Chapitre 38

Espérance et variance.

Sommaire.

Les propositions marquées de ★ sont au programme de colles.

1 Espérance.

1.1 Définition et exemples.

Définition 1

Soit (Ω, P) un espace probabilisé fini et X une variable aléatoire à valeurs dans $\mathbb{K}$ définie sur Ω.
On appelle **espérance** de X et on note $E(X)$ le nombre

$$E(X) = \sum_{x \in X(\Omega)} P(X = x)x.$$

Interprétation

On réalise un certain nombre de fois un expérience conduisant à un résutat numérique : un nombre dans l'ensemble $\{x_1, ..., x_n\}$. Pour $i \in [\![1, n]\!]$, notons f_i la fréquence à laquelle on a obtenu x_i. La valeur moyenne obtenue lors de cette série d'expériences vaut

$$\sum_{i \in [\![1,n]\!]} f_i x_i.$$

Si X est une variable aléatoire à valeurs dans $\{x_1, ..., x_n\}$ et qu'on note $p_i := P(X = x_i)$, l'espérance de X s'écrit

$$\sum_{i \in [\![1,n]\!]} p_i x_i.$$

Dans cette moyenne pondérée, les probabilités ont remplacé les fréquences. Or, on se souvient que le nombre $P(X = x_i)$ est interprété comme la fréquence a priori de l'événement $(X = x_i)$.
Ainsi, le nombre $E(X)$ peut être interprété comme la **valeur moyenne** prise par X a priori.
SI X est comme dans l'exemple un gain à un jeu, $E(X)$ représente le gain moyen a priori, ce que l'on peut espérer gagner en jouant au jeu. On dira aussi que c'est un indicateur de position : $E(X)$ est la position moyenne de la variable X.

Exemple 2

On jette un dé équilibré, si le résultat est 1, 2 ou 3, on ne gagne rien. Si le résultat est 4 ou 5, on gagne dix euros. Si le résultat est 6, on gagne 100 euros. On note X le gain à ce jeu.
Considérons que X est une variable aléatoire dans un espace probabilisé (Ω, P). Calculer $E(X)$.

Solution :

On a :

$$\begin{aligned} E(X) &= \sum_{x \in X(\Omega)} P(X = x)x \\ &= P(X = 0) \cdot 0 + P(X = 10) \cdot 10 + P(X = 100) \cdot 100 \\ &= \frac{2}{6} \cdot 10 + \frac{1}{6} \cdot 100 = \frac{120}{6} \\ &= 20 \end{aligned}$$

Interprétation: À ce jeu, on gagne en moyenne 20 euros.

Proposition 3: Une évidence.

Si deux variables aléatoires ont la même loi, alors elles ont la même espérance.

Preuve :

Lire la définition de l'espérance: elle ne dépend **que** de la loi de X, pas de (Ω, P).

Proposition 4: Éspérance des lois usuelles.

Soit X, Y, Z des variables aléatoires définies sur un espace probabilisé (Ω, P), $n \in \mathbb{N}^*$ et $p \in]0,1[$.

1. **Variable constante.** Si X est la variable aléatoire constante égale à $a \in \mathbb{K}$, alors $\boxed{E(X) = a}$.
2. **Loi de Bernoulli.** Si $Y \sim \mathscr{B}(p)$, $\boxed{E(Y) = p}$. En particulier, $\forall A \in \mathcal{P}(\Omega)$, $\boxed{E(\mathbf{1}_A)} = P(A)$.
3. **Loi binomiale.** Si $Z \sim \mathscr{B}(n,p)$, $\boxed{E(Z) = np}$.

Preuve :

$\boxed{1.}$ $X(\Omega) = \{a\}$ donc $P(X = a) = 1$, d'où $E(X) = P(X = a)a = a$.
$\boxed{2.}$ $Y(\Omega) = \{0,1\}$ et $P(Y = 1) = p$, $P(Y = 0) = 1 - p$ d'où $E(Y) = \cancel{P(Y = 0)} + P(Y = 1) = p$.
$\boxed{3.}$ $Z(\Omega) = [\![0, n]\!]$ et $\forall k \in [\![0, n]\!]$ $P(Z = k) = \binom{n}{k} p^k (1-p)^{n-k}$.
Alors :

$$\begin{aligned}
E(Z) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = n \sum_{k=0}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\
&= np
\end{aligned}$$

Exemple 5: Le pari de Pascal.

