Chapitre 36

Probabilités 3 – Espérance, variance

Plan du chapitre

1 Espérance **1**
- 1.1 Définition 1
- 1.2 Espérances des lois usuelles 2
- 1.3 Propriétés de l'espérance 3
- 1.4 Formule de transfert 5
- 1.5 Espérance d'un produit de v.a. indépendantes 6

2 Variance et covariance **6**
- 2.1 Définition 6
- 2.2 Variance des lois usuelles 7
- 2.3 Propriétés de la variance 8
- 2.4 Écart-type 8

3 Covariance **9**
- 3.1 Définition 9
- 3.2 Propriétés de la covariance 10
- 3.3 Variables aléatoires décorrélées 11

4 Inégalités probabilistes **12**

5 Compléments **15**

Hypothèse

Dans tout ce chapitre, $(\Omega, \mathbb{P})$ désigne un espace probabilisé fini (i.e. Ω désigne un univers fini et $\mathbb{P}$ est une probabilité définie sur Ω).
E, F sont des ensembles quelconques.

1 Espérance

1.1 Définition

On rappelle qu'une v.a. est dite réelle (resp. complexe) si elle est à valeurs dans un ensemble $E \subset \mathbb{R}$ (resp. $E \subset \mathbb{C}$).

Définition 36.1

Soit $X : \Omega \to E$ une v.a. réelle (resp. complexe). On définit l'espérance de X comme étant le réel (resp. le complexe) noté

$$\mathbb{E}(X) := \sum_{x \in E} x\mathbb{P}(X = x)$$

Remarque.

- L'espérance d'une variable aléatoire ne dépend ainsi que de sa loi. Si $X \sim Y$, alors $\mathbb{E}[X] = \mathbb{E}[Y]$.
- Bien que E soit infini, la somme ci-dessus est finie : on a plus précisément

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

  où $X(\Omega)$ est fini car Ω l'est.

L'espérance d'une v.a. représente sa valeur moyenne, comme l'illustrent les exemples suivants :

Exemple 1. On lance un dé à six faces. On note X la v.a. qui correspond au chiffre obtenu. On a donc

$$\forall i \in [\![1,6]\!] \qquad \mathbb{P}(X = i) = \frac{1}{6}$$

Alors l'espérance de X vaut :

1.2 Espérances des lois usuelles

Propriété 36.2 (Espérances de lois usuelles)

Soit $a, b \in \mathbb{Z}$ tels que $a \leq b$. Soit $n \in \mathbb{N}^*$ et $p \in [0,1]$. Soit enfin X une v.a.

1. Si $X \sim \mathcal{U}([\![a,b]\!])$, alors $\mathbb{E}(X) = \dfrac{a+b}{2}$.
2. Si $X \sim \mathcal{B}(p)$, alors $\mathbb{E}(X) = p$.
3. Si $X \sim \mathcal{B}(n,p)$, alors $\mathbb{E}(X) = np$.

Démonstration.

1. Pour la loi uniforme, le calcul est similaire à l'exemple précédent.
2.
3. Conséquence de l'assertion précédente et de l'Exemple 3 vu plus loin.

□

Définition 36.3

On dit qu'une v.a. $X : \Omega \to E$ est <u>(presque sûrement) constante</u> s'il existe $a \in E$ telle que $\mathbb{P}(X = a) = 1$. On peut alors écrire (par abus) $X = a$.

Écrire $X = a$ est un double abus. D'une part car X et a ne représentent pas le même objet : $X \in E^{\Omega}$ tandis que $a \in E$. Ensuite, même si on écrivait “$X \equiv a$”, ce serait encore un abus. Il est possible d'avoir $\mathbb{P}(X = a) = 1$ sans que X soit l'application constante égale à a (mais dans ce cas la probabilité $\mathbb{P}$ est un peu pathologique et cela signifie sans doute que la probabilité $\mathbb{P}$ ou bien l'univers Ω ont été mal choisis).

