– Chapitre 20 : Espérance et variance. –

Toutes les variables aléatoires sont supposées discrètes et définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathrm{P})$.
On notera : $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

I. Espérance d'une variable aléatoire réelle ou complexe.

I.1. Définition.

Définition 1. Soit X une variable aléatoire discrète, à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$.
On appelle *espérance* de X, et l'on note $E(X)$, la somme de la famille $\big(x\mathrm{P}(X=x)\big)_{x\in X(\Omega)}$:

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

Remarque 1. Si $\mathrm{P}(X=+\infty) > 0$ alors, $E(X) = +\infty$.
Si $\mathrm{P}(X=+\infty) = 0$, la définition précédente utilise la convention habituelle : $(+\infty)\times 0 = 0$. Ainsi, dans ce cas, on a : $E(X) = E(X\mathbb{1}_{\{X<+\infty\}})$.

Remarque 2. Si X est une variable à valeurs dans $\mathbb{N}$, alors :

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

⚠ La série de terme général $n\mathrm{P}(X=n)$ n'est pas nécessairement convergente.

Définition 2. Soit X une variable aléatoire discrète, à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.
On dit que X *est d'espérance finie* si la famille $\big(x\mathrm{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable. Dans ce cas, on appelle *espérance* de X, et l'on note $E(X)$, la somme de la famille $\big(x\mathrm{P}(X=x)\big)_{x\in X(\Omega)}$:

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

L'espérance est aussi appelée *moment d'ordre* 1.

⚠ Bien noter que si X est à valeurs positive, alors elle possède toujours une espérance qui peut être finie ou non. Mais dans le cas où X est à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$, on ne parle d'espérance que dans le cas où la famille $\big(x\mathrm{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable.

Remarque 3.

- Dans le cas particulier où X est une variable aléatoire finie (i.e. $X(\Omega)$ est un ensemble fini), alors X est d'espérance finie et la définition de son espérance est alors la même que celle donnée en mpsi.
- Dans le cas où X est une variable à valeurs dans $\mathbb{N}$, alors X est d'espérance finie si, et seulement si, la série de terme général $n\mathrm{P}(X=n)$ est convergente.

Exemple 1.

1. Pour tout événement A, $E(\mathbb{1}_A) =$

2. Si X est une variable aléatoire presque sûrement égale à une constante $a \in \mathbb{K}$, alors $E(X) =$

Remarque 4. En fait, l'espérance de X ne dépend que de la loi que X suit : on aurait pu parler d'espérance d'une loi, mais l'usage veut qu'on parle d'espérance d'une variable aléatoire.

Définition 3. On note $L^1(\Omega, \mathbb{K})$ l'ensemble des variables aléatoires discrètes d'espérance finie.

Proposition 1. Si X est une variable à valeurs dans $\mathbb{N} \cup \{+\infty\}$, alors :

$$E(X) = \sum_{n=0}^{+\infty} \mathrm{P}(X > n) = \sum_{n=1}^{+\infty} \mathrm{P}(X \geqslant n).$$

Démonstration.

$\square$

I.2. ESPÉRANCE DES LOIS USUELLES.

Proposition 2. Si $X \sim \mathcal{U}(\llbracket a, b \rrbracket)$, alors : $E(X) = \dfrac{a+b}{2}$.

En particulier, si $X \sim \mathcal{U}(\llbracket 1, n \rrbracket)$, alors : $E(X) = \dfrac{n+1}{2}$.

Démonstration.

$\square$

Proposition 3. Si $X \sim \mathcal{B}(p)$, alors : $E(X) = p$.

Démonstration.

□

Proposition 4. Si $X \sim \mathcal{B}(n, p)$, alors : $E(X) = np$.

Démonstration.

□

Proposition 5. Si $X \sim \mathcal{G}(p)$, avec $p \in]0,1[$ alors : $E(X) = \dfrac{1}{p}$.

Démonstration.

□

Proposition 6. Si $X \sim \mathcal{P}(\lambda)$, avec $\lambda > 0$ alors : $E(X) = \lambda$.

Démonstration.

□

I.3. Propriétés de l'espérance.

Théorème 1. Formule de transfert.
Soit X une variable aléatoire discrète à valeurs dans un ensemble quelconque et f est une application de $X(\Omega)$ dans $\mathbb{C}$.
Alors la variable aléatoire $f(X)$ est d'espérance finie si, et seulement si, la famille $\big(f(x)\mathrm{P}(X=x)\big)_{x\in X(\Omega)}$ est sommable. Dans ce cas :

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

Démonstration.

