

Dans l'ensemble de ce chapitre,  $(\Omega, P)$  désigne un espace probabilisé fini.

## 1 Espérance d'une variable aléatoire réelle ou complexe

### Définition 1 – Espérance d'une variable aléatoire réelle ou complexe

Soit  $X$  une variable aléatoire à valeurs complexes sur  $\Omega$ .

- On appelle *espérance de  $X$*  et on note  $\mathbf{E}(X)$  le complexe défini par  $\mathbf{E}(X) = \sum_{x \in X[\Omega]} P(X = x)x$ .
- La variable aléatoire  $X$  est dite *centrée* lorsque  $\mathbf{E}(X) = 0$ .

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 comptabilisé en proportion de sa probabilité d'occurrence  $P(X = x)$  : plus  $P(X = x)$  est proche de 1, plus  $x$  a d'importance dans la somme.

**Exemple 2** Soit  $n \in \mathbb{N}^*$ . Si  $X$  est une variable aléatoire uniforme sur  $\llbracket 1, n \rrbracket$ , alors  $\mathbf{E}(X) = \frac{n+1}{2}$ .

**Remarque 3** Dans un jeu de hasard, l'espérance représente la chance de gain (algébrique) du joueur. Un jeu est dit *équitable* lorsque l'espérance de gain des joueurs est nulle.

### Théorème 4 – Espérance des lois usuelles

Soit  $X$  une variable aléatoire sur  $\Omega$ ,  $n \in \mathbb{N}^*$  et  $p \in [0, 1]$ .

- (i) **Variable aléatoire constante.** Si  $X$  est constante égale à  $c$ , alors  $\mathbf{E}(X) = c$ .
- (ii) **Loi uniforme.** Si  $X \sim \mathcal{U}(E)$ , avec  $E = \{x_1, \dots, x_n\} \subset \mathbb{C}$ , alors  $\mathbf{E}(X) = \frac{1}{n} \sum_{i=1}^n x_i$ , et l'espérance de  $X$  n'est autre que la moyenne arithmétique des  $x_i$ .
- (iii) **Loi de Bernoulli.** Si  $X \sim \mathcal{B}(p)$ , alors  $\mathbf{E}(X) = p$ .  
**Exemple fondamental.** Pour tout événement  $A \in \mathcal{P}(\Omega)$ ,  $\mathbf{E}(1_A) = P(A)$ .
- (iv) **Loi binomiale.** Si  $X \sim \mathcal{B}(n, p)$ , alors  $\mathbf{E}(X) = np$ .

*Démonstration. ...* ■

### Théorème 5 – Propriétés de l'espérance

Soit  $X$  et  $Y$  deux variables aléatoires complexes sur  $\Omega$ .

- (i) **Expression alternative.**  $\mathbf{E}(X) = \sum_{\omega \in \Omega} P(\{\omega\})X(\omega)$ .
- (ii) **Linéarité.** Pour tous  $a, b \in \mathbb{C}$ ,  $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$ .
- (iii) **Inégalité triangulaire.**  $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$ .

Les deux dernières propriétés concernent uniquement les variables aléatoires RÉELLES.

- (iv) **Positivité.** Si  $X \geq 0$ , alors  $\mathbf{E}(X) \geq 0$ .
- (v) **Croissance.** Si  $X \leq Y$ , alors  $\mathbf{E}(X) \leq \mathbf{E}(Y)$ .

*Démonstration. ...* ■

L'expression alternative de l'espérance donnée au point (i) offre un changement de point de vue, parfois fructueux, en se basant sur l'univers  $\Omega$  plutôt que sur l'ensemble des valeurs de  $X$ .

**Définition-théorème 6 – Variable centrée associée à une variable aléatoire**

Soit  $X$  une variable aléatoire complexe sur  $\Omega$ . On appelle *variable centrée associée* à  $X$  la variable aléatoire centrée  $X^+ = X - \mathbf{E}(X)$ .