Dans son texte célèbre dit du "pari", *(Pensées, fragment 397)* Pascal met en scène un dialogue avec un athée, qu'il veut convaincre de croire en Dieu. Celui qui croit *gage* son énergie, son temps, parfois sa vie entière, et au-dessus de lui *se joue un jeu [...] où il arrivera croix ou pile*, c'est-à-dire qu'à la fin, *Dieu est, ou il n'est pas.* S'il est, celui qui a cru sortira gagnant mais... il sera perdant si Dieu n'existe pas ! Et c'est bien ce qui inquiète l'interlocuteur fictif de Pascal, qui a peur de *gager trop*, de ne pas récupérer sa mise... Savoir si l'on gagnera ou pas à ce jeu revient à savoir si Dieu existe ou pas, et Pascal nous dit que *la raison n'y peut rien déterminer* : pour celui qui croit, il y a *pareil hasard de gain et de perte.*

L'argument de Pascal en faveur de la croyance est le suivant : si on gagne, on gagne l'infini, si on perd, on perd peut-être beaucoup mais on perd une quantité finie (finitude de l'homme...) En moyenne, on gagne l'infini : son raisonnement est un calcul d'espérance ! N'oublions pas que Pascal était mathématicien en plus d'être philosophe, et qu'il s'intéressait notamment au hasard. Posons le calcul de Pascal, en notant X ce que gagne le croyant. Le gain X vaut $+\infty$ si Dieu existe, la perte est finie s'il n'existe pas : disons qu'alors $X = -a$, où a est une quantité finie. Voici ce que le croyant gagne en moyenne :

$$E(X) = \frac{1}{2}(+\infty) + \frac{1}{2}(-a) = +\infty.$$

Notez que le choix de $(\frac{1}{2}, \frac{1}{2})$ comme distribution de probabilités n'a aucune importance, tant que l'on évite une probabilité nulle. Dans *Ma nuit chez Maud*, d'Éric Rohmer, Antoine Vitez (Vidal dans le film) en choisit une autre lorsqu'il fait le pari que l'Histoire a un sens. Jean- Louis Trintignant (Jean-Louis dans le film) lui parle alors d'espérance mathématique. L'extrait est disponible sur Youtube, mais on n'hésitera pas à regarder tout le film !

1.2 Propriétés de l'espérance.

Lemme 6

Soit X une variable aléatoir edéfinie sur (Ω, P) et à valeurs dans $\mathbb{K}$. Alors,

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

Preuve :

L'ensemble $X(\Omega)$ est fini puisque Ω l'est.
Un événement s'écrit comme la «réunion des singletons». Ainsi pour x fixé dans $X(\Omega)$, on a

$$(X = x) = \bigcup_{\omega \in (X = x)} \{\omega\}.$$

La réunion précédente est bien sûr envisagée comme une réunion disjointe, ainsi

$$P(X = x) = \sum_{\omega \in (X = x)} P(\{\omega\}).$$

Réinjectons :

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

Or, $(X = x)_{x \in X(\Omega)}$ est un système complet d'événements, une partition de Ω. La double somme précédente est une somme sur Ω. On a bien

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

Proposition 7: Espérance et inégalités.

Soient X et Y deux variables aléatoires à valeurs *réelles* sur (Ω, P). Alors,

1. Si $\forall \omega \in \Omega,\ X(\omega) \geq 0$, alors $E(X) \geq 0$ (positivité).
2. Si $\forall \omega \in \Omega,\ X(\omega) \geq 0$, et $E(X) = 0$, alors $P(X = 0) = 1$ (cas d'égalité).
3. Si $\forall \omega \in \Omega,\ X(\omega) \leq Y(\omega)$, alors $E(X) \leq E(Y)$ (croissance).
4. $|E(X)| \leq E(|X|)$ (inégalité triangulaire).

