Position et dispersion d'une variable aléatoire

Dans tout ce chapitre, (Ω, P) est un espace probabilisé fini.

1 Espérance d'une variable aléatoire complexe

Définition (Espérance d'une variable aléatoire complexe) Soit X une variable aléatoire complexe sur Ω. On appelle *espérance de* X le nombre complexe $\mathrm{E}(X) = \sum_{x \in X(\Omega)} P(X = x)\, x$.
On dit que X est *centrée* si $\mathrm{E}(X) = 0$.

L'espérance de X est une moyenne des valeurs de X, donc un *indicateur de position*. Précisément, chaque valeur x, x décrivant $X(\Omega)$, s'y trouve comptabilisée en proportion de sa probabilité d'occurrence $P(X = x)$. Ainsi, plus $P(X = x)$ est proche de 1, plus x a d'importance dans le calcul.

Exemple Un dé à 6 faces numérotées de 1 à 6 a été truqué de telle sorte que les faces 1, 2 et 3 tombent avec probabilité $\frac{1}{6}$, les faces 4 et 5 avec probabilité $\frac{1}{12}$ et la face 6 avec probabilité $\frac{1}{3}$. Quel numéro obtient-on en moyenne ?

Démonstration Cet énoncé a du sens car $3 \times \frac{1}{6} + 2 \times \frac{1}{12} + \frac{1}{3} = 1$. Notons X le numéro obtenu avec un tel dé après un lancer. Par hypothèse : $P(X = 1) = P(X = 2) = P(X = 3) = \frac{1}{6}$, $P(X = 4) = P(X = 5) = \frac{1}{12}$ et $P(X = 6) = \frac{1}{3}$, donc $\mathrm{E}(X) = \sum_{k=1}^{6} P(X = k)\, k = \frac{1+2+3}{6} + \frac{4+5}{12} + \frac{6}{3} = 1 + \frac{3}{4} + 2 = 3,75$.

Théorème (Espérance des lois usuelles) Soit X une variable aléatoire sur Ω.

(i) **Variables aléatoires constantes :** Si X est constante de valeur m : $\mathrm{E}(X) = m$.

(ii) **Loi uniforme :** Soit $E = \{x_1, \ldots, x_n\}$ une partie de $\mathbb{C}$. Si $X \sim \mathscr{U}(E)$, $\mathrm{E}(X)$ est la moyenne « naturelle » des valeurs $x_1, \ldots, x_n$ de X : $\mathrm{E}(X) = \frac{1}{n} \sum_{k=1}^{n} x_k$.

(iii) **Loi de Bernoulli :** Soit $p \in [0, 1]$. Si $X \sim \mathscr{B}(p)$, alors $\mathrm{E}(X) = p$.

Exemple fondamental : Pour tout événement $A \in \mathscr{P}(\Omega)$: $\mathrm{E}(\mathbb{1}_A) = P(A)$.

(iv) **Loi binomiale :** Soient $n \in \mathbb{N}^*$ et $p \in [0, 1]$. Si $X \sim \mathscr{B}(n, p)$, alors $\mathrm{E}(X) = np$.

Démonstration

(i) Par hypothèse $X(\Omega) = \{m\}$ et $P(X = m) = P(\Omega) = 1$, donc $\mathrm{E}(X) = m \times 1 = m$.

(ii) Pour tout $k \in [\![1, n]\!]$: $P(X = x_k) = \frac{1}{n}$, donc $\mathrm{E}(X) = \sum_{k=1}^{n} \frac{x_k}{n} = \frac{1}{n} \sum_{k=1}^{n} x_k$.

(iii) Par hypothèse $X(\Omega) = \{0, 1\}$, donc $\mathrm{E}(X) = P(X = 1) \times 1 + P(X = 0) \times 0 = p$. Dans le cas où $X = \mathbb{1}_A$ pour un certain $A \in \mathscr{P}(\Omega)$, nous savons déjà que $\mathbb{1}_A \sim \mathscr{B}\big(P(A)\big)$, donc $\mathrm{E}(\mathbb{1}_A) = P(A)$.