*Démonstration.* Par linéarité de l'espérance,  $\mathbf{E}(X^+) = \mathbf{E}(X) - \mathbf{E}(\mathbf{E}(X)) = 0$ . ■

**Exemple 7** Qu'obtient-on en moyenne quand on lance 2 fois un dé à 6 faces et qu'on additionne les résultats obtenus?

**Exemple 8 – Retour sur l'espérance d'une loi binomiale** La linéarité de l'espérance permet de déduire l'espérance d'une loi binomiale à partir de celle d'une loi de Bernoulli. Précisément, pour tous  $p \in [0, 1]$  et  $n \in \mathbb{N}^*$ , si  $X_1, \dots, X_n$  sont des variables de Bernoulli indépendantes de paramètre  $p$  définies sur un même espace probabilisé, alors  $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$ , or

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

**Exemple 9 – Une preuve de la formule du crible** Pour tous  $A_1, \dots, A_n \in \mathcal{P}(\Omega)$ ,

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

**Théorème 10 – Formule de transfert**

Soit  $X$  une variable aléatoire et  $f : X[\Omega] \rightarrow \mathbb{C}$  une fonction. L'espérance de la variable aléatoire  $f(X)$  est entièrement déterminée par  $f$  et la loi de  $X$  :

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

*Démonstration.* La famille  $((X = x))_{x \in X[\Omega]}$  est un système complet d'événements, ainsi

$$\mathbf{E}(f(X)) = \sum_{\omega \in \Omega} \mathbf{P}(\{\omega\}) f(X(\omega)) = \sum_{x \in X[\Omega]} \sum_{\omega \in (X=x)} \mathbf{P}(\{\omega\}) f(X(\omega)) = \sum_{x \in X[\Omega]} \left( \sum_{\omega \in (X=x)} \mathbf{P}(\{\omega\}) \right) f(x) = \sum_{x \in X[\Omega]} \mathbf{P}(X = x) f(x). \quad \blacksquare$$

**En pratique** On peut donc calculer  $\mathbf{E}(f(X))$  directement à partir de la loi de  $X$ , sans avoir à déterminer celle de  $f(X)$ .

**Exemple 11** La formule de transfert permet de donner une nouvelle preuve de l'inégalité triangulaire pour l'espérance d'une variable aléatoire.

**En effet**, si  $X$  est une variable aléatoire complexe, alors, d'après la formule de transfert avec  $f : x \mapsto |x|$ ,

$$\mathbf{E}(|X|) = \sum_{x \in X[\Omega]} \mathbf{P}(X = x) |x| = \sum_{x \in X[\Omega]} |\mathbf{P}(X = x) x| \geq \left| \sum_{x \in X[\Omega]} \mathbf{P}(X = x) x \right| = |\mathbf{E}(X)|.$$

**Remarque 12** La formule de transfert vaut bien sûr pour les couples et plus généralement pour les  $n$ -uplets de variables aléatoires complexes. Par exemple, si  $X$  et  $Y$  sont deux variables aléatoires complexes et si  $f : X[\Omega] \times Y[\Omega] \rightarrow \mathbb{C}$ , alors

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

**Exemple 13** Une urne contient  $n$  boules numérotées de 1 à  $n$ , avec  $n \geq 2$ . On effectue deux tirages successifs et sans remise dans cette urne et on note  $X_1$  (resp.  $X_2$ ) le numéro de la première (resp. seconde) boule tirée. On a  $\mathbf{E}(X_1 X_2) = \frac{(3n+2)(n+1)}{12}$ .

Terminons par un résultat concernant l'espérance de deux variables aléatoires indépendantes.

