Omega} = \{\omega_n \mid n \in I\}$ (avec $I \subset \mathbb{N}$) un ensemble au plus dénombrable et $(p_n)_{n \in I}$ une famille de réels positifs vérifiant $\sum_{i \in I} p_i = 1$. Alors il existe une unique probabilité sur $(\tilde{\Omega}, \mathcal{P}(\tilde{\Omega}))$ vérifiant :

$$\forall i \in I, \mathbb{P}(\{\omega_i\}) = p_i.$$

On a alors pour tout $A \subset \tilde{\Omega}$:

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{\omega_i \in A} p_i.$$

Démonstration. (Du lemme).

Par analyse et synthèse. Soit $A \in \mathcal{P}(\Omega)$; alors A est au plus dénombrable et par σ -additivité si \mathbb{P} s'étend sur $\mathcal{P}(\Omega)$ en une probabilité on a nécessairement $\mathbb{P}(A) = \sum_{i \in A} p_i \in [0, 1]$. Cela définit une application $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ qui est σ -additive. De plus :

$$\mathbb{P}(\Omega) = \sum_{\omega_i \in \Omega} \mathbb{P}(\{\omega_i\}) = \sum_{i \in I} p_i = 1$$

Donc \mathbb{P} ainsi définie est bien une probabilité, et c'est la seule vérifiant $\forall i \in I, \mathbb{P}(\{\omega_i\}) = p_i$. ■

Revenons à la preuve. On a défini un espace probabilisé $(\tilde{\Omega}, \mathcal{P}(\tilde{\Omega}), \mathbb{P})$ d'univers $\tilde{\Omega}$ au plus dénombrable. En notant :

$$\begin{array}{ccc} \tilde{X} : \tilde{\Omega} & \longrightarrow & \mathbb{C} \\ (x, y) & \longmapsto & x \end{array} \quad \text{et} \quad \begin{array}{ccc} \tilde{Y} : \tilde{\Omega} & \longrightarrow & \mathbb{C} \\ (x, y) & \longmapsto & y \end{array}$$

\tilde{X} et \tilde{Y} sont des VAD, puisque :

$$\tilde{X}(\tilde{\Omega}) = X(\Omega) \quad \tilde{Y}(\tilde{\Omega}) = Y(\Omega)$$

sont au plus dénombrables et $\forall x \in \tilde{X}(\tilde{\Omega}), \tilde{X}^{-1}(\{x\}) \in \mathcal{P}(\tilde{\Omega})$ et $\forall y \in \tilde{Y}(\tilde{\Omega}), \tilde{Y}^{-1}(\{y\}) \in \mathcal{P}(\tilde{\Omega})$ sont bien des événements.

Le couple de VAD (\tilde{X}, \tilde{Y}) suit la même loi conjointe que le couple (X, Y) , puisque $(\tilde{X}, \tilde{Y})(\Omega') = \Omega' = (X, Y)(\Omega)$ et

$$\forall (x, y) \in \Omega', \mathbb{P}(\tilde{X} = x, \tilde{Y} = y) = \mathbb{P}(\{(x, y)\}) = \mathbb{P}(X = x, Y = y)$$

et donc \tilde{X} et X suivent la même première loi marginale, \tilde{Y} et Y suivent la même seconde loi marginale, et $a\tilde{X} + b\tilde{Y}$ suit la même loi que $aX + bY$ (la loi de la somme étant uniquement déterminée par la loi conjointe).

Les VAD \tilde{X} et \tilde{Y} étant définie sur un univers au plus dénombrable, la première partie de la preuve s'applique pour montrer que $a\tilde{X} + b\tilde{Y}$ est d'espérance finie et :

$$\mathbb{E}(a\tilde{X} + b\tilde{Y}) = a\mathbb{E}(\tilde{X}) + b\mathbb{E}(\tilde{Y})$$

Donc $aX + bY$ est aussi d'espérance finie et

$$\mathbb{E}(aX + bY) = \mathbb{E}(a\tilde{X} + b\tilde{Y}) = a\mathbb{E}(\tilde{X}) + b\mathbb{E}(\tilde{Y}) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

ce qui conclut la preuve. ■

Croissance de l'espérance

Propriété

Soient X et Y deux VAD sur un même espace probabilisé et à valeurs réelles. Si X et Y sont d'espérance finie, et si $X \leq Y$ alors :

$$\mathbb{E}(X) \leq \mathbb{E}(Y)$$

avec égalité si et seulement si $X = Y$ presque sûrement.

Démonstration. On applique la positivité de l'espérance à $Y - X$ pour obtenir $\mathbb{E}(Y - X) \geq 0$ avec égalité si et seulement si $Y = X$ presque sûrement, puis la linéarité de l'espérance pour en déduire $\mathbb{E}(Y) \geq \mathbb{E}(X)$ avec égalité si et seulement si $Y = X$ presque sûrement. ■

Espérance du produit de VAD indépendantes

Propriété

Soient X et Y deux VAD d'espérance finie sur un même espace probabilisé. Si X et Y sont indépendantes, alors la VAD XY est d'espérance finie et :

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

Démonstration. Les deux familles $(x \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ et $(y \times \mathbb{P}(Y = y))_{y \in \text{Im } Y}$ étant sommable, la famille $(xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y))_{(x,y) \in \text{Im } X \times \text{Im } Y}$ est sommable (Théorème 8) et :

$$\sum_{(x,y) \in \text{Im } X \times \text{Im } Y} xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{x \in \text{Im } X} x \times \mathbb{P}(X = x) \times \sum_{y \in \text{Im } Y} y \times \mathbb{P}(Y = y) \quad (1)$$

Or X et Y étant indépendantes :

$$\sum_{(x,y) \in \text{Im } X \times \text{Im } Y} xy \times \mathbb{P}(X = x)\mathbb{P}(Y = y) = \sum_{(x,y) \in \text{Im}(X,Y)} xy \times \mathbb{P}(X = x, Y = y) \quad (2)$$

