

Chapitre 5 - Variables aléatoires discrètes réelles : espérance et variance

Dans ce chapitre, les commentaires *en italique* ne font pas partie intégrante du cours mais aident à sa compréhension, en particulier en faisant le lien entre une approche intuitive des probabilités et le formalisme de la théorie des probabilités.

I – Espérance

Dans cette section, les variables aléatoires seront supposées à valeurs réelles ou complexes.

1) Définition

Définition I.1.1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète **réelle ou complexe** définie sur $(\Omega, \mathcal{A})$.

- On dit que X est d'espérance finie si la famille $(xP(X=x))_{x \in X(\Omega)}$ est finie ou sommable. Dans ce cas, on appelle espérance de X et on note $E(X)$ la somme de cette famille.

On dit aussi que X admet un moment d'ordre 1.

- Si $X(\Omega)$ est fini, X est toujours d'espérance finie (probabilité de première année) et $E(X) = \sum_{x \in X(\Omega)} xP(X=x)$.
- Si $X(\Omega)$ est dénombrable et s'écrit $X(\Omega) = \{x_n, n \in \mathbb{N}\}$, alors X est d'espérance finie si la série $\sum x_n P(X=x_n)$ est **absolument** convergente et dans ce cas $E(X) = \sum_{n=0}^{+\infty} x_n P(X=x_n)$.
Si X est à valeurs réelles positives, X est d'espérance finie si $\sum_{n=0}^{+\infty} x_n P(X=x_n) < +\infty$. On peut alors noter $E(X) < +\infty$.
- Si X est d'espérance finie et $E(X) = 0$, on dit que X est une variable aléatoire centrée.

Remarque. (1) L'espérance de X peut être interprétée comme la moyenne des valeurs de X si on répète un grand nombre de fois la même expérience aléatoire.

(2) Si X est constante égale à $b \in \mathbb{C}$, alors $E(X) = b$.

(3) La définition ci-dessus se généralise au cas d'une variable aléatoire à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$.

Si $P(X = +\infty) > 0$, on convient que X n'est pas d'espérance finie.

Si $P(X = +\infty) = 0$, on convient que $xP(X=x) = 0$ si $x = +\infty$.

On a rencontré une telle variable aléatoire dans un jeu de Pile ou Face étudié dans le chapitre précédent. La convention ci-dessus consiste à ne pas considérer un événement négligeable dans un calcul d'espérance.

Proposition I.1.2. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète définie sur $(\Omega, \mathcal{A})$ telle que $X(\Omega) \subset \mathbb{N} \cup \{+\infty\}$.

On rappelle que, par convention : pour tout $n \in \mathbb{N}$, $n \leq +\infty$.

X est d'espérance finie si et seulement si la série $\sum_{n \geq 1} P(X \geq n)$ est convergente.

Dans ce cas : $E(X) = \sum_{n=1}^{+\infty} P(X \geq n)$. En cas de divergence, cette égalité reste vraie puisque $E(X) = \sum_{n=1}^{+\infty} P(X \geq n) = +\infty$.

Remarque. ✌

Dans la démonstration précédente, le point-clé est la relation $P(X = n) = P(X \geq n) - P(X \geq n + 1)$ (on peut aussi écrire $P(X = n) = P(X \leq n) - P(X \leq n - 1)$). En utilisant cette relation, on peut parfois obtenir la loi de X à partir des probabilités $P(X \geq n)$ (ou $P(X \leq n)$). On utilise en particulier cette astuce pour obtenir la loi d'un maximum ou d'un minimum. C'est le cas de l'exercice suivant.

Exercice I.1.3. Soient X et Y deux variables aléatoires indépendantes suivant respectivement des lois géométriques de paramètre $p \in]0; 1[$ et $q \in]0; 1[$. On note $Z = \min\{X, Y\}$. Déterminer la loi de Z .

2) Espérances des lois usuelles

On ne rappelle pas ici les résultats de première année sur les espérances des lois uniformes, de Bernoulli et binomiale.