#### Théorème 14 – Espérance du produit de deux variables aléatoires indépendantes

Soit  $X$  et  $Y$  deux variables aléatoires complexes sur  $\Omega$ . Si  $X$  et  $Y$  sont INDÉPENDANTES, alors

$$\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y).$$

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

*Démonstration.* D'après la formule de transfert (avec  $f : (x, y) \mapsto xy$ ),

$$\mathbf{E}(XY) = \sum_{(x,y) \in X[\Omega] \times Y[\Omega]} \mathbf{P}((X=x) \cap (Y=y)) xy \stackrel{\text{indép.}}{=} \sum_{(x,y) \in X[\Omega] \times Y[\Omega]} \mathbf{P}(X=x) \mathbf{P}(Y=y) xy = \sum_{x \in X[\Omega]} \mathbf{P}(X=x) x \sum_{y \in Y[\Omega]} \mathbf{P}(Y=y) y = \mathbf{E}(X) \mathbf{E}(Y). \quad \blacksquare$$

✘ **ATTENTION ! ✘** L'identité «  $\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y)$  » est fautive en général. Par exemple, pour toute variable aléatoire  $X$  de loi uniforme sur  $\{-1, 1\}$ ,  $\mathbf{E}(X) = 0$  et donc  $\mathbf{E}(X)^2 = 0$ , tandis que  $\mathbf{E}(X^2) = \mathbf{E}(1) = 1$ .

## 2 Variance d'une variable aléatoire réelle, écart-type et covariance

### 2.1 Variance d'une variable aléatoire réelle

Conformément au programme, la notion de variance est seulement introduite pour les variables aléatoires réelles.

#### Définition 15 – Variance et écart-type d'une variable aléatoire réelle

Soit  $X$  une variable aléatoire réelle sur  $\Omega$ .

- On appelle *variance de  $X$*  et on note  $\mathbf{V}(X)$  le réel positif défini par  $\mathbf{V}(X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right)$ .
- On appelle *écart-type de  $X$*  le réel  $\sigma(X) = \sqrt{\mathbf{V}(X)}$ . • La variable  $X$  est dite *réduite* lorsque  $\mathbf{V}(X) = 1$ .

#### Remarque 16

- L'espérance de  $X$  est un indicateur de position, on peut alors se demander si les valeurs de  $X$  sont plutôt proches de cette valeur moyenne ou plutôt éloignées. Toute mesure de cette proximité à la moyenne est appelée un *indicateur de dispersion*. À proprement parler, l'écart naturel entre  $X$  et son espérance est  $|X - \mathbf{E}(X)|$ , ainsi l'écart moyen de  $X$  à sa moyenne est  $\mathbf{E}(|X - \mathbf{E}(X)|)$ , mais ce n'est pas l'indicateur de dispersion que les mathématiciens ont choisi d'utiliser couramment. Ces derniers lui ont préféré la variance, c'est-à-dire l'écart QUADRATIQUE moyen à la moyenne – « quadratique » parce que l'on passe au carré. Ce choix, moins naturel au premier abord, se justifie par une plus grande simplicité de manipulation.
- Quelle différence entre écart-type et variance? Si  $X$  représente par exemple une longueur,  $\mathbf{V}(X)$  représente une longueur AU CARRÉ. Ainsi il n'est pas possible de comparer directement la position moyenne  $\mathbf{E}(X)$  et sa dispersion  $\mathbf{V}(X)$  – un physicien dirait que l'espérance et la variance ne sont pas « homogène ». L'écart-type, au contraire, est homogène à une longueur, donc comparable à l'espérance, d'où son intérêt.

#### Théorème 17 – Propriétés de la variance

Soit  $X$  une variable aléatoire réelle sur  $\Omega$

- (i) **Formule de König-Huygens.**  $\mathbf{V}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$ .
- (ii) **Nullité.**  $\mathbf{V}(X) = 0$  si et seulement si  $\mathbf{P}(X = \mathbf{E}(X)) = 1$ , i.e.  $X$  est dite *presque sûrement constante*.
- (iii) **Effet d'une transformation affine.** Pour tous  $a, b \in \mathbb{R}$ ,  $\mathbf{V}(aX + b) = a^2 \mathbf{V}(X)$ .

En particulier, si  $\sigma(X) \neq 0$ , alors la variable aléatoire  $X^* = \frac{X - \mathbf{E}(X)}{\sigma(X)}$  est centrée-réduite.