Preuve :

$\boxed{1.}$ Supposons que $X(\Omega) \subset \mathbb{R}_+$. Alors :

$$E(X) = \sum_{\omega \in \Omega(P)} P(\{\omega\})X(\omega) \geq 0 \quad \text{car somme de termes positifs.}$$

$\boxed{2.}$ Supposons que $X(\Omega) \subset \mathbb{R}_+$ et $E(X) = 0$. Alors :

$$\sum_{\omega \in \Omega(P)} P(\{\omega\})X(\omega) = 0$$

Alors $\forall \omega \in \Omega \setminus (X = 0),\ P(\{\omega\})X(\omega) = 0$ et $X(\omega) \neq 0$ donc $P(\{\omega\}) = 0$.
Par conséquent, $P(X \neq 0) = 0$, donc $P(X = 0) = 1$.

$\boxed{3.}$ Supposons que $\forall \omega \in \Omega,\ X(\omega) \leq Y(\omega)$. Alors :

$$E(Y) - E(X) = \sum_{\omega \in \Omega} P(\{\omega\})(Y(\omega) - X(\omega)) \geq 0.$$

Donc $E(X) \leq E(Y)$.

$\boxed{4.}$ Ici, on autorise $X : \Omega \to \mathbb{C}$. Alors :

$$\Big|E(X)\Big| = \Big|\sum_{\omega \in \Omega} P(\{\omega\})X(\omega)\Big| \leq \sum_{\omega \in \Omega} P(\{\omega\})|X(\omega)| = E(|X|).$$

Théorème 8: L'espérance est linéaire.

Soient X et Y deux variables aléatoires sur (Ω, P) et à valeurs dans $\mathbb{K}$. Alors,

$$\forall \lambda, \mu \in \mathbb{K},\ E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$$

Notamment, $\forall \lambda, \mu \in \mathbb{K},\ E(\lambda X + \mu) = \lambda E(X) + \mu$.

Dans le cas de n variables aléatoires $X_1, ..., X_n$ et n scalaires $\lambda_1, ..., \lambda_n \in \mathbb{K}$,

$$E\left(\sum_{i=1}^{n} \lambda_i X_i\right) = \sum_{i=1}^{n} \lambda_i E(X_i).$$

Preuve :

Soient $\lambda, \mu \in \mathbb{K}$. On a :

$$\begin{aligned} E(\lambda X + \mu) &= \sum_{\omega \in \Omega} P(\{\omega\})(\lambda X(\omega) + \mu Y(\omega)) \\ &= \lambda \sum_{\omega \in \Omega} P(\{\omega\})X(\omega) + \mu \sum_{\omega \in \Omega} P(\{\omega\})Y(\omega) \\ &= \lambda E(X) + \mu E(Y). \end{aligned}$$

Proposition 9

Soit X une variable aléatoire définie sur (Ω, P) et à valeurs dans $\mathbb{K}$. La variable

$$\overline{X} = X - E(X)$$

est appelée variable aléatoire centrée. On a $E(\overline{X}) = 0$.

Preuve :

On a :

$$E(\overline{X}) = E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0.$$

Théorème 10: Formule du transfert.gc ngcfcn ygnjnfyfngt fn

Soit une v.a. $X : \Omega \to E$ où E est un ensemble sur lequel est défini une application $f : E \to \mathbb{K}$. Alors:

$$E(f(X)) = \sum_{x \in X(\Omega)} P(X = x)f(x).$$

Remarque: On n'a pas besoin de connaître la loi de $f(X)$ pour calculer $E(f(X))$.