(iv) On peut travailler dans l'anneau $\mathbb{R}[T]$: $(T + 1 - p)^n = \sum_{k=0}^{n} \binom{n}{k} T^k (1-p)^{n-k}$, dériver, puis évaluer en p, ou bien utiliser la formule du capitaine.

$$\mathrm{E}(X) = \sum_{k=0}^{n} P(X = k)\, k = \sum_{k=0}^{n} \binom{n}{k} k p^k (1-p)^{n-k} \overset{\text{Capitaine}}{=} n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$\overset{i = k-1}{=} np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{(n-1)-i} = np \times \big(p + (1-p)\big)^{n-1} = np.$$

Théorème (Propriétés de l'espérance) Soient X et Y deux variables aléatoires complexes sur Ω.

(i) **Une autre expression :** $\mathrm{E}(X)=\sum_{\omega\in\Omega}P\big(\{\omega\}\big)X(\omega)$.

(ii) **Linéarité :** Pour tous $\lambda,\mu\in\mathbb{C}$: $\mathrm{E}(\lambda X+\mu Y)=\lambda\mathrm{E}(X)+\mu\mathrm{E}(Y)$.

(iii) **Inégalité triangulaire :** $\big|\mathrm{E}(X)\big|\leqslant\mathrm{E}\big(|X|\big)$.

(iv) **Lien avec les parties réelle et imaginaire :** $\mathrm{E}(X)=\mathrm{E}\big(\mathrm{Re}(X)\big)+\mathrm{i}\,\mathrm{E}\big(\mathrm{Im}(X)\big)$.

La dernière propriété concerne les variables aléatoires réelles.

(v) **Positivité :** Si $X\geqslant 0$, alors $\mathrm{E}(X)\geqslant 0$. **Croissance :** Si $X\leqslant Y$, alors $\mathrm{E}(X)\leqslant\mathrm{E}(Y)$.

Démonstration

(i) Les événements $\{X=x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements, donc :

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

(ii) $$\mathrm{E}(\lambda X+\mu Y)\overset{\text{(i)}}{=}\sum_{\omega\in\Omega}P\big(\{\omega\}\big)(\lambda X+\mu Y)(\omega)=\lambda\sum_{\omega\in\Omega}P\big(\{\omega\}\big)X(\omega)+\mu\sum_{\omega\in\Omega}P\big(\{\omega\}\big)Y(\omega)=\lambda\mathrm{E}(X)+\mu\mathrm{E}(Y).$$

(iii) $$\big|\mathrm{E}(X)\big|\overset{\text{(i)}}{=}\left|\sum_{\omega\in\Omega}P\big(\{\omega\}\big)X(\omega)\right|\leqslant\sum_{\omega\in\Omega}P\big(\{\omega\}\big)\big|X(\omega)\big|\overset{\text{(i)}}{=}\mathrm{E}\big(|X|\big).$$

(iv) $$\mathrm{E}(X)\overset{\text{(i)}}{=}\sum_{\omega\in\Omega}P\big(\{\omega\}\big)X(\omega)=\sum_{\omega\in\Omega}P\big(\{\omega\}\big)\Big(\mathrm{Re}(X)(\omega)+\mathrm{i}\,\mathrm{Im}(X)(\omega)\Big)$$
$$=\sum_{\omega\in\Omega}P\big(\{\omega\}\big)\mathrm{Re}(X)(\omega)+\mathrm{i}\sum_{\omega\in\Omega}P\big(\{\omega\}\big)\mathrm{Im}(X)(\omega)\overset{\text{(i)}}{=}\mathrm{E}\big(\mathrm{Re}(X)\big)+\mathrm{i}\,\mathrm{E}\big(\mathrm{Im}(X)\big).$$

(v) Positivité évidente par définition de l'espérance. Pour la croissance, si $X\leqslant Y$, alors $Y-X\geqslant 0$, donc par linéarité et positivité : $\mathrm{E}(Y)-\mathrm{E}(X)=\mathrm{E}(Y-X)\geqslant 0$.

Exemple Qu'obtient-on en moyenne quand on lance 2 fois un dé à 6 faces et qu'on additionne les résultats obtenus ?

Démonstration Notons X_1 (resp. X_2) le résultat du premier (resp. deuxième) lancer et $S=X_1+X_2$ leur somme. On cherche à calculer $\mathrm{E}(S)$. Bien sûr : $X_1\sim\mathscr{U}\big([\![1,6]\!]\big)$ et $X_2\sim\mathscr{U}\big([\![1,6]\!]\big)$.