Exemple 2. Si $X = a$ est la v.a. constante égale à a (on peut alors noter $X = a$), alors

$$\mathbb{E}(X) = \mathbb{E}(a) = a$$

1.3 Propriétés de l'espérance

Le lemme qui suit n'a que peu d'intérêt en pratique. Il est néanmoins utile pour les démonstrations des propriétés qui suivent. La preuve de ce lemme se trouve dans la partie Compléments

Lemme 36.4

Soit $X : \Omega \to E$ une v.a. Alors

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

Propriété 36.5 (Linéarité)

Soit X, Y deux v.a. complexes et $\lambda \in \mathbb{C}$.

1. Linéarité :
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
$$\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$$
2. Positivité : si X est à valeurs dans $\mathbb{R}_+$ alors
$$\mathbb{E}(X) \geq 0$$
3. Croissance : si X, Y sont des v.a.r. telles que $X \leq Y$, alors
$$\mathbb{E}(X) \leq \mathbb{E}(Y)$$
4. Inégalité triangulaire :
$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$$

Démonstration. On ne montre que la première et la dernière assertion. On a

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

De plus, par l'inégalité triangulaire "classique" :

$$\begin{aligned}
|\mathbb{E}(X)| &= \left|\sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\{\omega\})\right| \\
&\leq \sum_{\omega\in\Omega} |X(\omega)\,\mathbb{P}(\{\omega\})| \\
&= \sum_{\omega\in\Omega} |X(\omega)|\,\mathbb{P}(\{\omega\}) \qquad \text{car } \mathbb{P}(\{\omega\}) \geq 0 \\
&= \mathbb{E}(|X|)
\end{aligned}$$

□

Remarque. Ces propriétés ressemblent énormément à celles des intégrales. C'est tout à fait normal, l'espérance est en fait une "intégrale déguisée". Une autre propriété connue des intégrales est la suivante : si $f \in C^0([a,b],\mathbb{R}_+)$ est $\boxed{\text{positive}}$, alors

$$\int_a^b f = 0 \qquad \Longrightarrow \qquad f \equiv 0$$

L'adaptation aux v.a. donne le résultat suivant : si X est une v.a. à valeurs dans $\boxed{\mathbb{R}_+}$, alors

$$\mathbb{E}(X) = 0 \quad \Longrightarrow \quad \mathbb{P}(X=0) = 1$$

Si $\mathbb{P}(X=0)=1$, on dit alors que X est <u>presque sûrement nulle</u> (ou encore presque sûrement égale à 0).

A noter : $\mathbb{E}(X)$ est une constante qui ne dépend que de la loi de X. Ainsi, par linéarité,

$$\mathbb{E}\left(\mathbb{E}(X)\right) = \mathbb{E}(X)\times\mathbb{E}(1) = \mathbb{E}(X)\times 1 = \mathbb{E}(X)$$

Exemple 3. Si $X \sim \mathcal{B}(n,p)$, alors $\mathbb{E}(X) = np$.

Définition 36.6

Une v.a. X telle que $\mathbb{E}(X)=0$ est dite une <u>v.a. centrée</u>.

Exemple 4. Si X est une v.a. quelconque, $X-\mathbb{E}(X)$ est ainsi une v.a. centrée.

1.4 Formule de transfert

Théorème 36.7 (Formule de transfert)

Soit $X : \Omega \to E$ une v.a. et $f : E \to \mathbb{C}$ une application. Alors

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

Remarque. Cette formule est incontournable. Elle permet de calculer l'espérance de $f(X)$ uniquement à partir de la loi de X (et non celle de $f(X)$).

Exemple 5. On suppose que $X \sim \mathcal{B}(n, p)$ avec $p > 0$. Déterminer la loi de $Y = \dfrac{1}{1+X}$.

1.5 Espérance d'un produit de v.a. indépendantes

Propriété 36.8

Soit X, Y deux v.a. (réelles ou) complexes **indépendantes**. Alors

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

Plus généralement, si $X_1, \cdots, X_n$ sont n v.a. **indépendantes**, alors

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n)$$

Exemple 6. On considère $X_1, \cdots, X_n$ des v.a. indépendantes suivant la même loi de Bernoulli $p \in [0,1]$. Alors

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

2 Variance et covariance

2.1 Définition

Définition 36.9 (Variance)

Soit X une v.a.r. définie sur Ω. On appelle <u>variance</u> de X le réel positif :

$$\mathbb{V}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2 \right) \qquad \geq 0$$

Autrement dit, on considère la v.a. centrée $X - \mathbb{E}(X)$, et on calcule l'espérance de son carré. Plus $\mathbb{V}(X)$ est grand, plus X prend des valeurs dispersées autour de sa moyenne.