Proposition I.2.4. Soit $p \in]0; 1[$. Soit X une variable aléatoire qui suit la loi géométrique de paramètre p . Alors X est d'espérance finie et $E(X) = \frac{1}{p}$.

Exemple I.2.5. On reprend l'exercice du jeu de Pile ou Face où X est le rang de la première apparition d'un Pile. On a déjà dit que $X \sim \mathcal{G}\left(\frac{1}{2}\right)$ donc $E(X) = 2$. On peut dire qu'en moyenne, le premier Pile arrive au deuxième lancer.

Proposition I.2.6. Soit $\lambda \in]0; +\infty[$. Soit X une variable aléatoire qui suit la loi de Poisson de paramètre λ . Alors X est d'espérance finie et $E(X) = \lambda$.

3) Propriétés de l'espérance

Théorème I.7 (Théorème de transfert - ADMIS). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit X une variable aléatoire discrète définie sur $(\Omega, \mathcal{A})$. Soit f une fonction définie sur $X(\Omega)$ et à valeurs réelles ou complexes. La variable aléatoire $f(X)$ est d'espérance finie si et seulement si la famille $(f(x)P(X = x))_{x \in X(\Omega)}$ est finie ou sommable. On a alors :

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

Cette somme est une somme finie ou la somme d'une famille sommable selon la nature de $X(\Omega)$.

Remarque. ✌ Il est donc inutile de connaître la loi de $f(X)$ pour en calculer l'espérance, la loi de X suffit!

Ce résultat s'applique au cas d'un couple (X, Y) de variables aléatoires discrètes et à une fonction f de ces variables à valeurs réelles :

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

Pour écrire cette égalité, on devra d'abord s'assurer que la famille est finie ou sommable.

Exercice I.3.8. Soit $p \in]0; 1[$. On considère une variable aléatoire X qui suit la loi géométrique $\mathcal{G}(p)$. Montrer l'existence et calculer $E(X^2)$.

Corollaire I.3.9. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit X une variable aléatoire discrète définie sur $(\Omega, \mathcal{A})$. Soit f une fonction définie sur $X(\Omega)$ et à valeurs réelles ou complexes. X est d'espérance finie si et seulement si $|X|$ est d'espérance finie.

Proposition I.3.10 (ADMIS). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes définies sur $(\Omega, \mathcal{A})$.

On suppose que Y est à valeurs réelles **positives** et que X est à valeurs réelles ou complexes.

Si Y est d'espérance finie et $|X| \leq Y$, alors X est d'espérance finie.

Proposition I.3.11 (Linéarité de l'espérance - ADMIS). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles ou complexes définies sur $(\Omega, \mathcal{A})$. Soit $\lambda \in \mathbb{C}$.

Si X et Y sont d'espérance finie, alors $X + \lambda Y$ est d'espérance finie et $E(X + \lambda Y) = E(X) + \lambda E(Y)$.

Exemple I.3.12. Soit $n \geq 2$. Soit $p \in]0; 1[$.

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant toutes la même loi de Bernoulli $\mathcal{B}(p)$.

On note $X = \sum_{i=1}^n X_i$ et on rappelle que X suit la loi binomiale $\mathcal{B}(n, p)$. Retrouver l'espérance de X à partir de l'espérance des X_i .

Corollaire I.3.13. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

On note L^1 l'ensemble des variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$ et d'espérance finie.

L^1 est un espace vectoriel et $X \mapsto E(X)$ est une forme linéaire sur L^1 .

Proposition I.3.14 (Positivité - Croissance). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes **réelles** définies sur $(\Omega, \mathcal{A})$.

(1) Si $X \geq 0$ (c'est-à-dire $X(\Omega) \subset \mathbb{R}_+$) et si X est d'espérance finie, alors $E(X) \geq 0$.

(2) Si $X \leq Y$ et si X et Y sont d'espérance finie, alors $E(X) \leq E(Y)$.