Il ne faut SURTOUT PAS ici calculer la loi de S pour en déduire son espérance. Ce serait se donner beaucoup de peine pour rien. Plus simplement, par linéarité : $\mathrm{E}(S)=\mathrm{E}(X_1)+\mathrm{E}(X_2)=\frac{1}{6}\sum_{k=1}^{6}k+\frac{1}{6}\sum_{k=1}^{6}k=2\times\frac{1}{6}\times\frac{6\times 7}{2}=7$.

Exemple Soient $p\in[0,1]$ et $X_1,\ldots,X_n$ des variables de Bernoulli indépendantes de paramètre p définies sur un même espace probabilisé fini. Nous savons déjà que $X_1+\ldots+X_n\sim\mathscr{B}(n,p)$. Un nouveau calcul de l'espérance de la loi binomiale en découle : $\mathrm{E}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{E}(X_i)=\sum_{i=1}^{n}p=np$.

Exemple (Formule du crible) Pour tous $A_1,\ldots,A_n\in\mathscr{P}(\Omega)$: $P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}P\big(A_{i_1}\cap\ldots\cap A_{i_k}\big)$.

Démonstration Cette formule hors programme a été énoncée sans preuve au chapitre « Espaces probabilisés finis et variables aléatoires ». On commence par un calcul d'indicatrices :

$$\mathbb{1}_{A_1\cup\ldots\cup A_n}=1-\mathbb{1}_{\overline{A_1\cup\ldots\cup A_n}}=1-\mathbb{1}_{\overline{A_1}\cap\ldots\cap\overline{A_n}}=1-\prod_{i=1}^{n}\mathbb{1}_{\overline{A_i}}=1-\prod_{i=1}^{n}\big(1-\mathbb{1}_{A_i}\big)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}}$$

et on conclut par linéarité de l'espérance : $$P\left(\bigcup_{i=1}^{n}A_i\right)=\mathrm{E}\big(\mathbb{1}_{A_1\cup\ldots\cup A_n}\big)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\mathrm{E}\big(\mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}}\big)$$
$$=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}P\big(A_{i_1}\cap\ldots\cap A_{i_k}\big).$$

Théorème (Formule de transfert) Soient X une variable aléatoire sur Ω et $f:X(\Omega)\longrightarrow\mathbb{C}$ une fonction. L'espérance de $f(X)$ est entièrement déterminée par f et la loi de X : $\mathrm{E}\big(f(X)\big)=\sum_{x\in X(\Omega)}P(X=x)f(x)$.

Démonstration Les événements $\{X = x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements, donc :
$$\mathrm{E}\big(f(X)\big) = \sum_{\omega\in\Omega} P\big(\{\omega\}\big) f\big(X(\omega)\big) = \sum_{x\in X(\Omega)} \sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big) f\big(X(\omega)\big) = \sum_{x\in X(\Omega)} \sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big) f(x) = \sum_{x\in X(\Omega)} P(X = x) f(x).$$

On dispose bien sûr d'une version de la formule de transfert pour les couples. Si X et Y sont deux variables aléatoires sur Ω et si $f : (X,Y)(\Omega) \longrightarrow \mathbb{C}$ est une fonction, alors $\mathrm{E}\big(f(X,Y)\big) = \sum_{(x,y)\in(X,Y)(\Omega)} P(X = x \text{ et } Y = y) f(x,y)$.

Exemple Soit X_n une variable uniforme sur $[\![1,n]\!]$. La variable aléatoire $U_n = \dfrac{X_n}{n}$ est alors clairement uniforme sur $\left\{\dfrac{k}{n} \;\middle|\; k\in[\![1,n]\!]\right\}$ et pour toute fonction $f \in \mathscr{CM}\big([0,1],\mathbb{C}\big)$, d'après la formule de transfert et le théorème sur les sommes de Riemann : $\mathrm{E}\big(f(U_n)\big) = \sum_{k=1}^{n} P(X_n = k) f\left(\dfrac{k}{n}\right) = \dfrac{1}{n}\sum_{k=1}^{n} f\left(\dfrac{k}{n}\right) \underset{n\to+\infty}{\longrightarrow} \int_0^1 f(x)\,\mathrm{d}x.$

Théorème (Espérance du produit de deux variables aléatoires indépendantes) Soient X et Y deux variables aléatoires complexes sur Ω. Si X et Y sont INDÉPENDANTES : $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

Ce résultat s'étend naturellement à un nombre fini quelconque de variables aléatoires indépendantes.