□

L'intérêt de la formule de transfert est de permettre de calculer $E\big(f(X)\big)$ à partir de la loi de X sans avoir besoin de déterminer celle de $f(X)$.

Remarque importante. La formule de transfert s'applique pour des variables aléatoires à valeurs dans un ensemble quelconque. Elle s'applique en particulier à une variable à valeurs dans un produit cartésien i.e. à un couple de variables aléatoires : $Z=(X,Y)$.
La formule de transfert appliquée au couple $Z=(X,Y)$ s'écrit de la forme :

$$E\big(f(X,Y)\big) =$$

où $Z(\Omega)$ est une partie de $X(\Omega)\times Y(\Omega)$.
Si $(x,y)\in X(\Omega)\times Y(\Omega)$ mais que $(x,y)\notin Z(\Omega)$ on a :
On peut donc écrire :

Exercice 1. Soit X et Y deux variables aléatoires indépendantes de loi $\mathcal{U}(\llbracket 1, 6 \rrbracket)$.
Déterminer l'espérance de $S = \mathrm{Sup}(X, Y)$.

Proposition 7. Inégalité triangulaire.
Pour tout $X \in L^1(\Omega, \mathbb{K})$:
$$\big|E(X)\big| \leqslant E\big(|X|\big).$$

Démonstration.

$\square$

Proposition 8. Soit X et Y deux variables aléatoires discrètes, respectivement complexe et réelle, telles que $|X| \leqslant Y$. Si Y est d'espérance finie, alors X est d'espérance finie.

Démonstration.
C'est une conséquence immédiate d'un résultat de comparaison de familles sommables.

$\square$

Théorème 2. Linéarité de l'espérance.
L'ensemble $L^1(\Omega, \mathbb{K})$ des variables aléatoires discrètes d'espérance finie est un $\mathbb{K}$-espace vectoriel, et l'espérance est une forme linéaire sur $L^1(\Omega, \mathbb{K})$.
Autrement-dit, si $(X, Y) \in \big(L^1(\Omega, \mathbb{K})\big)^2$, et si $\alpha \in \mathbb{K}$, alors :
$$E(X + Y) = E(X) + E(Y) \quad \text{et} \quad E(\alpha X) = \alpha E(X).$$

Démonstration.

□

Remarque 6. La linéarité de l'espérance donne une autre preuve de l'espérance d'une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, p)$.

Corollaire 1. Si X est une variable aléatoire d'espérance finie et si $(a, b) \in \mathbb{C}^2$, alors :

$$E(aX + b) = aE(X) + b.$$

Démonstration. Par linéarité de l'espérance : □

Définition 4. Une variable aléatoire réelle d'espérance nulle est dite *centrée*.

Proposition 9. Si X est une variable aléatoire d'espérance finie alors la variable aléatoire $(X - E(X))$ est une variable aléatoire centrée appelée la variable aléatoire centrée associée à X.

Démonstration. Il suffit d'appliquer le corollaire précédent avec : □

Proposition 10. Soit X une variable aléatoire réelle positive. Alors :

1. $E(X) \geqslant 0$,

2. $E(X) = 0$ si, et seulement si, $\mathrm{P}(X = 0) = 1$. On dit alors que X est *presque sûrement nulle.*

Démonstration.

□

Proposition 11. Croissance de l'espérance.
Soit X et Y deux variables aléatoires réelles d'espérance finie telles que $X \leqslant Y$. Alors $E(X) \leqslant E(Y)$.

Démonstration.

□

Remarque 7. Espérance d'un produit XY de deux variables aléatoires discrètes.

Théorème 3. Si X et Y sont deux variables aléatoires discrètes indépendantes et d'espérance finie, alors XY est d'espérance finie et :

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

Démonstration.

$\square$

⚠ Cette propriété se généralise au cas de n variables aléatoires mutuellement indépendantes.

II. Variance d'une variable aléatoire réelle.

II.1. Définition.

Dans cette partie, toutes les variables aléatoires sont supposées discrètes et à valeurs réelles.

Définition 5. On dit qu'un variable aléatoire X *admet un moment d'ordre* 2 si la variable aléatoire X^2 est d'espérance finie.

Définition 6. On note $L^2(\Omega, \mathbb{R})$ l'ensemble des variables aléatoires réelles discrètes admettant un moment d'ordre 2.