Preuve :

On a:

$$\begin{aligned} E(f(X)) &= \sum_{\omega \in \Omega} P(\{\omega\})f(X(\omega)) \\ &= \sum_{x \in X(\Omega)} \sum_{\omega \in (X=x)} P(\{\omega\})f(\underbrace{X(\omega)}_{=x}) \\ &= \sum_{x \in X(\Omega)} f(x) \underbrace{\sum_{\omega \in (X=x)} P(\{\omega\})}_{=P(X=x)} \\ &= \sum_{x \in X(\Omega)} f(x)P(X = x) \end{aligned}$$

Exemple 11: Un calcul simple.

Calcul de $E\left(\frac{1}{1+X}\right)$ où X suit la loi binomiale $\mathscr{B}(n,p)$.

Solution :

On applique la formule du transfert avec $f : x \mapsto \frac{1}{x+1}$, alors :

$$E\left(\frac{1}{X+1}\right) = \sum_{k=0}^{n} P(X = k)f(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{k+1}$$

Or, $\binom{n}{k}(n+1) = \binom{n+1}{k+1}(k+1)$ donc $\binom{n}{k}\frac{1}{k+1} = \binom{n+1}{k+1}\frac{1}{n+1}$ et :

$$\begin{aligned} E\left(\frac{1}{X+1}\right) &= \frac{1}{n+1} \sum_{k=0}^{n} \binom{n+1}{k+1} p^k (1-p)^{n-k} \\ &= \frac{1}{p(n+1)} \sum_{j=1}^{n+1} \binom{n+1}{j} p^j (1-p)^{n+1-j} \\ &= \left(\frac{1}{p(n+1)} \left(1 - (1-p)^{n+1}\right)\right) \end{aligned}$$

Exemple 12: Appliquer la formule de transfert à un couple.

Calcul de $E[(2^X - 1)2^{XY}]$ où X et Y sont deux v.a. indépendantes de loi uniforme sur $[\![0, n-1]\!]$.
On pourra appliquer la formule de transfet au couple (X,Y) dont on connaît la loi conjointe.

1.3 Espérance d'un produit et indépendance.

Théorème 13

Soient X, Y deux variables aléatoires définies sur un même espace probabilisé fini (Ω, P), et à valeurs dans $\mathbb{K}$. SI X et Y sont **indépendantes**, alors

$$E(XY) = E(X)E(Y).$$

Plus généralement, si $(X_i)_{i\in[\![1,n]\!]}$ est une famille de variables aléatoires **indépendantes**,

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i).$$

Preuve :

On applique la formule de transfert à $f : (x,t) \mapsto x \times y$.

$$E(XY) = E[f(X,Y)] = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} P\left((X,Y) = (x,y)\right) f(x,y).$$

Or, pour $(x,y) \in X(\Omega) \times Y(\Omega)$,

$$P\left((X,Y) = (x,y)\right) = P(X = x, Y = y) = P(X = x)P(Y = y).$$

Donc :

$$\begin{aligned} E(XY) &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} P(X = x)P(Y = y)xy \\ &= \left(\sum_{x\in X(\Omega)} P(X = x)x\right) \times \left(\sum_{y\in Y(\Omega)} P(Y = y)y\right) \\ &= E(X)E(Y) \end{aligned}$$

Exemple 14

Soient $X_1, ..., X_n$ des variables aléatoires sur (Ω, P), indépendantes et de même loi de Rademacher, donnée par

$$\forall k \in [\![1,n]\!],\ X_k(\Omega) = \{-1,1\} \quad \text{et} \quad P(X_k = 1) = P(X_k = -1) = \frac{1}{2}.$$

Notons $S_n = \sum_{k=1}^{n} X_k$. Pour $t \in \mathbb{R}$, démontrer que $E(\cos(tS_n)) = \cos^n(t)$.

Solution :

On a $\cos(x) = \mathrm{Re} e^{ix}$.

$$\begin{aligned} E(\cos(tS_n)) &= E\left[\mathrm{Re}(e^{itS_n})\right] \\ &= \mathrm{Re}(E(e^{itS_n})) \\ &= \mathrm{Re}\left(E\left(e^{it\sum_{j=1}^{n} X_j}\right)\right) \\ &= \mathrm{Re}\left(E\left(\prod_{j=1}^{n} e^{itX_j}\right)\right) \\ &= \mathrm{Re}\left(\prod_{j=1}^{n} E(e^{itX_j})\right) \end{aligned}$$

Or $E(e^{itX_j}) = P(X_j = 1)e^{it} + P(X_j = -1)e^{-it} = \frac{1}{2}e^{it} + \frac{1}{2}e^{-it} = \cos t$.
Alors $E\left[\cos(tS_n)\right] = (\cos t)^n$.