On considère $f : (x, y) \mapsto x \times y$ et on applique le théorème de transfert à $f \circ (X, Y) = XY$; on obtient que XY est d'espérance finie et :

$$\begin{aligned} \mathbb{E}(XY) &= \sum_{(x,y) \in \text{Im}(X,Y)} xy \times \mathbb{P}(X = x, Y = y) \\ &= \sum_{(x,y) \in \text{Im } X \times \text{Im } Y} xy \times \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &\stackrel{(1)}{=} \sum_{x \in \text{Im } X} x \times \mathbb{P}(X = x) \times \sum_{y \in \text{Im } Y} y \times \mathbb{P}(Y = y) \\ &\stackrel{(2)}{=} \mathbb{E}(X) \mathbb{E}(Y) \end{aligned}$$



Remarque. Attention l'indépendance est nécessaire pour que XY admette une espérance finie. Il est facile de construire un exemple où X et Y sont d'espérance finie sans que XY ne le soit :

Soit $X = Y : \Omega \rightarrow \mathbb{N}^*$ de loi :

$$\forall n \in \mathbb{N}^*, \mathbb{P}(X = n) = \frac{a}{n^3} \quad \text{avec } a = \left(\sum_{k \in \mathbb{N}^*} \frac{1}{n^3} \right)^{-1}$$

Alors X et Y sont d'espérance finie puisque $\mathbb{E}(X) = \mathbb{E}(Y) = \sum \frac{an}{n^3} = \sum \frac{a}{n^2}$ converge, tandis que $XY = X^2$ ne l'est pas puisque $\mathbb{E}(XY) = \mathbb{E}(X^2) = \sum \frac{an^2}{n^3} = \sum \frac{a}{n}$ diverge.

Ce résultat se généralise aux vecteurs de variables aléatoires indépendantes :

Propriété

Soient X_1, \dots, X_n des VAD d'espérance finie sur un même espace probabilisé. Si X_1, \dots, X_n sont indépendantes, alors la VAD $X_1 \times \dots \times X_n$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \dots \times X_n) = \mathbb{E}(X_1) \times \dots \times \mathbb{E}(X_n)$$

Démonstration. Par récurrence sur $n \geq 2$.

(I) Pour $n = 2$ c'est le résultat précédent.

(H) Supposons l'assertion vraie au rang n et soient X_1, \dots, X_n, X_{n+1} ($n+1$) variables indépendantes. Par hypothèse de récurrence $X_1 \times \dots \times X_n$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \dots \times X_n) = \mathbb{E}(X_1) \times \dots \times \mathbb{E}(X_n)$$

D'après le lemme de coalition $X_1 \times \dots \times X_n$ et X_{n+1} sont indépendantes. Alors d'après la propriété précédente on a que $X_1 \times \dots \times X_n \times X_{n+1}$ est d'espérance finie et :

$$\mathbb{E}(X_1 \times \dots \times X_n \times X_{n+1}) = \mathbb{E}(X_1 \times \dots \times X_n) \times \mathbb{E}(X_{n+1}) = \mathbb{E}(X_1) \times \dots \times \mathbb{E}(X_n) \times \mathbb{E}(X_{n+1})$$

Variance et covariance (VAD réelle)

Dans cette partie toutes les VAD sont à valeurs réelles.

Définition

Soit X une variable aléatoire réelle et soit $p \in \mathbb{N}^*$; on dit que X admet un moment d'ordre p si X^p est d'espérance finie. Le moment d'ordre p de X est alors $\mathbb{E}(X^p)$; c'est un réel.

Proposition

Si X admet un moment d'ordre p , alors X admet un moment d'ordre q pour tout $q \leq p$.

Démonstration. Soit $p, q \in \mathbb{N}^*$ avec $q \leq p$; alors $|X^q| \leq 1 + |X^p|$. Puisque X admet un moment d'ordre p , d'après le théorème de transfert, la famille $(|x|^p \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ est sommable. La famille $(\mathbb{P}(X = x))_{x \in \text{Im } X}$ est aussi sommable, puisque $\sum_{x \in \text{Im } X} \mathbb{P}(X = x) = 1$. Ainsi par linéarité de la sommation et par comparaison, la famille $(|x|^q \times \mathbb{P}(X = x))_{x \in \text{Im } X}$ est sommable ; d'après le théorème de transfert, X^q est d'espérance finie. ■

Variance d'une VADR

Proposition-Définition

Soit X une variable aléatoire réelle discrète.

Si X admet un moment d'ordre 2, alors X est d'espérance finie et on définit la variance de X par :

$$V(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

qui peut aussi s'écrire (formule de Koenig-Hugens) :

$$\mathbb{E}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

La variance de X est un réel positif, et on appelle écart-type le réel

$$\sigma(X) = \sqrt{V(X)}.$$

Démonstration. Puisque X admet un moment d'ordre 2, d'après la proposition précédente, X^2 et X sont d'espérance finie, et par linéarité de l'espérance $(X - \mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2$ est aussi d'espérance finie. Ainsi $V(X)$ est bien définie, et par positivité de l'espérance est à valeurs dans \mathbb{R}_+ . De plus par linéarité de l'espérance :

$$\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

ce qui prouve la formule de Koenig-Hugens.

Remarques.

- Par positivité de l'espérance :

$$V(X) = 0 \quad \text{si et seulement si} \quad X = \mathbb{E}(X) \text{ presque sûrement}$$

Ainsi une VARD X est de variance nulle si et seulement si elle suit une loi quasi-certaine, c'est à dire $\exists c \in \mathbb{R}, \mathbb{P}(X = c) = 1$. (Mais pour autant on n'a pas nécessairement $\text{Im } X = \{c\}$.)

- Si X admet un moment d'ordre 2 alors X admet une variance finie. Réciproquement, si X admet une variance finie, alors par définition X admet une espérance finie, et par linéarité de l'espérance X^2 admet une espérance finie, c'est à dire X admet un moment d'ordre 2. Ainsi une VARD admet un moment d'ordre 2 si et seulement si elle admet (une espérance et) une variance finie(s).