*Démonstration.* ... ■

**Remarque 18**

- Le résultat ci-dessus, concernant l'effet d'une transformation affine sur la variance, se comprend aisément si l'on a à l'esprit que la variance mesure la dispersion de la loi d'une variable aléatoire réelle. En effet, lorsque l'on translate une variable aléatoire d'une constante  $b$ , on ne modifie pas la dispersion de sa loi et la variance est donc constante :  $\mathbf{V}(X + b) = \mathbf{V}(X)$ .
- Centrer et réduire permet d'obtenir des variables représentant des données indépendantes des unités ou de l'échelle choisies et des variables ayant même moyenne et même dispersion. Autrement dit, des variables plus facilement comparables. Il s'agit d'un procédé de normalisation.
- Le point **(ii)** du théorème 17 exprime qu'une variable aléatoire réelle sur  $\Omega$  suit une loi certaine si et seulement si sa variance est nulle.

**Exemple 19** Soit  $n \in \mathbb{N}^*$ . Si  $X$  une variable aléatoire uniforme sur  $\llbracket 1, n \rrbracket$ , alors  $\mathbf{V}(X) = \frac{n^2 - 1}{12}$ .

**Théorème 20 – Variances des lois de Bernoulli et binomiale**

Soit  $X$  un variable aléatoire sur  $\Omega$ ,  $p \in [0, 1]$  et  $n \in \mathbb{N}^*$ .

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

*Démonstration. ...* ■

**2.2 Covariance de deux variables aléatoires réelles**

La *covariance* de deux variables aléatoires réelles est destinée à mesurer le défaut par rapport à l'indépendance.

**Définition 21 – Covariance de deux variables aléatoires réelles**

Soit  $X$  et  $Y$  deux variables aléatoires réelles sur  $\Omega$ . On appelle *covariance de  $X$  et  $Y$*  et on note  $\text{Cov}(X, Y)$  le réel

$$\text{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))).$$

Les variable aléatoires  $X$  et  $Y$  sont dites *décorrélées* lorsque  $\text{Cov}(X, Y) = 0$ .

**Théorème 22 – Propriétés de la covariance**

Soit  $X, Y, Z, X_1, \dots, X_n$  des variables aléatoires réelles sur  $\Omega$  et  $(a, b) \in \mathbb{R}^2$ .

- (i) Expression développée.**  $\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ .  
**(ii) Positivité.**  $\text{Cov}(X, X) = \mathbf{V}(X) \geq 0$ .      **(iii) Symétrie.**  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ .  
**(iv) Bilinéarité.**

$$\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z) \quad \text{et} \quad \text{Cov}(Z, aX + bY) = a \text{Cov}(Z, X) + b \text{Cov}(Z, Y).$$

- (v) Lien avec la variance.**  $\mathbf{V}(X + Y) = \mathbf{V}(X) + 2 \text{Cov}(X, Y) + \mathbf{V}(Y)$ .

Plus généralement : 
$$\mathbf{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \quad (\star)$$

- (vi) Lien avec l'indépendance.** Si  $X$  et  $Y$  sont indépendantes, alors  $\text{Cov}(X, Y) = 0$ .

D'après **(v)** et **(vi)**, si  $X$  et  $Y$  sont indépendantes (ou seulement décorréélées), alors  $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$ . Plus généralement, si  $X_1, \dots, X_n$  sont DEUX À DEUX indépendantes (ou seulement deux à deux décorréélées), alors

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

*Démonstration. ...* ■

**✘ ATTENTION ! ✘** La réciproque de **(vi)** est fautive ! La covariance de  $X$  et  $Y$  peut être nulle sans que  $X$  et  $Y$  soient indépendantes. Considérons par exemple une variable aléatoire  $X$  suivant la loi uniforme sur  $\{-1, 0, 1\}$  et la variable aléatoire  $Y = 1 - X^2$  de support  $Y[\Omega] = \{0, 1\}$ . Les variables  $X$  et  $Y$  ne sont pas indépendantes, puisque

$$P((X = 1) \cap (Y = 1)) = P((X = 1) \cap (X = 0)) = P(\emptyset) = 0 \quad \text{et} \quad P(X = 1)P(Y = 1) = P(X = 1)P(X = 0) = \frac{1}{9}.$$