Propriété 36.10

Soit X une v.a.r. définie sur Ω. Alors

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

C'est en général avec cette formule qu'on calcule la variance d'une v.a.r.

Démonstration.

□

Exemple 7. Si X est une v.a.r. constante égale à $a \in \mathbb{R}$, alors $\mathbb{V}(X) = 0$.

Voici des exemples de v.a.r avec une forte variance et une faible variance :

2.2 Variance des lois usuelles

Propriété 36.11

Si $X \sim \mathcal{B}(p)$, alors $\mathbb{V}(X) = p(1-p)$.

Démonstration.

□

Propriété 36.12

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

Démonstration. Vue ultérieurement □

2.3 Propriétés de la variance

Propriété 36.13

Soit $a, b \in \mathbb{R}$ et X une v.a.r. définie sur Ω. Alors

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

Démonstration. On sait que $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$, de sorte que

$$\begin{aligned}
\mathbb{V}(aX+b) &= \mathbb{E}\left((aX+b-\mathbb{E}(aX+b))^2\right) \\
&= \mathbb{E}\left((aX+b-a\mathbb{E}(X)-b)^2\right) \\
&= \mathbb{E}\left(a^2(X-\mathbb{E}(X))^2\right) \\
&= a^2\mathbb{E}\left((X-\mathbb{E}(X))^2\right) \\
&= a^2\mathbb{V}(X)
\end{aligned}$$

□

Attention cependant : en général, $\mathbb{V}(X+Y) \neq \mathbb{V}(X)+\mathbb{V}(Y)$! Pour traiter la variance d'une somme, on a besoin d'une autre notion : la covariance (cf partie suivante).

2.4 Écart-type

Définition 36.14 (Écart-type)

Soit X une v.a.r. définie sur Ω. On appelle écart-type de X le réel positif :

$$\sigma(X) := \sqrt{\mathbb{V}(X)}$$

Définition 36.15

On dit qu'une v.a.r. X est centrée réduite si $\mathbb{E}(X) = 0$ et $\mathbb{V}(X) = 1$ (ou encore $\sigma(X) = 1$).

Propriété 36.16

Si X est une v.a.r. telle que $\sigma(X) > 0$, alors

$$Y = \frac{X-\mathbb{E}(X)}{\sigma(X)}$$

est une v.a.r. centrée réduite.

Démonstration.

□

3 Covariance

3.1 Définition

Définition 36.17

Soit X, Y deux v.a.r. définies sur Ω. On appelle covariance de X, Y le réel

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

En d'autres termes, on considère les deux v.a.r. centrées $X-\mathbb{E}(X)$ et $Y-\mathbb{E}(Y)$ et on calcule l'espérance de leur produit.

- Lorsque $\mathrm{Cov}(X,Y) > 0$, cela signifie que ces v.a. centrées prennent, lorsque ω parcourt Ω, des valeurs de même signe : elles sont alors dites positivement corrélées. Ce phénomène est d'autant plus marqué que $\mathrm{Cov}(X,Y)$ est grand.
- À l'inverse, si $\mathrm{Cov}(X,Y) < 0$, cela signifie que ces v.a. centrées prennent, lorsque ω parcourt Ω, des valeurs de signes opposés : elles sont alors dites négativement corrélées. Ce phénomène est d'autant plus marqué que $\mathrm{Cov}(X,Y)$ est petit.

Par exemple, la température extérieure et la consommation de glaces sont des v.a. positivement corrélées. À l'inverse, la température extérieure et la consommation de chauffage sont des v.a. négativement corrélées.

Propriété 36.18

Soit X, Y deux v.a.r. définies sur Ω. Alors :

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

Démonstration.

□

Exemple 8. Si $X = a$ est une v.a. presque sûrement constante, alors pour toute v.a.r. Y, on a $\mathrm{Cov}(X,Y) = 0$.