Exemples. Variance de lois usuelles.

- La variance d'une loi quasi-certaine est nulle :

$$V(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = (\mathbb{E}(X) - \mathbb{E}(X))^2 \times 1 + \sum_{x \neq \mathbb{E}(X)} (x - \mathbb{E}(X))^2 \times 0 = 0.$$

- Variance d'une loi de Bernoulli.

Soit $X \hookrightarrow \mathcal{B}(p)$ alors $V(X) = p(1-p)$. En effet : $\text{Im } X = \{0, 1\}$ et $\mathbb{E}(X) = p$, donc d'après la formule de transfert :

$$\begin{aligned} V(X) &= \mathbb{E}((X - p)^2) = (0 - p)^2 \mathbb{P}(X = 0) + (1 - p)^2 \mathbb{P}(X = 1) \\ &= p^2 \times (1 - p) + (1 - p)^2 \times p = p(1 - p)(p + 1 - p) \\ &= p(1 - p) \end{aligned}$$

- Variance d'une loi géométrique.

Soit $X \hookrightarrow \mathcal{G}(p)$ avec $p \in]0, 1[$; alors avec la formule de transfert, X admet un moment d'ordre 2 puisque la série

$$\mathbb{E}(X^2) = \sum_{n \in \mathbb{N}^*} n^2 (1 - p)^{n-1} p$$

converge (par exemple en appliquant d'Alembert).

Calculons sa variance : par linéarité de l'espérance, $X(X-1)$ admet une espérance finie, et avec la formule de transfert :

$$\begin{aligned}\mathbb{E}(X(X-1)) &= \sum_{n \in \mathbb{N}^*} n(n-1)(1-p)^{n-1}p \\ &= p(1-p) \sum_{n \geq 2} n(n-1)(1-p)^{n-2}\end{aligned}$$

on reconnaît la série géométrique dérivée seconde :

$$\begin{aligned}&= p(1-p) \times \frac{d^2}{dp^2} \left(\frac{1}{1-(1-p)} \right) \\ &= p(1-p) \times \frac{2}{p^3} \\ &= \frac{2(1-p)}{p^2}\end{aligned}$$

Donc par linéarité de l'espérance :

$$\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

Et d'après Koenig-Huygens : $V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$

Exercice. Montrer que la variance d'une loi de Poisson de paramètre $\lambda > 0$ est égale à λ .

Résolution.

Soit $X \leftrightarrow \mathcal{P}(\lambda)$; alors $n^2 \mathbb{P}(X = n) = n^2 \frac{\lambda^n}{n!} e^{-\lambda}$ est le terme général d'une série convergente (par d'Alembert par exemple) et donc X admet un moment d'ordre 2.

Donc par linéarité, $X(X - 1)$ admet une espérance finie, et avec la formule de transfert :

$$\mathbb{E}(X(X-1)) = \sum_{n \in \mathbb{N}} n(n-1) \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \lambda^2 \times \sum_{n \geq 2} n(n-1) \frac{\lambda^{n-2}}{n!} = e^{-\lambda} \lambda^2 \times \sum_{n \geq 0} \frac{\lambda^n}{n!} = \lambda^2$$

et d'après Koenig-Huygens et par linéarité de l'espérance :

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Propriété

Soit $(a, b) \in \mathbb{R}^2$ et X une variable aléatoire admettant un moment d'ordre 2; alors

$$V(aX + b) = a^2 V(X)$$

Démonstration. Par linéarité de l'espérance :

$$aX + b - \mathbb{E}(aX + b) = a(X - \mathbb{E}(X))$$

et X et X^2 admettant une espérance finie, il en va de même de $(a(X - \mathbb{E}(X)))^2$; on a alors :

$$V(aX + b) = \mathbb{E}((a(X - \mathbb{E}(X)))^2) = a^2 \mathbb{E}((X - \mathbb{E}(X))^2) = a^2 V(X)$$



Remarque. Une VADR est dite centrée lorsque son espérance est nulle ; elle est dite réduite lorsque sa variance vaut 1. Pour toute variable X admettant un moment d'ordre 2, la variable aléatoire

$$Y = \frac{X - m}{\sigma} \quad \text{avec } m = \mathbb{E}(X) \text{ et } \sigma = \sqrt{V(X)}$$

vérifie : $\mathbb{E}(Y) = 0$ et $V(Y) = 1$, c'est une variable aléatoire centrée réduite.

Proposition

(Condition suffisante d'espérance finie du produit)

Soient X et Y deux VADR sur un même espace probabilisé et admettant un moment d'ordre 2; alors XY est d'espérance finie.

Démonstration. Puisque $X^2 + Y^2 - 2|XY| = (|X| - |Y|)^2 \geq 0$, on a alors $|XY| \leq \frac{1}{2}(X^2 + Y^2)$. Par linéarité de l'espérance $\frac{1}{2}(X^2 + Y^2)$ admet un espérance finie, et avec la proposition 30, XY aussi. ■

Cette propriété nous assure en corollaire qu'une combinaison linéaire de VARD admettant un moment d'ordre 2, admet aussi un moment d'ordre 2. Autrement dit :

Corollaire

(Structure de \mathbb{R} -espace vectoriel)

L'ensemble des variables aléatoires sur un même espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et admettant un moment d'ordre 2 est un \mathbb{R} -espace vectoriel noté $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$.

Démonstration. Soient X et Y deux VADR sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ admettant un moment d'ordre 2. Alors pour tout $\lambda \in \mathbb{R}$, $\lambda \cdot X + Y$ est une VADR sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et puisque :

$$(\lambda \cdot X + Y)^2 = \lambda^2 \cdot X^2 + Y^2 + 2\lambda \cdot XY$$

X^2 , Y^2 et XY admettant une espérance finie, il en va de même, par linéarité de l'espérance, de $(\lambda \cdot X + Y)^2$; la VADR $(\lambda \cdot X + Y)$ admet donc un moment d'ordre 2. ■

Covariance de X et Y

L'espérance finie du produit XY de deux VARD X et Y admettant un moment d'ordre 2 permet surtout de définir la notion de covariance de X et Y .