En revanche, puisque  $XY = X - X^3$  est constante égale à 0 et puisque  $X$  est d'espérance nulle,

$$\text{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y) = \mathbf{E}(0) - 0 \times \mathbf{E}(Y) = 0.$$

**Remarque 23**

- Dans la terminologie du chapitre suivant, la covariance est une *forme bilinéaire symétrique positive*.
- La formule **(★)** du point **(v)** rappelle fortement la formule classique

$$\forall x_1, \dots, x_n \in \mathbb{R}, \quad \left( \sum_{k=1}^n x_k \right)^2 = \sum_{k=1}^n x_k^2 + 2 \sum_{1 \leq i < j \leq n} x_i x_j,$$

qui elle-même est une généralisation de l'identité remarquable  $(x + y)^2 = x^2 + 2xy + y^2$ . Il s'agit ainsi de retenir que « Le carré de la somme est égal à la somme des carrés plus la somme des doubles produits ».

**Exemple 24 – Retour sur la variance d'une loi binomiale** La linéarité de la variance pour des variables aléatoires INDÉPENDANTES permet à nouveau de déduire la variance d'une loi binomiale à partir de celle d'une loi de Bernoulli. Précisément, pour tous  $p \in [0, 1]$  et  $n \in \mathbb{N}^*$ , si  $X_1, \dots, X_n$  sont des variables de Bernoulli indépendantes de paramètre  $p$  définies sur un même espace probabilisé, alors  $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$ , or

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

Voyons sur un exemple comment les propriétés de la covariance permettent en pratique de calculer des variances ou covariances en évitant de recourir à la formule de transfert.

**Exemple 25** Une urne contient des boules de  $N$  types différents (bleues, jaunes, ...) mais indiscernables au toucher. Le type  $i$  est présent en proportion  $p_i$ . On tire  $n$  boules au hasard avec remise et on note  $X_i$  le nombre de boules de type  $i$  obtenues, alors, pour tous  $1 \leq i < j \leq N$ ,  $\text{Cov}(X_i, X_j) = -np_i p_j$ .

**En effet**, pour tout  $1 \leq i \leq N$ , les tirages ayant lieu avec remise, la variable aléatoire  $X_i$  suit la loi binomiale  $\mathcal{B}(n, p_i)$ . En outre, pour tous  $1 \leq i < j \leq N$ ,  $X_i + X_j$  représente le nombre de boules de type «  $i$  ou  $j$  » tirées. Ces dernières étant en proportion  $p_i + p_j$ , la variable aléatoire  $X_i + X_j$  suit la loi binomiale  $\mathcal{B}(n, p_i + p_j)$ . Or,

$$\mathbf{V}(X_i + X_j) = \mathbf{V}(X_i) + 2 \text{Cov}(X_i, X_j) + \mathbf{V}(X_j),$$

d'où

$$\text{Cov}(X_i, X_j) = \frac{1}{2} [\mathbf{V}(X_i + X_j) - \mathbf{V}(X_i) - \mathbf{V}(X_j)] = \frac{n}{2} [(p_i + p_j)(1 - p_i - p_j) - p_i(1 - p_i) - p_j(1 - p_j)]$$

et le résultat annoncé, après simplification.

### 3 Inégalités probabilistes

Parmi les inégalités probabilistes qui jouent un grand rôle théorique, les inégalités de Markov et de Bienaymé-Tchebychev se distinguent, notamment par la simplicité de leurs hypothèses.

**Théorème 26 – Inégalité de Markov†**

Si  $X$  est une variable aléatoire sur  $\Omega$  à valeurs POSITIVES, alors

$$\forall a > 0, \quad P(X \geq a) \leq \frac{\mathbf{E}(X)}{a}.$$

*Démonstration.* ... ■

†. Andreï Andreïevitch Markov (1856 à Riazan – 1922 à Pétrograd) est un mathématicien russe, considéré comme le fondateur de la théorie des processus stochastiques.