Démonstration
$$\mathrm{E}(X)\,\mathrm{E}(Y) = \sum_{x\in X(\Omega)} P(X = x)x \times \sum_{y\in Y(\Omega)} P(Y = y)y = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} P(X = x)P(Y = y)xy$$
$$\overset{\text{Indép.}}{=} \sum_{(x,y)\in(X,Y)(\Omega)} P(X = x \text{ et } Y = y)xy \overset{\text{Transfert}}{=} \mathrm{E}(XY).$$

Attention ! L'identité $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$ est fausse en général. Par exemple, pour toute variable de Rademacher X, i.e. uniforme sur $\{-1,1\}$: $\mathrm{E}(X) = 0$ donc $\mathrm{E}(X)^2 = 0$, mais $\mathrm{E}\big(X^2\big) = \mathrm{E}(1) = 1$.

2 Variance, écart-type, covariance

Nous avons introduit l'espérance d'une variable aléatoire COMPLEXE, mais nous nous contenterons d'introduire la variance d'une variable aléatoire RÉELLE.

Définition (Moments, variance et écart-type d'une variable aléatoire réelle) Soit X une variable aléatoire réelle.

- **Moments :** Pour tout $k \in \mathbb{N}$, $\mathrm{E}\big(X^k\big)$ est appelé le *moment d'ordre k de X* et $\mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^k\Big)$ son *moment centré d'ordre k*.
- **Variance :** On appelle *variance de X* le réel positif $\mathrm{V}(X) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big)$ et *écart-type de X* le réel $\sigma(X) = \sqrt{\mathrm{V}(X)}$. On dit que X est *réduite* si $\mathrm{V}(X) = 1$.

L'espérance de X est un indicateur de position, mais les valeurs de X sont-elles plutôt proches de cette valeur moyenne ou plutôt éloignées ? Toute mesure de cette proximité à la moyenne est appelée un *indicateur de dispersion*. L'écart le plus naturel entre X et son espérance est $\big|X - \mathrm{E}(X)\big|$, donc l'écart moyen de X à sa moyenne est $\mathrm{E}\Big(\big|X - \mathrm{E}(X)\big|\Big)$, mais ce n'est pourtant pas l'indicateur de dispersion que les mathématiciens ont choisi de mettre en avant. Ils lui ont préféré la variance, c'est-à-dire l'écart QUADRATIQUE moyen à la moyenne — quadratique parce qu'on passe au carré. Pourquoi ce choix moins naturel au premier abord ? Parce que la variance est plus facile à manipuler d'un point de vue calculatoire, comme on va le voir. Vive les identités remarquables !

Et l'écart-type, quelle différence avec la variance ? Si par exemple X représente une longueur, $\mathrm{V}(X)$ représente une longueur AU CARRÉ, donc il n'est pas possible de comparer directement la position moyenne $\mathrm{E}(X)$ et sa dispersion moyenne $\mathrm{V}(X)$ — un physicien dirait que l'espérance et la variance ne sont pas homogènes. L'écart-type, au contraire, est homogène à une longueur, donc comparable à l'espérance — d'où son intérêt.

Théorème **(Propriétés de la variance)** Soit X une variable aléatoire réelle sur Ω.

(i) **Expression développée :** $\mathrm{V}(X)=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2$.

(ii) **Effet d'une transformation affine :** Pour tous $a,b\in\mathbb{R}$: $\mathrm{V}(aX+b)=a^2\,\mathrm{V}(X)$.

En particulier, si $\sigma(X)>0$, la variable aléatoire $\dfrac{X-\mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.

(iii) **Nullité :** $\mathrm{V}(X)=0$ si et seulement si l'événement $\big\{X=\mathrm{E}(X)\big\}$ est *presque certain*, i.e. de probabilité 1. On dit alors que X est *presque sûrement constante*.