2 Variance.

2.1 Définition et exemples.

Définition 15

Soit (Ω, P) un espace probabilisé fini et X une variable aléatoire réelle sur Ω.
On appele **variance** de X et on note $V(X)$ le réel

$$V(X) = E\left[(X - E(X))^2\right].$$

On appelle **écart-type**, parfois noté $\sigma(X)$, le réel $\sqrt{V(X)}$.

Interprétation

L'expression de $V(X)$ donne **l'écart quadratique moyen** de la variable X, par rapport à sa moyenne.

$\boxed{\text{Plus la variance est grande, plus les valeurs prises pas } X \text{ sont «loin» (en moyenne) de } E(X).}$

La variance et l'écart-type sont des indicateurs de dispersion. Il aurait pu sembler plus naturel de considérer la quantité $E(|X - E(X)|)$, mais on verra que le carré qui se trouve dans la définition est bien mieux adapté à la linéarité de l'espérance (voir plus loin le travail sur les sommes de v.a.).

Proposition 16: La variance est quadratique.

Soit X une variable aléatoire réelle sur un espace probabilisé fini. On a, pour tous $a, b \in \mathbb{R}$,

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

Preuve :

Soient $a, b \in \mathbb{R}$.

$$\begin{aligned} V(aX+b) &= E[(aX+b-E(aX+b))^2] \\ &= E[(aX+b-(aE(X)+b))^2] \\ &= E[a^2(X-E(X))^2] \\ &= a^2 E((X-E(X))^2) \\ &= a^2 V(X) \end{aligned}$$

Proposition 17

Soit X une variable aléatoire réelle sur un espace probabilisé fini telle que $\sigma(X) > 0$.
La variable aléatoire $\frac{1}{\sigma(X)} X$ est de variance 1 : elle est dite **réduite**.
La variable aléatoire $\frac{X-E(X)}{\sigma_X}$.

Preuve :

◎ $E\left(\frac{X-E(X)}{\sigma_X}\right) = \frac{1}{\sigma_x}(E(X)-E(X)) = 0.$
◎ $V\left(\frac{X-E(X)}{\sigma_X}\right) = \frac{1}{\sigma_X^2} V(X) = 1$

Proposition 18: Lien avec le moment d'ordre 2.

Soit X une variable aléatoire réelle sur un espace probabilisé fini. On a

$$V(X) = E(X^2) - E(X)^2.$$

Pour $k \in \mathbb{N}$, le nombre $E(X^k)$ est appelé **moment** d'ordre k de la variable X.

Preuve :

On a :

$$\begin{aligned} V(X) &= E\left[(X-E(X))^2\right] \\ &= E\left[X^2 - 2E(X) + (E(X))^2\right] \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2. \end{aligned}$$

Proposition 19: Une évidence.

Si deux variables aléatoires ont la même loi, elles ont la même variance.

Proposition 20: Variance des lois usuelles.

Soit X, Y, Z des variables aléatoires définies sur un espace probabilisé (Ω, P), $n \in \mathbb{N}^*$ et $p \in]0,1[$.

1. **Variable constante.** Si X est la variable aléatoire constante égale à $a \in \mathbb{K}$, alors $\boxed{V(X) = 0}$.
2. **Loi de Bernoulli.** Si $Y \sim \mathscr{B}(p)$, $\boxed{V(Y) = p(1-p)}$.
3. **Loi binomiale.** Si $Z \sim \mathscr{B}(n,p)$, $\boxed{V(Z) = np(1-p)}$.