Remarque. On peut remarquer l'analogie entre les résultats précédents et les résultats sur les intégrales.

Proposition I.3.15. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète **réelle et positive** définie sur $(\Omega, \mathcal{A})$.

Alors : $E(X) = 0$ si et seulement si $0 \in X(\Omega)$ et $P(X = 0) = 1$ (on dit que X est presque sûrement nulle).

Exercice I.3.16. Démontrer qu'une variable aléatoire *presque sûrement* constante a une espérance finie égale à cette constante.

Proposition I.3.17. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles ou complexes définies sur $(\Omega, \mathcal{A})$.

Si X et Y sont **indépendantes** et d'espérance finie, alors XY est d'espérance finie et $E(XY) = E(X)E(Y)$.

Corollaire I.3.18. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $n \geq 2$.

Soient $X_1, \dots, X_n$ des variables aléatoires discrètes réelles ou complexes définies sur $(\Omega, \mathcal{A})$.

Si $X_1, \dots, X_n$ sont **indépendantes** et d'espérance finie, alors $X_1 \dots X_n$ est d'espérance finie et $E(X_1 \dots X_n) = E(X_1) \dots E(X_n)$.

II – Variance - Écart-type

Dans cette section, les variables aléatoires seront supposées à **valeurs réelles**.

Proposition II.0.1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{A})$.

Si X^2 est d'espérance finie, alors X est d'espérance finie.

Définition II.0.2. On note L^2 l'ensemble des variables aléatoires réelles telles que X^2 est d'espérance finie. Le résultat précédent assure que $L^2 \subset L^1$.

Proposition II.0.3 (Inégalité de Cauchy-Schwarz). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$.

Si X et Y sont dans L^2 , alors XY est dans L^1 et

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

Il y a égalité si et seulement s'il existe $\lambda \in \mathbb{R}$ tel que $P(X = \lambda Y) = 1$ ou $P(Y = \lambda X) = 1$ (on peut dire que X et Y sont presque sûrement colinéaires).

Corollaire II.0.4. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. L'ensemble L^2 est un espace vectoriel.

Exemple II.0.5. Les variables aléatoires constantes (ou seulement constantes presque sûrement) appartiennent à L^2 .

1) Définition

Définition - Théorème II.1.6. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{A})$. **On suppose que $X \in L^2$.**

On dit aussi que X admet un moment d'ordre 2.

Alors la variable aléatoire $X - E(X)$ appartient à L^2 et on appelle variance de X le réel positif

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

On appelle écart-type de X le réel

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

On a :

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

Cette somme est une somme finie ou la somme d'une famille sommable selon la nature de $X(\Omega)$.

On dit que X est réduite si $V(X) = \sigma(X) = 1$.

Remarque. (1) La variance et l'écart-type sont appelés des *paramètres de dispersion* de la variable aléatoire X .

Une "petite" valeur de la variance traduit que les valeurs de $(X - E(X))^2$ sont "petites" (ou du moins "fréquemment petites"), c'est-à-dire que les valeurs de X sont peu éloignées ("dispersées") de leur moyenne (l'espérance).

(2) Si X est une variable aléatoire réelle presque sûrement constante, alors $V(X) = 0$.

2) Propriétés

Proposition II.2.7. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{A})$. **On suppose que $X \in L^2$.**

$$(1) \quad V(X) = E(X^2) - (E(X))^2.$$

(2) Soient a et b deux réels.

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

En particulier : $V(X + b) = V(X)$ et $V(aX) = a^2 V(X)$.

(3) Supposons que $\sigma(X) \neq 0$.

La variable aléatoire $Y = \frac{X - E(X)}{\sigma(X)}$ est centrée réduite : $E(Y) = 0$ et $\sigma(Y) = 1$.

3) Variances des lois usuelles

On ne rappelle pas ici les résultats de première année sur les variances des lois uniformes, de Bernoulli et binomiale.