Démonstration

(i) $\mathrm{V}(X)=\mathrm{E}\Big(X^2-2\overbrace{\mathrm{E}(X)}^{\text{Constante}}X+\overbrace{\mathrm{E}(X)^2}^{\text{Constante}}\Big)=\mathrm{E}\big(X^2\big)-2\,\mathrm{E}(X)\mathrm{E}(X)+\mathrm{E}(X)^2=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2$.

(ii) $\mathrm{E}(aX+b)=a\,\mathrm{E}(X)+b$, donc $\mathrm{V}(aX+b)=\mathrm{E}\Big(\big(aX+b-a\,\mathrm{E}(X)-b\big)^2\Big)=a^2\,\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)=a^2\,\mathrm{V}(X)$.

Enfin, si $\sigma(X)>0$: $\mathrm{E}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{E}(X)-\mathrm{E}(X)}{\sigma(X)}=0$ et $\mathrm{V}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{V}(X)}{\sigma(X)^2}=1$.

(iii) D'après la formule de transfert : $\mathrm{V}(X)=\displaystyle\sum_{x\in X(\Omega)}P(X=x)\big(x-\mathrm{E}(X)\big)^2$ et les termes sommés sont tous positifs, donc :

$$\begin{aligned}\mathrm{V}(X)=0 \quad&\Longleftrightarrow\quad \forall x\in X(\Omega),\quad P(X=x)=0 \quad\text{ou}\quad x=\mathrm{E}(X) \quad\Longleftrightarrow\quad \forall x\in X(\Omega)\setminus\big\{\mathrm{E}(X)\big\},\quad P(X=x)=0\\ &\Longleftrightarrow\quad P\big(X=\mathrm{E}(X)\big)=1 \qquad \text{car} \sum_{x\in X(\Omega)}P(X=x)=1.\end{aligned}$$

Attention ! L'événement certain Ω est de probabilité 1 : $P(\Omega)=1$, mais un événement peut être de probabilité 1 sans être égal à Ω tout entier.

— Vous travaillerez en deuxième année avec des espaces probabilisés quelconques, éventuellement infinis. Intéressons-nous au tirage aléatoire d'un point M dans le disque $\mathscr{D}$ de $\mathbb{C}$ de centre 0 et de rayon 1. Intuitivement, et sous réserve qu'on puisse donner un sens à tout cela, la probabilité de tirer un point dans une partie $\mathscr{A}$ de $\mathscr{D}$ vaut :
$$P(M\in\mathscr{A})=\frac{\mathrm{Aire}(\mathscr{A})}{\mathrm{Aire}(\mathscr{D})}=\frac{\mathrm{Aire}(\mathscr{A})}{\pi}.$$
En particulier $P(M=0)=0$, donc $P(M\neq 0)=1$. L'événement $\big\{M\neq 0\big\}$ **N'est PAS** certain car rien n'empêche M de valoir 0, mais il est proche de Ω en ce sens qu'il est de probabilité 1, on dit qu'il est presque certain. On dit ausi que $M\neq 0$ *presque sûrement*.

— Autre exemple très différent. Soit $A\in\mathscr{P}(\Omega)$ un événement pour lequel $P(A)\in\,]0,1[$. En particulier $A\neq\Omega$, autrement dit A n'est pas l'événement certain. Pourtant $P_A(A)=1$, donc A est presque certain au sens de la probabilité P_A.

Définition-théorème **(Covariance de deux variables aléatoires réelles)** Soient X et Y deux variables aléatoires réelles sur Ω. On appelle *covariance de X et Y* le réel $\mathrm{cov}(X,Y)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\Big)$.

On dit que X et Y sont *décorrélées* si $\mathrm{cov}(X,Y)=0$.

Clairement : $\mathrm{V}(X)=\mathrm{cov}(X,X)$ et $\mathrm{cov}(Y,X)=\mathrm{cov}(X,Y)$.

(i) **Expression développée :** $\mathrm{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$.

(ii) **Lien avec l'indépendance :** Si X et Y sont indépendantes, X et Y sont décorrélées.

(iii) **Variance d'une somme :** $\mathrm{V}(X+Y)=\mathrm{V}(X)+2\,\mathrm{cov}(X,Y)+\mathrm{V}(Y)$.