Preuve :

$\boxed{1.}$ Si X est constante égale à $a \in \mathbb{R}$, alors $V(X) = E[(X-E(X))^2] = E[(a-a)^2] = 0$.
$\boxed{2.}$ Si $Y \sim \mathscr{B}(p)$, alors :

$$V(Y) = E(Y^2) - E(Y)^2 = p - p^2 = p(1-p) \quad \text{car } Y^2 = Y.$$

$\boxed{3.}$ Si $Z \sim \mathscr{B}(n,p)$, on introduit $Z_1, ..., Z_n$ indépendantes de loi $\mathscr{B}(p)$, on note $\widetilde{Z}$ la somme de ces variables.
On a $\widetilde{Z} \sim \mathscr{B}(n,p) \sim Z$ donc $V(\widetilde{Z}) = V(Z)$. Ainsi :

$$V(Z) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = np(1-p).$$

On a utilisé la proposition 26.

Définition 21

On appelle **covariance** de deux variables aléatoires X et Y, et on note $\text{cov}(X,Y)$ le nombre

$$\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))].$$

Lorsque ce nombre est nul, on dit qu'elles sont **décorrélées**.

Interprétation

Lorsque X et Y ont une covariance positive, alors le produit $(X - E(X))(Y - E(Y))$ est positif en moyenne. Cela signifie que lorsque X est supérieure à sa moyenne, Y a tendance à l'être aussi.

Lorsque X et Y ont une covariance négative, alors le produit $(X - E(X))(Y - E(Y))$ est négatif en moyenne. Cela signifie que lorsque X est supérieure à sa moyenne, Y a tendance à être inférieure à la sienne.

Proposition 22

Si X et Y sont deux variables aléatoires réelles, leur covariance s'exprime comme

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y).$$

Si X et Y sont indépendantes, alors elles sont décorrélées :

$$X \perp\!\!\!\perp Y \Longrightarrow \text{cov}(X,Y) = 0.$$

Preuve :

On a :

$$\begin{aligned}
\text{cov}(X,Y) &= E[(X - E(X))(Y - E(Y))] \\
&= E[XY - E(X)Y - E(Y)X + E(X)E(Y)] \\
&= E(XY) - E(X)E(Y) - \cancel{E(Y)E(X)} + \cancel{E(X)E(Y)} \\
&= E(XY) - E(X)E(Y) \\
&= 0 \quad \text{car } E(XY) = E(X)E(Y) \text{ si } X \perp\!\!\!\perp Y.
\end{aligned}$$

Donc si X et Y indépendantes, elles sont décorrélées.

Exemple 23: Décorrélées mais pas indépendantes.

Soient X et Y deux v.a. définies sur un espace probabilisé, indépendantes et de même loi de Bernoulli de paramètre $p \in]0,1[$. On note $U = X + Y$ et $V = X - Y$.
Vérifier que U et V sont décorrélées puis justifier qu'elles ne sont pas indépendantes.

Solution :

On a $E(V) = E(X) - E(Y) = 0$ car mêmes lois.
On a $E(UV) = E(X^2) - E(Y^2) = 0$ car mêmes lois donc même moment d'ordre 2.
Donc $\text{cov}(U,V) = 0$.

On a $P(U = 1, V = 0) = 0$ car $(V = 0) \subset (U \text{ pair})$.
On a $P(U = 1) = 2p(1-p)$ et $P(V = 0) = (1-p)^2 + p^2$.
Donc $P(U = 1)P(V = 0) \neq P(U = 1, V = 0)$, elles ne sont pas indépendantes.

Lemme 24: La covariance est presque un produit scalaire.

Soient $X, \widetilde{X}, Y$ trois variables aléatoires sur (Ω, P).

1. $\text{cov}(X,Y) = \text{cov}(Y,X)$.
2. Pour $(\lambda,\mu) \in \mathbb{R}^2$, $\text{cov}(\lambda X + \mu\widetilde{X}, Y) = \lambda\text{cov}(X,Y) + \mu\text{cov}(\widetilde{X},Y)$.
3. $\text{cov}(X,X) = V(X) \geq 0$.
4. $|\text{cov}(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}$.