Proposition-Définition

Soient X et Y deux variables aléatoires sur un même espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et admettant un moment d'ordre 2. On définit la covariance de X et Y comme le réel :

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

En particulier, $V(X) = \text{Cov}(X, X)$.

La covariance peut aussi s'écrire (Formule de Koenig-Huygens) :

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Démonstration. Puisque X et Y admettent un moment d'ordre 2, alors X et Y admettent une espérance finie, ainsi que XY (proposition 40). Par linéarité de l'espérance, $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) = XY - \mathbb{E}(Y)X - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)$ admet donc aussi une espérance finie. La covariance de X , Y est donc bien définie. De plus :

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[XY - \mathbb{E}(Y)X - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)] \\ &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

ce qui établit la formule de Koenig-Huygens. ■

Propriété

(Forme bilinéaire symétrique positive)

Sur $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ la covariance est une forme bilinéaire, symétrique et positive. Autrement dit pour tout $X, Y, Z \in \mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, et pour tout $(a, b) \in \mathbb{R}^2$:

- $\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$,
- $\text{Cov}(X, aY + bZ) = a \text{Cov}(X, Y) + b \text{Cov}(X, Z)$,
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,
- $\text{Cov}(X, X) = V(X) \geq 0$.

Démonstration. Elle est triviale ; la bilinéarité découle de la linéarité de l'espérance, la symétrie de la commutativité de la multiplication, et la positivité découle de la définition. ■

Remarque. La covariance n'est pas un produit scalaire, par absence de définie positivité ; en effet $\text{Cov}(X, X) = 0$ implique que X suit une loi quasi-certaine.

Une propriété très utile de la covariance est la suivante :

Propriété

Deux variables indépendantes ont une covariance nulle

Soient X et Y dans $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$; si X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

Démonstration. Pour deux variables indépendantes $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (propriété 36), d'où la nullité de leur covariance, avec Koenig-Huygens. ■

Remarque. Ainsi la covariance peut être utile pour montrer que deux variables ne sont pas indépendantes. (Car la réciproque est fautive : $\text{Cov}(X, Y)$ peut être nulle sans X et Y ne soient indépendants.).

Inégalité de Cauchy-Schwarz

La covariance n'est pas très loin d'être un produit scalaire, de "norme associée" l'écart-type. L'indépendance implique alors "l'orthogonalité" pour cette analogie. Et on dispose encore d'une inégalité de Cauchy-Schwarz :

Théorème

Soient X et Y deux variables aléatoires dans $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, alors :

$$\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2) \times \mathbb{E}(Y^2).$$

avec égalité si et seulement si X et Y sont proportionnels presque sûrement.

Et :

$$|\text{Cov}(X, Y)| \leq \sigma(X) \times \sigma(Y)$$

avec égalité si et seulement si il existe $(a, b) \in \mathbb{R}^2$ tel que $Y = aX + b$ p.s. ou $X = aY + b$ p.s..

Démonstration. On reprend la preuve de l'inégalité de Cauchy-Schwarz avec le pseudo-produit scalaire $(X, Y) \mapsto \mathbb{E}(XY)$ (on n'a pas le caractère défini, car $\mathbb{E}(X^2) = 0 \implies X = 0$ p.s. (et pas partout) mais la preuve est la même). Supposons que $\mathbb{E}(X^2) \neq 0$; soit $\lambda \in \mathbb{R}$; par positivité de l'espérance :

$$0 \leq \mathbb{E}((\lambda \cdot X + Y)^2) = \lambda^2 \mathbb{E}(X^2) + 2\lambda \cdot \mathbb{E}(XY) + \mathbb{E}(Y^2)$$

C'est un trinôme en λ , dont le discriminant est négatif ou nul :

$$\Delta = 4\mathbb{E}(XY)^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0 \implies \mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

De plus en cas d'égalité, $\Delta = 0$, et donc il existe $\lambda \in \mathbb{R}$, tel que $\mathbb{E}((\lambda \cdot X + Y)^2) = 0$, c'est à dire $\lambda \cdot X + Y = 0$ presque sûrement (par positivité, propriété 31).

Si $\mathbb{E}(X^2) = 0$ alors, toujours d'après la propriété 31, $X = 0$ presque sûrement et là encore $X = 0 \cdot Y$ presque sûrement et $0 = \mathbb{E}(XY)^2 = \mathbb{E}(X^2) \times \mathbb{E}(Y^2)$. Réciproquement si $X = aY$ (ou $Y = aX$) presque sûrement, il est immédiat de vérifier que l'inégalité est une égalité.

En particulier, en appliquant le résultat à $(X - \mathbb{E}(X))$ et $(Y - \mathbb{E}(Y))$:

$$\begin{aligned} \text{Cov}(X, Y)^2 &= \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]^2 \leq \mathbb{E}((X - \mathbb{E}(X))^2) \times \mathbb{E}((Y - \mathbb{E}(Y))^2) \\ \iff |\text{Cov}(X, Y)| &\leq \sigma(X) \times \sigma(Y) \end{aligned}$$

avec égalité si et seulement si il existe $a \in \mathbb{R}$ tel que presque sûrement $X - \mathbb{E}(X) = aY - a\mathbb{E}(Y)$ (ou $Y - \mathbb{E}(Y) = aX - a\mathbb{E}(X)$), et en particulier il existe $(a, b) \in \mathbb{R}^2$ tels que presque sûrement $X = aY + b$ (ou $Y = aX + b$). Réciproquement, si (par exemple) $X = aY + b$ presque sûrement, alors $\mathbb{E}(X) = a\mathbb{E}(Y) + b$ et donc $X - \mathbb{E}(X) = aY + b - a\mathbb{E}(Y) - b = aY - a\mathbb{E}(Y)$ et il y a donc égalité. ■

Coefficient de corrélation

Remarque. Soient $X, Y \in \mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ avec $V(X) \neq 0$ et $V(Y) \neq 0$; on définit leur coefficient de corrélation par la formule :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

c'est un réel dans $[-1, 1]$ qui vérifie :

$$\rho(X, Y) = \begin{cases} 0 & \text{lorsque } X \text{ et } Y \text{ sont indépendantes} \\ \pm 1 & \iff Y = aX + b \text{ p.s. ou } X = aY + b \text{ p.s.} \end{cases}$$

Il est utilisé notamment en statistique pour la régression linéaire. Lorsqu'il est proche de ± 1 , les séries statistiques, que l'on peut voir comme des VADR, sont proches de suivre une relation affine.