Plus généralement, pour des variables aléatoires $X_1,\dots,X_n$ sur Ω : $\mathrm{V}\left(\displaystyle\sum_{i=1}^n X_i\right)=\displaystyle\sum_{i=1}^n\mathrm{V}(X_i)+2\sum_{1\leqslant i<j\leqslant n}\mathrm{cov}(X_i,X_j)$.

D'après (ii) et (iii), si X et Y sont indépendantes (ou seulement décorrélées) : $\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)$. De même, si $X_1,\dots,X_n$ sont indépendantes (ou seulement deux à deux décorrélées) : $\mathrm{V}\left(\displaystyle\sum_{i=1}^n X_i\right)=\displaystyle\sum_{i=1}^n\mathrm{V}(X_i)$.

Nous donnerons une interprétation géométrique à la covariance et à la corrélation de deux variables aléatoires au prochain chapitre « Espaces préhilbertiens réels ».

Attention ! Deux variables décorrélées ne sont pas indépendantes en général. Par exemple, pour toute variable aléatoire X uniforme sur $\big\{-1,0,1\big\}$: $\mathrm{cov}\big(X,X^2\big)=\mathrm{E}\big(X^3\big)-\mathrm{E}(X)\,\mathrm{E}\big(X^2\big)=0-0\times\mathrm{E}\big(X^2\big)=0$, mais X et X^2 ne sont pas indépendantes car $P\big(X=0 \text{ et } X^2=1\big)=0$ alors que $P(X=0)\,P\big(X^2=1\big)=\dfrac{1}{3}\times\dfrac{2}{3}\neq 0$.

Démonstration

(i) $\operatorname{cov}(X,Y) = \mathrm{E}\Big(XY - X\overbrace{\mathrm{E}(Y)}^{\text{Constante}} - \overbrace{\mathrm{E}(X)}^{\text{Constante}}Y + \overbrace{\mathrm{E}(X)\mathrm{E}(Y)}^{\text{Constante}}\Big) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(Y) + \mathrm{E}(X)\mathrm{E}(Y)$
$= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$

(ii) $$\mathrm{V}(X+Y) = \mathrm{E}\Big(\big(X - \mathrm{E}(X) + Y - \mathrm{E}(Y)\big)^2\Big)$$
$$= \mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2 + 2\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big) + \big(Y-\mathrm{E}(Y)\big)^2\Big) = \mathrm{V}(X) + 2\operatorname{cov}(X,Y) + \mathrm{V}(Y).$$

Théorème (Variance de la loi de Bernoulli et de la loi binomiale) Soient X une variable aléatoire et $p \in [0,1]$.
(i) Si $X \sim \mathscr{B}(p)$, alors $\mathrm{V}(X) = p(1-p)$. (ii) Si $X \sim \mathscr{B}(n,p)$, alors $\mathrm{V}(X) = np(1-p)$.

Démonstration

(i) $\mathrm{E}(X^2) = P(X=1)\times 1^2 + P(X=0)\times 0^2 = p$, donc $\mathrm{V}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = p - p^2 = p(1-p)$.

(ii) Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes sur Ω de même loi de Bernoulli $\mathscr{B}(p)$. Comme $X_1+\ldots+X_n \sim \mathscr{B}(n,p)$, alors par indépendance $\mathrm{V}(X_1+\ldots+X_n) = \mathrm{V}(X_1)+\ldots+\mathrm{V}(X_n) = np(1-p)$. Ce calcul ne concerne a priori pas la variable aléatoire X de l'énoncé, dont on ne sait pas si elle peut être décomposée ou non comme une somme de n variables aléatoires indépendantes de loi $\mathscr{B}(p)$, mais l'espérance d'une variable aléatoire ne dépend que de sa loi, donc le calcul que nous venons de faire dans le cas particulier de la somme $X_1+\ldots+X_n$ est en fait emblématique de toute situation binomiale.

3 Un premier pas vers les grands nombres

Première question : Jusqu'où les valeurs d'une variable aléatoire réelle peuvent-elles être grandes en valeur absolue ?