Proposition II.3.8. Soit $p \in]0; 1[$. Soit X une variable aléatoire qui suit la loi géométrique de paramètre p .

Alors $X \in L^2$ et $V(X) = \frac{1-p}{p^2}$.

Proposition II.3.9. Soit $\lambda \in]0; +\infty[$. Soit X une variable aléatoire qui suit la loi de Poisson de paramètre λ .

Alors $X \in L^2$ et $V(X) = \lambda$.

4) Covariance et variance d'une somme de variables aléatoires

Définition - Théorème II.4.10. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$. **On suppose que $X \in L^2$ et $Y \in L^2$.**

Alors la quantité $E(XY) - E(X)E(Y)$ est bien définie et est appelée covariance de X et Y .

On note $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. En particulier : $V(X) = \text{Cov}(X, X)$.

Remarque. ✌ Pour calculer $E(XY)$ (et ensuite la covariance de X et Y), il est souvent inutile de déterminer la loi de XY .

On peut utiliser le théorème de transfert : $E(XY) = \sum_{(i,j) \in X(\Omega) \times Y(\Omega)} ij P(X = i, Y = j)$.

Proposition II.4.11. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$. **On suppose que $X \in L^2$ et $Y \in L^2$.**

Si X et Y sont **indépendantes** alors $\text{Cov}(X, Y) = 0$.

Remarque. Si deux variables aléatoires sont indépendantes, leur covariance est nulle.

En revanche, on peut avoir $\text{Cov}(X, Y) = 0$ (on dit que X et Y sont *décorrélées*) avec X et Y qui ne sont pas indépendantes.

On peut par exemple considérer la variable aléatoire X à valeurs dans $\{-1, 0, 1\}$ dont la loi est donnée par

$$P(X = 0) = \frac{1}{5}, \quad P(X = -1) = P(X = 1) = \frac{2}{5},$$

ainsi que $Y = X^2$. Ces deux variables aléatoires ont une covariance nulle et ne sont pas indépendantes. Le prouver!

Proposition II.4.12. (1) Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soient X et Y deux variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$.

On suppose que $X \in L^2$ **et** $Y \in L^2$.

Alors $X + Y \in L^2$ et on a

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

En particulier, si X et Y sont **indépendantes** (ou simplement **décorrélées**), alors $V(X + Y) = V(X) + V(Y)$.

(2) Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit $n \geq 2$.

Soient $X_1, \dots, X_n$ des variables aléatoires discrètes réelles définies sur $(\Omega, \mathcal{A})$.

On suppose que pour tout $1 \leq i \leq n$, **on a** $X_i \in L^2$.

Alors $X_1 + \dots + X_n \in L^2$ et on a :

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n) + 2 \sum_{1 \leq k < \ell \leq n} \text{Cov}(X_k, X_\ell).$$

En particulier, si les variables X_k sont **deux à deux indépendantes** (ou simplement **décorrélées**), on a

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

Exemple II.4.13. Soit $n \geq 2$. Soit $p \in]0; 1[$.

Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant toutes la même loi de Bernoulli $\mathcal{B}(p)$.

On note $X = \sum_{i=1}^n X_i$ et on rappelle que X suit la loi binomiale $\mathcal{B}(n, p)$. Retrouver la variance de X à partir de la variance des X_i .