Preuve :

$\boxed{1.}$ clair.
$\boxed{2.}$

$$\begin{aligned}
\text{cov}(\lambda X + \mu\widetilde{X}, Y) &= E\left[(\lambda X + \mu\widetilde{X} - E(\lambda X + \mu\widetilde{X}))(Y - E(Y))\right] \\
&= E\left[(\lambda(X - E(X)) + \mu(\widetilde{X} - E(\widetilde{X})))(Y - E(Y))\right] \\
&= ... = \lambda\text{cov}(X,Y) + \mu\text{cov}(\widetilde{X},Y).
\end{aligned}$$

$\boxed{3.}$ $\text{cov}(X,X) = E\left[(X - E(X))^2\right] = V(X) \geq 0$.
$\boxed{4.}$ Posons $f : \lambda \mapsto V(X + \lambda Y)$ positive et $f(\lambda) = \text{cov}(X + \lambda Y, X + \lambda Y) = \text{cov}(X,X) + 2\lambda\text{cov}(X,Y) + \lambda^2\text{cov}(Y,Y)$.
C'est un polynôme de degré inférieur à 2.
Si $\text{cov}(Y,Y) \neq 0$, f est positif de discriminant négatif.
Donc $\Delta = 4(\text{cov}(X,Y)^2 - V(X)V(Y)) \neq 0$ donc $\text{cov}(X,Y)^2 \leq V(X)V(Y)$ donc $|\text{cov}(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}$.

Sinon, $\text{cov}(Y,Y) = 0$ et $Y = E(Y)$, l'inégalité est alors $0 \leq 0$.

Proposition 25: Variance d'une somme de deux variables aléatoires.

Soient X et Y deux variables aléatoires réelles sur (Ω, P). Alors,

$$V(X+Y) = V(X) + V(Y) + 2\text{cov}(X,Y).$$

Dans le cas où X et Y sont décorrélés, on a

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

Cette égalité est notablement vraie lorsque X et Y sont indépendantes.

Preuve :

On a:

$$V(X,Y) = \text{cov}(X+Y, X+Y) = \text{cov}(X,X) + 2\text{cov}(X,Y) + \text{cov}Y,Y = V(X) + 2\text{cov}(X,Y) + V(Y).$$

Si $\text{cov}(X,Y) = 0$...

Proposition 26: Variance d'une somme.

Soit $(X_i)_{i \in [\![1,n]\!]}$ une famille de variables aléatoires réelles sur (Ω, P). On a

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

Si les variables $X_1, ..., X_n$ sont deux à deux décorrélées, alors

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i).$$

Cette égalité est notablement vraie lorsque les variables $X_1, ..., X_n$ sont deux-à-deux indépendantes.

Preuve :

On a:

$$\begin{aligned}
V\left(\sum_{i=1}^{n} X_i\right) &= \text{cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \underbrace{\text{cov}(X_i, X_i)}_{=V(X_i)} + \underbrace{\sum_{i<j} \text{cov}(X_i, X_j) + \sum_{i>j} \text{cov}(X_i, X_j)}_{\text{égales par symétrie.}} \\
&= \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} \text{cov}(X_i, X_j)
\end{aligned}$$

2.2 Inégalités probabilistes.

Proposition 27: Inégalité de Markov.

Soit une variable aléatoire réelle positive $X : \Omega \to \mathbb{R}_+$ et $a \in \mathbb{R}_+$, alors:

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

Preuve :

Soit $\omega \in \Omega$.
Si $\omega \in (X \geq a)$, alors $a\mathbb{1}_{X \geq a}(\omega) = a \leq X(\omega)$.
Si $\omega \notin (X \geq a)$, alors $a\mathbb{1}_{X \geq a}(\omega) = 0 \leq X(\omega)$
Donc pour tout ω, $a\mathbb{1}_{X \geq a}(\omega) \leq X(\omega)$.