Théorème (Inégalité de Markov) Soit X une variable aléatoire réelle. Pour tout $a > 0$: $$P\big(|X| \geqslant a\big) \leqslant \frac{\mathrm{E}\big(|X|\big)}{a}.$$

Comme $\big\{|X| > a\big\} \subset \big\{|X| \geqslant a\big\}$, on peut remplacer $P\big(|X| \geqslant a\big)$ par $P\big(|X| > a\big)$ si on veut.

Démonstration Tout repose sur les inégalités suivantes : $\mathbb{1}_{\{|X|\geqslant a\}}a \leqslant \mathbb{1}_{\{|X|\geqslant a\}}|X| \leqslant |X|$. Aussitôt, par croissance de l'espérance : $a\,P\big(|X| \geqslant a\big) = \mathrm{E}\big(\mathbb{1}_{\{|X|\geqslant a\}}a\big) \leqslant \mathrm{E}\big(|X|\big)$.

Deuxième question : Jusqu'où les valeurs d'une variable aléatoire réelle peuvent-elles être éloignées de son espérance ?

Théorème (Inégalité de Bienaymé-Tchebychev) Soit X une variable aléatoire réelle. Pour tout $a > 0$:
$$P\Big(\big|X - \mathrm{E}(X)\big| \geqslant a\Big) \leqslant \frac{\mathrm{V}(X)}{a^2}.$$

Démonstration D'après l'inégalité de Markov : $P\Big(\big(X-\mathrm{E}(X)\big)^2 \geqslant a^2\Big) \leqslant \dfrac{\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)}{a^2} = \dfrac{\mathrm{V}(X)}{a^2}$ et il nous suffit dès lors d'observer que $\Big\{\big(X-\mathrm{E}(X)\big)^2 \geqslant a^2\Big\} = \Big\{\big|X-\mathrm{E}(X)\big| \geqslant a\Big\}$. Pour justifier cette égalité d'événements, on peut se contenter d'observer que par stricte croissance de $x \longmapsto x^2$ sur $\mathbb{R}_+$:
$$\big(X-\mathrm{E}(X)\big)^2 \geqslant a^2 \quad\Longleftrightarrow\quad \big|X-\mathrm{E}(X)\big| \geqslant a,$$
mais quelques détails ne seront peut-être pas de trop. Pour tout $\omega \in \Omega$:
$$\omega \in \Big\{\big(X-\mathrm{E}(X)\big)^2 \geqslant a^2\Big\} \iff \big(X(\omega)-\mathrm{E}(X)\big)^2 \geqslant a^2 \iff \big|X(\omega)-\mathrm{E}(X)\big| \geqslant a \iff \omega \in \Big\{\big|X-\mathrm{E}(X)\big| \geqslant a\Big\}.$$
N'oublions pas que tout événement est une partie de Ω.

Exemple On dispose d'une pièce éventuellement truquée dont la probabilité d'obtention de pile est notée p. Pour connaître p, on lance cette pièce n fois et on note F la fréquence d'apparition de pile obtenue. On cherche une valeur de n à partir de laquelle la probabilité pour que F soit une approximation de p à 10^{-2} près est supérieure à 0,9 — i.e. à partir de laquelle $P\big(|F-p| \geqslant 10^{-2}\big) \geqslant 0,9$.

Démonstration En notant N le nombre de piles obtenus : $F = \frac{N}{n}$ avec $N \sim \mathscr{B}(n,p)$ par indépendance des lancers, donc $\mathrm{E}(N) = np$, puis $\mathrm{E}(F) = p$. Ce résultat nous incite à choisir F pour estimer p. D'après l'inégalité de Bienaymé-Tchebychev :

$$P\Big(|F-p| \geqslant 10^{-2}\Big) = P\Big(\big|N-\mathrm{E}(N)\big| \geqslant 10^{-2}n\Big) \leqslant \frac{\mathrm{V}(N)}{\left(10^{-2}n\right)^2} = \frac{np(1-p)}{10^{-4}n^2} = \frac{10^4}{n} \times p(1-p) \leqslant \frac{10^4}{n} \times \frac{1}{4} = \frac{2500}{n}.$$

Or : $P\Big(|F-p| < 10^{-2}\Big) \geqslant 0,9 \iff P\Big(|F-p| \geqslant 10^{-2}\Big) \leqslant 0,1$, donc il nous suffit de choisir n avec la condition $\frac{2500}{n} \leqslant 0,1$ pour garantir que F soit une approximation de p à 10^{-2} près. Conclusion : il faut quand même lancer 25 000 fois la pièce ! — ou bien trouver une meilleure majoration de $P\Big(|F-p| \geqslant 10^{-2}\Big)$, et il en existe, mais là nous débutons.