Variance d'une somme

Propriété

Pour des VAD dans $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$:

$$V(X + Y) = V(X) + V(Y) + 2 \operatorname{Cov}(X, Y)$$

et plus généralement :

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k) + \sum_{i \neq j} \operatorname{Cov}(X_i, X_j) = \sum_{k=1}^n V(X_k) + 2 \sum_{i < j} \operatorname{Cov}(X_i, X_j)$$

Démonstration. On démontre la deuxième formule, plus générale ; pour cela on rappelle que :

$$\left(\sum_{i=1}^n x_i\right)^2 = \sum_{i=1}^n x_i^2 + 2 \times \sum_{1 \leq i < j \leq n} x_i x_j \quad (*)$$

d'après le formule de Koenig-Huygens :

$$\begin{aligned}
 V\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right) - \left[\mathbb{E}\left(\sum_{i=1}^n X_i\right)\right]^2 \\
 &= \mathbb{E}\left(\sum_{i=1}^n X_i^2 + 2 \times \sum_{1 \leq i < j \leq n} X_i X_j\right) - \left(\sum_{i=1}^n \mathbb{E}(X_i)\right)^2 && (*) \text{ et linéarité de } \mathbb{E} \\
 &= \sum_{i=1}^n \mathbb{E}(X_i^2) + 2 \times \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i X_j) - \sum_{i=1}^n \mathbb{E}(X_i)^2 - 2 \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i) \mathbb{E}(X_j) && \text{idem} \\
 &= \sum_{i=1}^n \left(\mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2\right) + 2 \times \sum_{1 \leq i < j \leq n} \left(\mathbb{E}(X_i X_j) - \mathbb{E}(X_i) \mathbb{E}(X_j)\right) \\
 &= \sum_{i=1}^n V(X_i) + 2 \times \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) && \text{Koenig-Huygens}
 \end{aligned}$$



Variance d'une somme de variables indépendantes

Corollaire

Pour des VARD dans $\mathcal{L}^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ deux à deux indépendantes :

$$V(X + Y) = V(X) + V(Y)$$

et plus généralement :

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k)$$

Remarque. Noter que l'indépendance mutuelle n'est pas nécessaire, l'indépendance 2 à 2 suffit.

Exemple. Variance d'une loi binomiale.

Soit $X \leftrightarrow \mathcal{B}(n, p)$; alors $X = X_1 + \dots + X_n$ avec X_1, \dots, X_n des variables i.i.d. suivant la même loi de Bernoulli $\mathcal{B}(p)$. Alors :

$$V(X) = \sum_{k=1}^n V(X_k) = n \times p \times (1 - p)$$

Exercice. *Navale 2015.* On distribue, aléatoirement, uniformément et indépendamment r boules dans n boîtes B_1, \dots, B_n . On note Y_i la variable aléatoire prenant la valeur 1 si B_i est vide et 0 sinon.

1. Déterminer la loi de Y_i .
2. Calculer $\text{Cov}(Y_i, Y_j)$ pour $i \neq j$.
3. Calculer l'espérance et la variance de la variable aléatoire donnant le nombre de boîtes vides.

Résolution.

1. Chaque Y_i suit une loi de Bernoulli de paramètre :

$$p = \mathbb{P}(Y_i = 1) = \left(\frac{n-1}{n}\right)^r = \left(1 - \frac{1}{n}\right)^r$$

2. On applique la formule de Koenig-Huygens : $Y_i Y_j$ suit aussi une loi de Bernoulli de paramètre :

$$\mathbb{P}(Y_i = 1, Y_j = 1) = \left(\frac{n-2}{n}\right)^r = \left(1 - \frac{2}{n}\right)^r$$

et donc pour $i \neq j$:

$$\text{Cov}(Y_i, Y_j) = \mathbb{E}(Y_i Y_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j) = \left(1 - \frac{2}{n}\right)^r - \left(1 - \frac{1}{n}\right)^{2r}$$

3. Le nombre de boîtes vides est donné par la VARD : $X = \sum Y_i$.

Par linéarité de l'espérance :

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(Y_i) = n \times \left(1 - \frac{1}{n}\right)^r$$

et :

$$V(X) = \sum_{i=1}^n V(Y_i) + 2 \sum_{i < j} \text{Cov}(Y_i, Y_j)$$

Puisque (variance d'une loi de Bernoulli) :

$$V(Y_i) = \left(1 - \left(1 - \frac{1}{n}\right)^r\right) \times \left(1 - \frac{1}{n}\right)^r = \left(1 - \frac{1}{n}\right)^r - \left(1 - \frac{1}{n}\right)^{2r}$$

On obtient :

$$\begin{aligned}V(X) &= n \left(1 - \frac{1}{n}\right)^r - n \left(1 - \frac{1}{n}\right)^{2r} + 2 \times \binom{n}{2} \left(\left(1 - \frac{2}{n}\right)^r - \left(1 - \frac{1}{n}\right)^{2r} \right) \\ &= n \left(1 - \frac{1}{n}\right)^r - n \left(1 - \frac{1}{n}\right)^{2r} + n(n-1) \times \left(\left(1 - \frac{2}{n}\right)^r - \left(1 - \frac{1}{n}\right)^{2r} \right) \\ &= n \left(1 - \frac{1}{n}\right)^r - n^2 \left(1 - \frac{1}{n}\right)^{2r} + n(n-1) \times \left(1 - \frac{2}{n}\right)^r\end{aligned}$$

Fonction génératrice d'une VAD à valeurs dans \mathbb{N}

Dans cette partie toutes les VAD sont à valeurs dans \mathbb{N} .