On a : $a\mathbb{1}_{X \geq a} \leq X$, par passage à l'espérance : $aE(\mathbb{1}_{X \geq a}) \leq E(X)$.
Or $\mathbb{1}_{X \geq a} \sim \mathscr{B}(p)$ où $p = P(X \geq a)$ donc $E(\mathbb{1}_{X \geq a}) = P(X \geq a)$ donc $aP(X \geq a) \leq E(X)$.
Ainsi, $P(X \geq a) \leq \frac{E(X)}{a}$.

Proposition 28: Inégalité de Bienaymé-Tchebychev.

Soit une variable aléatoire réelle X sur Ω et $a \in \mathbb{R}_+^*$, alors:

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

Preuve :

On a $((|X - E(X|) \geq a) = ((X - E(X))^2 \geq a^2)$.
Soit $Y = (X - E(X))^2$, variable aléatoire positive à laquelle on applique Markov :

$$P(Y \geq a^2) \leq \frac{E(Y)}{a^2}$$

Or $E(Y) = E\left[(X - E(X))^2\right] = V(X)$. Alors:

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

Exemple 29: Des inégalités pas si bonnes dans la pratique.

Soit X une v.a. sur (Ω, P) de loi binomiale $\mathscr{B}(10^3, \frac{1}{2})$. Majorer la probabilité de l'événement $(X \geq 600)$ d'abord avec l'inégalité de Markov, puis avec l'inégalité de Bienaymé-Tchebychev. L'utilisation d'une machine nous donne que cette probabilité est de l'ordre de 10^{-10}. Commenter.

Solution :

On sait le calculer:

$$P(X \geq 600) = \frac{1}{2^{1000}} \sum_{k=600}^{1000} \binom{1000}{k} \sim 10^{-10}$$

L'inégalité de Markov donne :

$$P(X \geq 600) \leq \frac{E(X)}{600} \leq \frac{5}{6}$$

C'est claqué au sol, on ne voit même pas que l'évènement est rare.

Et avec l'inégalité de Bienaymé-Tchebychev ?
On a $(X \geq 600) = (X - E(X) \geq 100) \subset (|X - E(X)| \geq 100)$.
De plus, $V(X) = 10^3 \cdot \frac{1}{2} \cdot \frac{1}{2}$. Alors :

$$P(X \geq 600) \leq P(|X - E(X)| \geq 100) \leq \frac{V(X)}{100^2} \leq \frac{1}{40}.$$

C'est mauvais mais un peu meilleur.

Exemple 30: Inégalité de concentration : distance entre moyennes empiriques et théoriques.

Soit $n \in \mathbb{N}^*$. On considère n variables aléatoires $X_1, ..., X_n$ sur (Ω, P).
On les suppose indépendantes et de même loi. Notons $m = E(X_1)$ et $\sigma^2 = V(X_1)$. On a

$$\forall \varepsilon > 0, \quad P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

La majoration par un $O(\frac{1}{n})$ montre que la probabilité que la moyenne empirique $\frac{1}{n}\sum\limits_{i=1}^{n} X_i$ des n variables aléatoires soit éloignée de sa moyenne théorique μ est plus petite lorsque n est grand.

On observe donc un phénomène de **concentration**: plus n devient grand, plus la variable aléatroie $\frac{1}{n}\sum\limits_{i=1}^{n} X_i$ prend des valeurs concentrées autour de m. Cette convergence est une loi de la nature observéee dans le monde physique. L'observer mathématiquement nous conduit à penser que le modèle probabiliste qui a été développé jusqu'ici n'est pas trop mauvais...

On retrouvera cette description de la concentration en spé avec la Loi faible des grands nombres.

Solution :

Posons $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Alors $E(Y_n) = E\left[\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right] = \frac{1}{n}\sum\limits_{i=1}^{n} E(X_i) = m$.
De plus, $V(Y_n) = V\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum\limits_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum\limits_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}$.
On applique Bienaymé-Tchebychev à Y_n:

$$P(|Y_n - E(Y_n)| \geq \varepsilon) \leq \frac{V(Y_n)}{\varepsilon^2} \quad \text{alors} \quad P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 n} = O\left(\frac{1}{n}\right)$$