Le paragraphe qui suit est l'aboutissement ultime de nos pérégrinations probabilistes en MPSI. On s'intéresse à une grandeur d'espérance m et d'écart-type σ. On la mesure dans un premier temps à l'occasion d'une unique expérience de résultat noté X_1. D'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$: $P\Big(|X_1 - m| \geqslant \varepsilon\Big) \leqslant \frac{\sigma^2}{\varepsilon^2}$. La mesure X_1 est donc à distance supérieure ou égale à ε de m avec probabilité au plus $\frac{\sigma^2}{\varepsilon^2}$.

Si l'on souhaite estimer m, l'expérience nous enseigne qu'une seule mesure a peu de chances de nous donner un résultat satisfaisant, nous obtiendrons une meilleure estimation de m en augmentant le nombre de mesures. La théorie des probabilités confirme-t-elle cette intuition ? Quel résultat quantitatif pouvons-nous en tirer ? Nous effectuons à présent n mesures $X_1,\ldots,X_n$ de la grandeur étudiée. Formellement, $X_1,\ldots,X_n$ sont des variables aléatoires réelles définies sur un même espace probabilisé fini, **INDÉPENDANTES** et **DE MÊME LOI**, d'espérance m et d'écart-type σ. Posons $S_n = \sum_{k=1}^{n} X_k$. Par linéarité de l'espérance : $\mathrm{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(X_k) = \frac{1}{n}\sum_{k=1}^{n} m = m$ et par indépendance : $\mathrm{V}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathrm{V}(X_k) = \frac{1}{n^2}\sum_{k=1}^{n}\sigma^2 = \frac{\sigma^2}{n}$, donc $\sigma\left(\frac{S_n}{n}\right) = \frac{\sigma}{\sqrt{n}}$. En résumé :

> La moyenne de n variables aléatoires **INDÉPENDANTES** et **DE MÊME LOI**, d'espérance m et d'écart-type σ, admet toujours m pour espérance, **MAIS** $\frac{\sigma}{\sqrt{n}}$ pour écart-type.

Avec n mesures, nous avons ainsi gagné un facteur $\frac{1}{\sqrt{n}}$ en écart-type — et $\frac{1}{n}$ en variance. Intuitivement, cela veut dire que chaque nouvelle mesure nous rapproche de la certitude que la moyenne empirique $\frac{S_n}{n}$ est proche de l'espérance m. Formellement, l'inégalité de Bienaymé-Tchebychev énonce cette fois que pour tout $\varepsilon > 0$:

$$P\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

En particulier, cette probabilité tend vers 0 par encadrement lorsque n tend vers $+\infty$, en accord avec notre intuition. Toute limite de ce genre est appelée une *loi des grands nombres*. Il en existe plusieurs, dont celle présentée ici qui est la plus faible de toutes, mais qui est facile à établir. Les lois des grands nombres sont fondamentales car elles justifient l'interprétation *fréquentiste* que nous avons des probabilités. Qu'avons-nous en tête quand nous disons que la probabilité d'apparition de chaque face d'un dé équilibré à 6 faces vaut $\frac{1}{6}$? Nous signifions par là une vérité sur le monde qui n'a rien à voir avec le fait que de tels dés existent et qu'on peut les lancer un nombre indéfini de fois. Cette probabilité vaudrait $\frac{1}{6}$ même si l'humanité s'interdisait rigoureusement tout lancer de dé. Cela dit, le fait est que nous avons une autre notion de probabilité en tête simultanément, nous avons aussi l'habitude de constater que lorsqu'on lance un dé de nombreuses fois, la fréquence d'apparition de chaque face tend vers $\frac{1}{6}$ à mesure que ce nombre augmente. Les lois des grands nombres réconcilient les deux points de vue, la fréquence d'un événement lorsqu'on répète indéfiniment une expérience aléatoire tend vers la probabilité a priori — i.e. hors de toute expérience — de cet événement.