Tableau récapitulatif des lois usuelles

$X \sim \dots$	b	$\mathcal{U}([1; n])$	$\mathcal{B}(p)$	$\mathcal{B}(n, p)$	$\mathcal{G}(p)$	$\mathcal{P}(\lambda)$
paramètre(s)	$b \in \mathbb{C}$	$n \in \mathbb{N}^*$	$p \in [0; 1]$	$n \in \mathbb{N}, p \in [0; 1]$	$p \in]0; 1[$	$\lambda \in \mathbb{R}_+^*$
$X(\Omega)$	$\{b\}$	$[1; n]$	$\{0, 1\}$	$[0; n]$	$\mathbb{N}^*$	$\mathbb{N}$
$P(X = k)$	1	$\frac{1}{n}$	p si $k = 1$; $1 - p$ si $k = 0$	$\binom{n}{k} p^k (1 - p)^{n-k}$	$p(1 - p)^{k-1}$	$e^{-\lambda} \frac{\lambda^k}{k!}$
$E(X)$	b	$\frac{n+1}{2}$	p	np	$\frac{1}{p}$	λ
$V(X)$	0	$\frac{n^2 - 1}{12}$	$p(1 - p)$	$np(1 - p)$	$\frac{1 - p}{p^2}$	λ

III – Inégalité de Bienaymé-Tchebychev - Loi faible des grands nombres

Proposition III.0.1 (Inégalité de Markov). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{A})$. On suppose que $X \in L^1$.

(1) Supposons que X est à valeurs positives. On a alors, pour tout $a > 0$:

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

(2) Dans le cas général, pour tout $a > 0$:

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

Théorème III.2 (Inégalité de Bienaymé-Tchebychev).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit X une variable aléatoire discrète réelle définie sur $(\Omega, \mathcal{A})$. On suppose que $X \in L^2$.

On a, pour tout réel $\varepsilon > 0$:

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

Remarque. Cet énoncé signifie que la probabilité que la variable aléatoire X prenne des valeurs éloignées de son espérance est relativement faible, d'autant plus faible que sa variance est petite.

En utilisant l'événement contraire, on en déduit que $P(|X - E(X)| < \varepsilon) \geq 1 - \frac{V(X)}{\varepsilon^2}$.

Par exemple, la probabilité que X soit dans $]E(X) - 2\sigma(X); E(X) + 2\sigma(X)[$ est supérieure à $\frac{3}{4}$.

Exemple III.0.3. On reprend l'exercice du jeu de Pile ou Face où X est le rang de la première apparition d'un Pile.

On a déjà dit que $X \sim \mathcal{G}\left(\frac{1}{2}\right)$ donc $E(X) = 2$ et $V(X) = 2$. Démontrer qu'avec une probabilité supérieure à $\frac{7}{8}$, le premier Pile apparaîtra strictement avant le sixième lancer. Que vaut exactement $P(X \leq 5)$?

Exercice III.0.4. On lance un dé cubique équilibré à 6 faces n fois. On note X le nombre d'apparitions de la valeur 6 lors des n lancers. On appelle fréquence d'apparition du 6 la variable aléatoire $F = \frac{X}{n}$.

(1) Reconnaître la loi de X ? En déduire l'espérance et la variance de F .

(2) Démontrer que :

$$\lim_{n \rightarrow +\infty} P\left(\left|F - \frac{1}{6}\right| \leq 0.01\right) = 1.$$

(3) Quelle valeur suffit-il de donner à n pour que F appartienne à $\left[\frac{1}{6} - 0.01; \frac{1}{6} + 0.01\right]$ avec une probabilité supérieure ou égale à 0.95?

Théorème III.5 (Loi faible des grands nombres). Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé.

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles discrètes indépendantes identiquement distribuées (suite iid).

On suppose que $X_1 \in L^2$ (donc toutes les variables X_i également) et on note $\mu = E(X_1)$ et $\sigma = \sigma(X_1)$.

Pour tout $n \in \mathbb{N}^*$, on note $S_n = \sum_{k=1}^n X_k$. On a alors, pour tout $\varepsilon > 0$:

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = 0.$$

Plus précisément :

$$\forall \varepsilon > 0, P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Remarque. En utilisant l'événement contraire, on peut aussi écrire

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) = 1.$$

On peut interpréter ce résultat dans le cadre de la répétition d'une même expérience aléatoire n fois.

Pour un ε fixé (même très petit), si n est choisi suffisamment grand, la moyenne des X_i sera très probablement dans l'intervalle $[\mu - \varepsilon; \mu + \varepsilon]$.