Proposition

(Série entière de T.G. la loi de $X : \Omega \rightarrow \mathbb{N}$)

Soit X une VAD à valeurs dans \mathbb{N} . Alors la série entière :

$$\sum \mathbb{P}(X = n)z^n$$

- a un rayon de convergence $R \geq 1$,
- converge normalement sur le disque fermé $\overline{D(0, 1)}$.

Démonstration. Puisque $\sum \mathbb{P}(X = n) = 1$, la série converge pour $z = 1$, donc $R \geq 1$. De plus pour tout $z \in \overline{D(0, 1)}$:

$$|z| \leq 1 \implies |\mathbb{P}(X = n)z^n| \leq \mathbb{P}(X = n)$$

et la série $\sum \mathbb{P}(X = n) = 1$ converge ; il y a donc convergence normale sur $\overline{D(0, 1)}$. ■

Fonction génératrice

Définition

Soit X une variable aléatoire discrète à valeurs dans \mathbb{N} . On pose en tout $t \in \mathbb{R}$ où c'est possible,

$$G_X(t) = \mathbb{E}(t^X) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n) \cdot t^n$$

Cette fonction est appelée fonction génératrice de la variable aléatoire X .

Remarques.

- L'égalité entre $\mathbb{E}(t^X)$ et la somme découle de la formule de transfert.
- On a toujours $G_X(1) = 1$ puisque $G_X(1) = \sum \mathbb{P}(X = n) = 1$.
- Si X est à valeurs finies, G_X est une fonction polynomiale.

Exemples. Il faut connaître et savoir retrouver la fonction génératrice des lois usuelles.

- Fonction génératrice d'une loi de Bernoulli.

Soit $X \hookrightarrow \mathcal{B}(p)$. Alors :

$$G_X(t) = \mathbb{P}(X = 0) \cdot t^0 + \mathbb{P}(X = 1) \cdot t^1 = (1 - p) + tp = q + tp$$

la fonction génératrice est affine. Réciproquement si une fonction génératrice G_X est affine, $G_X(t) = q + pt$, alors X suit presque sûrement une loi de Bernoulli, *i.e.* $\text{Im } X \supset \{0, 1\}$ et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = q$.

- Fonction génératrice d'une loi binomiale.

Soit $X \hookrightarrow \mathcal{B}(n, p)$. Alors :

$$G_X(t) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} t^k = (1-p+tp)^n = (q+tp)^n$$

- Fonction génératrice de $\mathcal{U}(\llbracket 1, n \rrbracket)$.

Soit $X \hookrightarrow \mathcal{U}(\llbracket 1, n \rrbracket)$, alors :

$$G_X(t) = \sum_{k=1}^n \frac{1}{n} \times t^k = \frac{t}{n} \times \frac{1-t^n}{1-t} \quad \text{si } t \neq 1 \text{ et } G_X(1) = 1$$

Propriété

(Analytique)

Sous les mêmes hypothèses :

- Le domaine de définition de G_X contient $[-1, 1]$,
- La fonction G_X est continue sur le segment $[-1, 1]$,
- La fonction G_X est de classe \mathcal{C}^∞ sur l'intervalle ouvert $] - 1, 1[$.

Démonstration. Elle découle de faits élémentaires sur les séries entières. Comme vu plus haut il y a convergence normale sur le segment $[-1, 1]$, et donc existence et continuité sur tout le segment. Le rayon de convergence est $R \geq 1$, la somme d'une série entière étant de classe \mathcal{C}^∞ sur l'intervalle ouvert de convergence, elle l'est au moins sur l'ouvert $] - 1, 1[$. ■

L'intérêt de la fonction génératrice G_X réside notamment dans le fait qu'elle caractérise la loi de X :

La fonction génératrice caractérise la loi

Proposition

Soit X une variable aléatoire à valeurs dans \mathbb{N} . La loi de X est entièrement déterminée par sa fonction génératrice G_X . Plus précisément :

$$\forall n \in \mathbb{N}, \mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}.$$

Remarque. En particulier deux variables aléatoires à valeurs dans \mathbb{N} ont même loi si et seulement si elles ont même fonction génératrice.

Démonstration. La fonction G_X est infiniment dérivable sur $] -1, 1[$. D'après le théorème de dérivation terme à terme, $\forall t \in] -1, 1[, \forall n \in \mathbb{N}$:

$$G_X^{(n)}(t) = \sum_{k=n}^{+\infty} \frac{k!}{(k-n)!} \mathbb{P}(X = k) t^{k-n}$$

en particulier $\forall n \in \mathbb{N}$:

$$G_X^{(n)}(0) = n! \times \mathbb{P}(X = n) \implies \mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

Espérance et variance d'une variable aléatoire à valeurs entières sont aussi facilement déterminés à l'aide de la fonction génératrice.

Théorème

(Calcul d'espérance et de variance)

La fonction $t \in [-1, 1] \mapsto G_X(t)$ est :

- dérivable en 1 ssi X est d'espérance finie, on a alors :

$$G'_X(1) = \mathbb{E}(X)$$

- deux fois dérivable en 1 ssi X admet une variance et :

$$G''_X(1) = \mathbb{E}(X(X-1))$$

Remarques.