Exemple 9. Soit $p \in]0,1[$ et X une v.a. telle que $X \sim \mathcal{B}(p)$. On pose $Y = 1 - X$. Calculer $\mathrm{Cov}(X,Y)$.

3.2 Propriétés de la covariance

Propriété 36.19

Soit X, X', Y, Y' des v.a.r. définies sur Ω.

1. $\mathrm{Cov}(X,X) = \mathbb{V}(X) \geq 0$
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
3. Cov est une forme bilinéaire : pour tout $\lambda \in \mathbb{R}$,

$$\mathrm{Cov}(\lambda X + X', Y) = \lambda \mathrm{Cov}(X,Y) + \mathrm{Cov}(X',Y)$$

$$\mathrm{Cov}(X, \lambda Y + Y') = \lambda \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Y')$$

Plus précisément, Cov est une forme bilinéaire de $\mathcal{E} \times \mathcal{E}$ dans $\mathbb{R}$, où $\mathcal{E}$ est l'e.v. de v.a.r. définies sur $(\Omega, \mathbb{P})$.

Propriété 36.20 ("Identité remarquable")

Soit X, Y des v.a.r. définies sur Ω. Alors

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

Démonstration. On a

$$\begin{aligned}
\mathbb{V}(X+Y) &= \mathbb{E}\left((X+Y)^2\right) - \mathbb{E}(X+Y)^2 \\
&= \mathbb{E}\left(X^2 + Y^2 + 2XY\right) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
&= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

□

On peut généraliser la Propriété précédente à n variables :

Propriété 36.21

Pour toute famille $(X_1, \cdots, X_n)$ de v.a.r. définies sur Ω, on a

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

Exemple 10. Voir exemple 11 plus loin.

3.3 Variables aléatoires décorrélées

Définition 36.22

Deux v.a.r. X, Y définies sur Ω sont dites <u>décorrélées</u> si $\mathrm{Cov}(X, Y) = 0$.
Dans ce cas, on a

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

Propriété 36.23

Si X, Y sont des v.a.r. indépendantes, alors elles sont décorrélées.

Démonstration. Si X, Y sont indépendantes, alors

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

□

Exemple 11. Soit X une v.a.r. telle que $X \sim \mathcal{B}(n, p)$. Montrer que $\mathbb{V}(X) = np(1 - p)$.

Attention, la réciproque de la Proposition 36.23 est fausse.

Exemple 12. Soit X,Y deux v.a. telles que $X \sim \mathcal{U}(\{-1,0,1\})$ et $Y = \mathbf{1}_{\{X=0\}}$. Alors X,Y sont décorrélées, mais ne sont pas indépendantes.

4 Inégalités probabilistes

Propriété 36.24 (Inégalité de Markov)

Soit X une v.a. à valeurs dans $\mathbb{R}_+$. Alors

$$\forall a > 0 \qquad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Démonstration. Soit $a > 0$. On affirme que les v.a. $\mathbf{1}_{\{X \geq a\}}$ et $\dfrac{X}{a}$ vérifient :

$$\mathbf{1}_{\{X \geq a\}} \leq \frac{X}{a}$$

Pour le prouver, on raisonne par disjonction de cas :

- Si $\omega \in \Omega$ est tel que $X(\omega) < a$, alors cette inégalité devient $0 \leq \dfrac{X(\omega)}{a}$, qui est vraie car X est positive.
- Si $\omega \in \Omega$ est tel que $X(\omega) \geq a$, alors cette inégalité devient $1 \leq \dfrac{X(\omega)}{a}$, qui est vraie car $X(\omega) \geq a$.