Remarque 8. On a donc : $X \in L^2(\Omega, \mathbb{R}) \Leftrightarrow X^2 \in L^1(\Omega, \mathbb{R})$.

⚠ Les variables aléatoires presque sûrement constantes admettent un moment d'ordre 2.

Théorème 4. Inégalité de Cauchy-Schwarz.
Si $(X, Y) \in L^2(\Omega, \mathbb{R})$ alors $XY \in L^1(\Omega, \mathbb{R})$ et :

$$E(XY)^2 \leqslant E(X^2)E(Y^2).$$

Il y a égalité si, et seulement si, X et Y sont proportionnelles presque sûrement.

Démonstration. L'inégalité et le cas d'égalité se démontrent comme l'inégalité de Cauchy-Schwarz dans un espace préhilbertien. Soit $(X, Y) \in L^2(\Omega, \mathbb{R})$. Montrons que $XY \in L^1(\Omega, \mathbb{R})$.

$\square$

Proposition 12. Si X admet un moment d'ordre 2, alors X est d'espérance finie.
Autrement-dit, $L^2(\Omega, \mathbb{R}) \subset L^1(\Omega, \mathbb{R})$.

Proposition 13. L'ensemble $L^2(\Omega, \mathbb{R})$ est un $\mathbb{R}$-espace vectoriel.

Démonstration.

$\square$

Définition 7. Si $X \in L^2(\Omega, \mathbb{R})$, on appelle *variance* de X le réel noté $V(X)$ et défini par :

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

On appelle *écart type* de X le réel noté $\sigma(X)$ et défini par : $\sigma(X) = \sqrt{V(X)}$.

⚠ Cette définition a bien un sens puisque :

Remarque 9. La variance, comme l'espérance ne dépend que de la loi de la variable aléatoire.

Remarque 10. La variance d'une variable aléatoire réelle est l'espérance du carré des écarts à l'espérance. La variance permet donc de mesurer la dispersion de X autour de $E(X)$. On dit que $V(X)$ est un *paramètre de dispersion*, alors que $E(X)$ est un *paramètre de position.*

II.2. Propriétés de la variance.

D'après la formule de transfert avec $f : x \mapsto (x - E(X))^2$:

Proposition 14. Si $X \in L^2(\Omega, \mathbb{R})$ alors :

$$V(X) = \sum_{x \in X(\Omega)} (x - E(X))^2 P(X = x).$$

Mais c'est surtout l'identité remarquable suivante, qui sera utilisée :

Théorème 5. Formule de Kœnig-Huygens.
Si $X \in L^2(\Omega, \mathbb{R})$ alors : $V(X) = E(X^2) - E(X)^2$.

Démonstration.

$\square$

Corollaire 2. Si $X \in L^2(\Omega, \mathbb{R})$ et si $(a, b) \in \mathbb{R}^2$, alors :

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

Démonstration.

□

Proposition 15. Soit $X \in L^2(\Omega, \mathbb{R})$. Alors :

1. $V(X) \geqslant 0$,

2. $V(X) = 0$ si, et seulement si, il existe un réel m tel que $P(X = m) = 1$. On dit alors que X est *presque sûrement constante.*

Démonstration.

□

II.3. Variance des lois usuelles.

Proposition 16. Si $X \sim \mathcal{U}\big([\![1, n]\!]\big)$, alors : $V(X) = \dfrac{n^2 - 1}{12}$.

Démonstration.

□

Proposition 17. Si $X \sim \mathcal{B}(p)$, alors : $V(X) = p(1 - p) = pq$.

Démonstration.

□

Proposition 18. Si $X \sim \mathcal{B}(n,p)$, alors : $V(X) = np(1-p) = npq$.

Démonstration.

$\square$

Proposition 19. Si $X \sim \mathcal{G}(p)$, avec $p \in]0,1[$ alors : $V(X) = \dfrac{1-p}{p^2} = \dfrac{q}{p^2}$.

Démonstration.

$\square$

Proposition 20. Si $X \sim \mathcal{P}(\lambda)$, avec $\lambda > 0$ alors : $V(X) = \lambda$.

Démonstration.

$\square$

II.4. Covariance de deux variables aléatoires réelles.