- Pour retrouver la formule, on retiendra

$$G'_X(t) = \mathbb{E}(Xt^{X-1}) \quad \text{donc} \quad G'_X(1) = \mathbb{E}(X).$$

$$G''_X(t) = \mathbb{E}(X(X-1)t^{X-2}) \quad \text{donc} \quad G''_X(1) = \mathbb{E}(X(X-1)).$$

- Lorsque G_X est deux fois dérivable en 1, la variance s'obtient alors par :

$$V(X) = G''_X(1) + G'_X(1) - G_X'^2(1)$$

puisque :

$$G''_X(1) = \mathbb{E}(X^2) - \mathbb{E}(X) = V(X) + (\mathbb{E}(X))^2 - \mathbb{E}(X) = V(X) + (G'_X(1))^2 - G'_X(1).$$

Démonstration. (La démonstration du sens réciproque est non exigible.)

On note $p_n = \mathbb{P}(X = n)$.

Pour l'espérance : Supposons que X est d'espérance finie, $\mathbb{E}(X) = \sum_{n=0}^{+\infty} n \times p_n$. Alors la série $\sum_{n \geq 1} n t^{n-1} \times p_n$ converge normalement sur $[0, 1]$. Puisque la série $\sum_{n \geq 0} t^n \times p_n$ converge simplement sur $[0, 1]$, d'après le théorème de dérivation \mathcal{C}^1 , $G'_X(1)$ existe et vaut $G'_X(1) = \sum_{n=0}^{+\infty} n \times p_n = \mathbb{E}(X)$.

Réciproquement, supposons que G_X est dérivable en 1 : on a $\lim_{t \rightarrow 1} \frac{G_X(1) - G_X(t)}{1-t} = \ell \in \mathbb{R}$.
Remarquons que pour $t < 1$:

$$\frac{G_X(1) - G_X(t)}{1-t} = \frac{\sum_{n=0}^{+\infty} p_n(1-t^n)}{1-t} = \sum_{n=0}^{+\infty} p_n \left(1 + t + \dots + t^{n-1}\right) \xrightarrow{t \rightarrow 1} \ell$$

Puisque $t \mapsto (1 + t + \dots + t^{n-1})$ est croissante, on a pour tout $t \in [0, 1]^2$

$$\sum_{n=0}^{+\infty} p_n \left(1 + t + \dots + t^{n-1}\right) \leq \ell.$$

Donc, pour tout $N \in \mathbb{N}$, en faisant tendre t vers 1 :

$$\sum_{n=0}^N n p_n \leq \ell$$

La série étant à terme positifs, d'après le théorème de la limite monotone, $\sum_{n \geq 0} n p_n$ converge, ainsi X admet une espérance.

Pour la variance : En utilisant l'argument précédent mais sur la série dérivée, on établit que la fonction $G'_X(t) = \sum_{n=1}^{+\infty} np_n t^{n-1}$ est dérivable en 1 ssi $\sum_{n=1}^{+\infty} n(n-1)p_n < +\infty$, et qu'en cas de convergence, $\lim_{t \rightarrow 1} G''_X(t) = G''_X(1) = \sum_{n=1}^{+\infty} n(n-1)p_n$.

En particulier $\mathbb{E}(X(X-1)) < +\infty$. De plus, puisque G_X est dérivable en 1, $\mathbb{E}(X) < +\infty$, donc $\mathbb{E}(X^2) < +\infty$ et X admet une variance. ■

Théorème

(Fonction génératrice de la somme de variables indépendantes) Soient X et Y deux variables aléatoires à valeurs dans \mathbb{N} indépendantes.

Alors pour tout $t \in [-1, 1]$

$$G_{X+Y}(t) = G_X(t) \times G_Y(t).$$

Exemple. Une loi binomiale $\mathcal{B}(n, p)$ étant somme de lois de Bernoulli $\mathcal{B}(p)$ i.i.d. on retrouve que puisque la fonction génératrice de $\mathcal{B}(p)$ est $pt + q$, celle de $\mathcal{B}(n, p)$ est $(pt + q)^n$.

Démonstration. Les variables X et Y étant indépendantes, d'après le lemme de coalition t^X et t^Y sont indépendantes donc pour $|t| \leq 1$, avec la propriété 36 :

$$G_{X+Y}(t) = \mathbb{E}(t^{X+Y}) = \mathbb{E}(t^X t^Y) = \mathbb{E}(t^X) \mathbb{E}(t^Y) = G_X(t) \times G_Y(t).$$



Exercice. Calculer les fonctions génératrices :

- d'une loi géométrique $\mathcal{G}(p)$
- d'une loi de Poisson $\mathcal{P}(\lambda)$.

Résolution.

- Soit $X \leftrightarrow \mathcal{G}(p)$; posons $q = 1 - p$. Pour tout $k \in \mathbb{N}^*$:

$$\mathbb{P}(X = k)t^k = q^{k-1}pt^k = (qt)^{k-1}pt$$

La série est géométrique, elle converge pour tout $|qt| < 1$ c'est à dire pour $|t| < \frac{1}{q}$.

$$G_X(t) = \sum_{k \geq 1} (qt)^{k-1}pt = pt \sum_{k \geq 0} (qt)^k = \frac{pt}{1 - qt}$$

- Soit $X \leftrightarrow \mathcal{P}(\lambda)$. Pour tout $k \in \mathbb{N}$:

$$\mathbb{P}(X = k)t^k = e^{-\lambda} \frac{(\lambda t)^k}{k!}$$

La série est exponentielle, elle converge pour tout t .

$$G_X(t) = \sum_{k \geq 0} e^{-\lambda} \frac{(\lambda t)^k}{k!} = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda t)^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}$$

- Formulaire à connaître

Loi de X	$\text{Im } X$	$\mathbb{P}(X = k)$	$\mathbb{E}(X)$	$V(X)$	$G_X(t)$
$\mathcal{B}(p)$	$\{0, 1\}$	$\begin{cases} p & \text{si } k = 1 \\ q & \text{si } k = 0 \end{cases}$	p	pq	$pt + q$
$\mathcal{B}(n, p)$	$[[0, n]]$	$\binom{n}{k} p^k q^{n-k}$	np	npq	$(pt + q)^n$
$\mathcal{U}([1, n])$	$[[1, n]]$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2 - 1}{12}$	$\frac{t}{n} \frac{1-t^n}{1-t} \quad (t \neq 1)$
$\mathcal{G}(p)$	\mathbb{N}^*	$q^{k-1} p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pt}{1-qt}$
$\mathcal{P}(\lambda)$	\mathbb{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(t-1)}$

en notant $q = 1 - p$. Il faut savoir retrouver rapidement les fonctions génératrices des lois usuelles.