Alors, en passant à l'espérance, on a

$$\mathbb{E}\left(\mathbf{1}_{\{X \geq a\}}\right) \leq \mathbb{E}\left(\frac{X}{a}\right)$$

donc

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

□

Propriété 36.25 (Inégalité de Bienaymé-Tchebychev)

Soit X une v.a.r. Alors

$$\forall \varepsilon > 0 \qquad \mathbb{P}\left(|X - \mathbb{E}(X)| \geq \varepsilon\right) \quad \leq \quad \frac{\mathbb{V}(X)}{\varepsilon^2}$$

En passant au complémentaire, on a donc

$$\mathbb{P}\left(|X - \mathbb{E}(X)| < \varepsilon\right) \quad \geq \quad 1 - \frac{\mathbb{V}(X)}{\varepsilon^2}$$

Démonstration. On pose la v.a. positive

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

ainsi que $a := \varepsilon^2 > 0$. On applique l'inégalité de Markov à Y en a : on a donc

$$\begin{aligned}
&\mathbb{P}(Y \geq a) \leq \frac{\mathbb{E}(Y)}{a} \\
\iff &\mathbb{P}\left((X - \mathbb{E}(X))^2 \geq \varepsilon^2\right) \leq \frac{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}{\varepsilon^2} \\
\iff &\mathbb{P}\left(\sqrt{(X - \mathbb{E}(X))^2} \geq \varepsilon\right) \leq \frac{\mathbb{V}(X)}{\varepsilon^2} \\
\iff &\mathbb{P}\left(|X - \mathbb{E}(X)| \geq \varepsilon\right) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}
\end{aligned}$$

□

Remarque. L'inégalité de Bienaymé-Tchebicheff est une *inégalité de concentration* : elle fournit une majoration de la probabilité que X n'appartienne pas à l'intervalle $[\mathbb{E}(X) - \varepsilon, \mathbb{E}(X) + \varepsilon]$, autrement dit que X s'écarte à plus de ε de son espérance.

Exemple 13. Soit $X_1, \cdots, X_n$ des v.a.r. indépendantes de même loi : on suppose que leur espérance (commune) est $M \in \mathbb{R}$ et leur variance (commune) est $V \in \mathbb{R}$. On pose

$$S_n = \frac{1}{n} \sum_{k=1}^{n} X_k$$

ce qui correspond à la moyenne des valeurs prises par les X_k. Alors

$$\mathbb{E}(S_n) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\mathbb{V}(S_n) = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

L'inégalité de Bienaymé-Tchebychev appliquée à S_n donne :

Interprétation :

On dit que la suite $(S_n)_{n \in \mathbb{N}}$ converge "en probabilité" vers M (hors-programme).

Exemple 14. On lance n fois une pièce équilibrée. Donner une minoration de la probabilité d'obtenir entre 40% et 60% de faces.

Note : si par exemple $n = 100$, la réponse exacte est

$$\sum_{k=40}^{60} \binom{100}{k} 0.5^{100} \approx 0.965$$

5 Compléments

On reproduit ici le Lemme 36.4 et on en donne une preuve.

Lemme 36.26

Soit $X : \Omega \to E$ une v.a. Alors
$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

Démonstration. On a
$$\Omega = X^{-1}(E) = X^{-1}\left(\bigcup_{x\in E}\{x\}\right) = \bigcup_{x\in E} X^{-1}(\{x\})$$
Par ailleurs, on sait que si $I_1, I_2, \cdots, I_n$ sont des ensembles disjoints, alors on a la formule
$$\sum_{i\in I_1\cup I_2\cup\ldots\cup I_n} a_i \quad = \quad \sum_{k=1}^{n}\sum_{i\in I_k} a_i$$
Comme les ensembles $X^{-1}(\{x\})$ sont disjoints pour deux valeurs de x distinctes, on peut donc écrire
$$\begin{aligned}
\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) &= \sum_{x\in E}\sum_{\omega\in X^{-1}(\{x\})} X(\omega)\mathbb{P}(\{\omega\}) \\
&= \sum_{x\in E}\sum_{\omega\in X^{-1}(\{x\})} x\mathbb{P}(\{\omega\}) \\
&= \sum_{x\in E} x \sum_{\omega\in X^{-1}(\{x\})} \mathbb{P}(\{\omega\}) \\
&= \sum_{x\in E} x\mathbb{P}\left(\bigcup_{\omega\in X^{-1}(\{x\})}\{\omega\}\right) && \text{car ces singletons } \{\omega\} \text{ sont disjoints deux à deux} \\
&= \sum_{x\in E} x\mathbb{P}\left(X^{-1}(\{x\})\right) \\
&= \sum_{x\in E} x\mathbb{P}(X = x) \\
&= \mathbb{E}(X)
\end{aligned}$$

□