Définition 8. Soit $(X,Y) \in \left(L^2(\Omega,\mathbb{R})\right)^2$. On appelle *covariance* de X et Y, ou *covariance* de (X,Y), le réel noté $\mathrm{Cov}(X,Y)$ et défini par :

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

⚠ Cette définition a bien un sens puisque :

Remarque 11. On peut remarquer que l'application Cov est une forme bilinéaire symétrique positive ! Elle est positive car : $\mathrm{Cov}(X,X) =$

Théorème 6. Formule de Kœnig-Huygens.
Si $(X,Y) \in \left(L^2(\Omega,\mathbb{R})\right)^2$, alors : $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

Démonstration.
$\mathrm{Cov}(X,Y) =$

$=$

$=$ □

Remarque 12. Cette formule permet de retrouver la formule de Kœnig-Huygens :

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

Théorème 7. Si $(X,Y) \in \left(L^2(\Omega,\mathbb{R})\right)^2$ alors : $V(X+Y) = V(X) + V(Y) + 2\mathrm{Cov}(X,Y)$.

Démonstration. L'application Cov étant une forme bilinéaire symétrique :
$V(X+Y) = \mathrm{Cov}(X+Y, X+Y) =$ □

Corollaire 3. Si $(X,Y) \in \left(L^2(\Omega,\mathbb{R})\right)^2$ et si X et Y sont indépendantes, alors :

$$\mathrm{Cov}(X,Y) = 0 \qquad \text{et} \qquad V(X+Y) = V(X) + V(Y).$$

Démonstration. □

Définition 9. Lorsque $\mathrm{Cov}(X,Y) = 0$, on dit que les variables aléatoires X et Y sont *non corrélées.*

⚠ On a donc : (X et Y sont indépendantes) $\Rightarrow$ (X et Y sont non corrélées). Mais la réciproque est fausse.

L'application Cov étant une forme bilinéaire symétrique :

Théorème 8. Si $(X_1, \dots, X_n) \in \left(L^2(\Omega, \mathbb{R})\right)^n$ alors :

$$V(X_1 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j).$$

Corollaire 4. Si de plus, les variables aléatoires sont 2 à 2 indépendantes alors :

$$V(X_1 + \cdots + X_n) = \sum_{k=1}^{n} V(X_k).$$

Remarque 13. On en déduit une autre preuve de la variance d'une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, p)$.

III. Inégalités probabilistes et loi faible des grands nombres.

Théorème 9. Inégalité de Markov.
Pour toute variable aléatoire réelle positive X, on a :

$$\forall a > 0,\ \mathrm{P}(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

Nous allons voir deux manières d'écrire la démonstration de l'inégalité de Markov.

Démonstration. 1.

$\square$

Démonstration. 2.

$\square$

Théorème 10. Inégalité de Bienaymé-Tchebychev.
Pour toute variable aléatoire réelle $X \in L^2(\Omega, \mathbb{R})$ on a :

$$\forall \varepsilon > 0,\ \mathrm{P}\big(|X - E(X)| \geqslant \varepsilon\big) \leqslant \frac{V(X)}{\varepsilon^2}.$$

Démonstration.

$\square$

L'inégalité de Bienaymé-Tchebychev est qualifiée d'*inégalité de concentration*, car elle majore la probabilité qu'une variable aléatoire dévie de son espérance.

Théorème 11. Loi faible des grands nombres.
Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires discrètes i.i.d. sur un même espace probabilisé.
On suppose que $X_1 \in L^2(\Omega, \mathbb{R})$ et on note $m = E(X_1)$.
Pour $n \in \mathbb{N}^*$, on définit : $S_n = X_1 + \cdots + X_n$. On a alors :

$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} \mathrm{P}\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) = 0.$$

Démonstration.

$\square$

⚠ La variable aléatoire $\dfrac{S_n}{n}$ désigne la moyenne empirique. La loi faible des grands nombres signifie, qu'en un certain sens, la suite $\left(\dfrac{S_n}{n}\right)_{n\in\mathbb{N}^*}$ converge vers m.

Exemple 2. On réalise une suite d'épreuves aléatoires identiques et indépendantes (par exemple un lancer de dé). On s'intéresse à la réalisation ou non d'un certain événement de probabilité p, et on note A_n l'événement correspondant lors de la n-ème épreuve (par exemple A_n serait de réaliser un 6 lors du n-ème lancer et $p = \frac{1}{6}$.).
On obtient alors une suite $(A_n)_{n\in\mathbb{N}^*}$ d'événements mutuellement indépendants, et une suite $(\mathbb{1}_{A_n})_{n\in\mathbb{N}^*}$ de variable aléatoire i.i.d. telles que pour tout $n \in \mathbb{N}^*$, $\mathbb{1}_{A_n} \sim \mathcal{B}(p)$.
Comme $E(\mathbb{1}_{A_1}) =$ on obtient :

Ici, la variable aléatoire $\dfrac{S_n}{n}$ désigne la *fréquence empirique* de réalisation de notre événement. La loi faible des grands nombres signifie, qu'en un certain sens, la suite $\left(\dfrac{S_n}{n}\right)_{n\in\mathbb{N}^*}$ converge vers p.
Ceci peut-être utile lorsque l'on ne connait pas la valeur de p. La variable aléatoire $\dfrac{S_n}{n}$ est alors appelée un *estimateur* de p.