Résultats asymptotiques

Dans cette partie toutes les variables aléatoires sont à valeurs réelles.

Théorème

(Inégalité de Markov)

Soit X une VADR positive d'espérance finie ; alors pour tout $\varepsilon > 0$,

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}(X).$$

Démonstration. On écrit :

$$\begin{aligned} X &= X \times \mathbb{1}_{X \geq \varepsilon} + X \times \mathbb{1}_{X < \varepsilon} \\ &\geq \varepsilon \times \mathbb{1}_{X \geq \varepsilon} \quad \text{car } X \geq 0 \end{aligned}$$

donc par croissance de l'espérance :

$$\mathbb{E}(X) \geq \mathbb{E}(\varepsilon \times \mathbb{1}_{X \geq \varepsilon}) = \varepsilon \times \mathbb{P}(X \geq \varepsilon).$$

Remarque. Si on ne suppose pas X positive, on peut écrire

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}(|X|).$$

Exemple. Les salaires étant positifs, la proportion de la population recevant plus de 3 fois le salaire moyen est au plus d'un tiers.

(Poser X qui à un salarié associe le rapport entre son salaire et le salaire moyen ; alors $\mathbb{E}(X) = 1$).

L'inégalité de Bienaymé-Tchebychev se déduit de l'inégalité de Markov ; elle est plus fine, faisant intervenir aussi la variance. Elle donne notamment un sens à la variance, comme mesure de la dispersion autour de la valeur moyenne (= espérance).

Théorème

(Inégalité de Bienaymé-Tchebychev)

Soit X une VADR admettant un moment d'ordre 2 ; alors pour tout $\varepsilon > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} V(X).$$

Démonstration. On applique l'inégalité de Markov à la VADR positive $|X - \mathbb{E}(X)|^2$:

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) &= \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq \varepsilon^2) \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E}(|X - \mathbb{E}(X)|^2) = \frac{1}{\varepsilon^2} V(X). \end{aligned}$$

■

Exemple. Soit X est une VADR vérifiant $\mathbb{E}(X) = 0$ et $V(X) = 2$; déterminons $\alpha > 0$ tel que $\mathbb{P}(|X| \geq \alpha) \leq \frac{3}{4}$.

$$\mathbb{P}(|X| \geq \alpha) \leq \frac{3}{4} \iff \mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{3}{4}$$

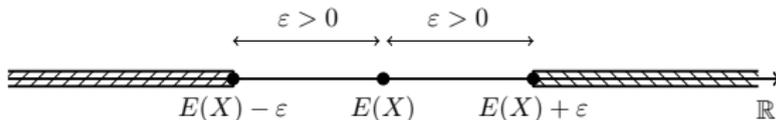
D'après Bienaymé-Tchebychev, il suffit de prendre α tel que $\frac{2}{\alpha^2} \leq \frac{3}{4}$ soit

$$\alpha \geq \frac{2\sqrt{6}}{3}.$$

Remarques.

- La condition $|X - \mathbb{E}(X)| \geq \varepsilon$ s'interprète :

" X s'éloigne de $\mathbb{E}(X)$ d'au moins ε "



• L'écart-type (ainsi que la variance) mesurent la dispersion des valeurs prises par X autour de $\mathbb{E}(X)$. Il est peu probable que X s'écarte de $\mathbb{E}(X)$ d'un écart bien supérieur à son écart-type. Un faible écart-type signifie donc une V.A.R. dont les valeurs sont concentrées autour de son espérance. Avec $\varepsilon = 2 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $2 \times \sigma(X)$ est $\leq \frac{1}{4}$.

Avec $\varepsilon = 5 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $5 \times \sigma(X)$ est $\leq \frac{1}{25}$.

Avec $\varepsilon = 10 \times \sigma(X)$, la probabilité que X s'écarte de $\mathbb{E}(X)$ d'au moins $10 \times \sigma(X)$ est $\leq \frac{1}{100}$.

• La formule de Bienaymé-Tchebychev est très générale : elle s'applique à toute V.A.R. ; en contrepartie la majoration obtenue est relativement faible. Par exemple pour $\varepsilon \leq \sigma(X)$ cette formule ne donne aucune information, puisque dans ce cas le majorant $\left(\frac{\sigma(X)}{\varepsilon}\right)^2$ est supérieur à 1 tandis qu'une probabilité est toujours inférieure à 1. Lorsqu'on connaît la loi de X (et pas seulement $\mathbb{E}(X)$ et $\sigma(X)$) on peut souvent obtenir une majoration plus fine de $\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon)$.

Loi faible des grands nombres

Théorème

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles deux à deux indépendantes et de même loi, admettant un moment d'ordre 2. On pose $S_n = \sum_{k=1}^n X_k$ et $m = \mathbb{E}(X_1)$. Alors :

$$\forall \varepsilon > 0, \quad \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0.$$

Remarque. Ce résultat exprime que la moyenne empirique des résultats d'une expérience aléatoire tend "en probabilité" vers son espérance mathématique.

Démonstration. Par linéarité de l'espérance : $\mathbb{E}(S_n) = n\mathbb{E}(X_1) = nm$ et par indépendance : $V(S_n) = n V(X_1)$, toutes les variables étant de même loi. Ainsi :

$$\mathbb{E}\left(\frac{S_n}{n}\right) = m \quad \text{et} \quad V\left(\frac{S_n}{n}\right) = \frac{V(X_1)}{n}$$

En appliquant l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} V\left(\frac{S_n}{n}\right) = \frac{V(X_1)}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