La figure ci-dessus (obtenue en Python) illustre la convergence de la fréquence empirique de l'événement "obtenir un 6" vers la probabilité $p = \frac{1}{6}$.

Exercice 2. Écrire une fonction Python simulant la répétition d'un lancer de dé équilibré et permettant d'obtenir la figure ci-dessus ; n et p seront des paramètres de la fonction.

IV. Fonctions génératrices.

IV.1. Définition.

Définition 10. Pour toute variable aléatoire X à valeurs dans $\mathbb{N}$, on appelle *fonction génératrice* de X la fonction G_X définie sur une partie de $\mathbb{R}$ par :

$$G_X(t) = E(t^X) = \sum_{k=0}^{+\infty} \mathrm{P}(X = k)\, t^k.$$

⚠ La deuxième égalité est une conséquence de la formule de transfert.

⚠ L'ensemble de définition de G_X est donc l'intervalle de convergence de la série entière $\sum \mathrm{P}(X = k)\, t^k$. Comme pour n'importe quelle série entière, il est donc intéressant de connaître son rayon de convergence.

IV.2. Propriétés.

Proposition 21. Soit X variable aléatoire à valeurs dans $\mathbb{N}$.

1. La série entière définissant G_X est de rayon de convergence $R \geqslant 1$.

2. La convergence est normale sur $[-1, 1]$.

3. G_X est définie et continue (au moins) sur $[-1, 1]$.

⚠ D'après les résultats sur les séries entières, G_X est même de classe $\mathcal{C}^\infty$ sur $]-R, R[$.

Démonstration.

□

Exercice 3. Que peut-on dire dans le cas où X est une variable finie ?

Proposition 22. La fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$ caractérise la loi.
Autrement-dit, pour toutes variables aléatoires X et Y à valeurs dans $\mathbb{N}$, X et Y ont même loi si, et seulement si, pour tout $t \in [-1,1]$, $G_X(t) = G_Y(t)$.

Démonstration. □

Théorème 12. Une variable aléatoire X à valeurs dans $\mathbb{N}$ est d'espérance finie si, et seulement si, G_X est dérivable en 1. Dans ce cas, on a :

$$E(X) = G'_X(1).$$

Démonstration. □

Théorème 13. Une variable aléatoire X à valeurs dans $\mathbb{N}$ appartient à $L^2(\Omega, \mathbb{R})$ si, et seulement si, G_X est deux fois dérivable en 1. Dans ce cas, on a :

$$G''_X(1) = E\big(X(X-1)\big).$$

Démonstration. Admis. □

⚠ Dans ce cas, on a : $V(X) =$

IV.3. Fonctions génératrices des lois usuelles.

IV.4. Fonction génératrice d'une somme de variables aléatoires indépendantes.

Les fonctions génératrices peuvent être particulièrement efficace pour déterminer la loi de la somme de variables aléatoires indépendantes :

Théorème 14. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes à valeurs dans $\mathbb{N}$.
On pose : $S_n = \displaystyle\sum_{k=1}^{n} X_k$. Soit un réel t tel que G_{X_k} soit défini en t pour tout $k \in [\![1, n]\!]$.
Alors, G_{S_n} est défini en t et :

$$G_{S_n}(t) = \prod_{k=1}^{n} G_{X_k}(t).$$

⚠ Ce théorème s'applique en particulier pour tout $t \in [-1, 1]$.

Démonstration.

□

Exercice 4. Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que $X_k \sim \mathcal{B}(n_k, p)$ pour tout $k \in [\![1, n]\!]$. Déterminer la loi de $S_n = \sum_{k=1}^{n} X_k$ en utilisant les fonctions génératrices.

Exercice 5. Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes telles que $X_k \sim \mathcal{P}(\lambda_k)$ pour tout $k \in [\![1, n]\!]$. Déterminer la loi de $S_n = \sum_{k=1}^{n} X_k$ en utilisant les fonctions génératrices.