**Exemple 27** Les salaires étant positifs, la part de la population percevant un salaire supérieur à 5 fois le salaire moyen est au maximum d'un cinquième.

**Exemple 28** Soit  $f$  une fonction croissante et positive sur un intervalle  $I$  et  $Y$  une variable aléatoire à valeurs dans  $I$  sur  $\Omega$ , alors

$$\forall b \in I, \quad \left( f(b) > 0 \implies P(Y \geq b) \leq \frac{\mathbf{E}(f(Y))}{f(b)} \right).$$

L'inégalité de Markov permet d'établir l'inégalité de concentration fondamentale suivante.

**Théorème 29 – Inégalité de Bienaymé-Tchebychev<sup>‡</sup>**

Si  $X$  est une variable aléatoire réelle sur  $\Omega$ , alors

$$\forall \varepsilon > 0, \quad P(|X - \mathbf{E}(X)| \geq \varepsilon) \leq \frac{\mathbf{V}(X)}{\varepsilon^2}.$$

*Démonstration. ...* ■

**Remarque 30**

- L'inégalité de Bienaymé-Tchebychev exprime clairement que la variance est une mesure de la dispersion d'une variable aléatoire autour de sa valeur moyenne : la probabilité pour que  $X$  prenne des valeurs éloignées de  $\mathbf{E}(X)$  d'au moins  $\varepsilon$  est d'autant plus faible que  $\mathbf{V}(X)$  est plus petite.
- L'inégalité de Bienaymé-Tchebychev est qualifiée d'inégalité de concentration dans la mesure où elle indique que la variable aléatoire  $X$  prendra avec une faible probabilité une valeur relativement lointaine de son espérance, ou encore que  $X$  fluctue dans l'intervalle  $]\mathbf{E}(X) - \varepsilon, \mathbf{E}(X) + \varepsilon[$  avec une probabilité supérieure à  $1 - \frac{\mathbf{V}(X)}{\varepsilon^2}$ .
- À l'instar de l'inégalité de Markov, celle de Bienaymé-Tchebychev nécessite relativement peu d'hypothèses sur  $X$  mais s'avère en général assez grossière (cf. exemple ci-après).
- La majoration établie par l'inégalité de Bienaymé-Tchebychev est triviale dès que  $\varepsilon \leq \sigma(X)$ .

**Exemple 31** On lance une pièce de monnaie équilibrée dix fois de suite et on note  $X$  le nombre d'apparition de la face pile. Ainsi  $X$  suit la loi binomiale  $\mathcal{B}(10, 1/2)$ , pour laquelle  $\mathbf{E}(X) = 5$  et  $\mathbf{V}(X) = \frac{5}{2}$ .

L'inégalité de Bienaymé-Tchebychev fournit alors, pour tout  $\varepsilon > 0$ , la majoration  $P(|X - 5| \geq \varepsilon) \leq \frac{5}{2\varepsilon^2}$ , et, pour  $\varepsilon = 2$ , on obtient donc  $P(|X - 5| \geq 2) \leq 0,625$ . Or un calcul exact donne

$$P(|X - 5| \geq 2) = 1 - (P(X = 4) + P(X = 5) + P(X = 6)) = 1 - \frac{1}{2^{10}} \left( \binom{10}{4} + \binom{10}{5} + \binom{10}{6} \right) \approx 0,344.$$

### 3.1 Application aux intervalles de confiance

**Exemple 32** On dispose d'une pièce éventuellement truquée pour laquelle la probabilité d'apparition de « Pile » est notée  $p$ . Pour connaître  $p$ , on lance cette pièce  $n$  fois et on note  $F$  la fréquence d'apparition de « Pile » ainsi obtenue. À partir de quelle valeur de  $n$  la probabilité pour que  $F$  soit une approximation de  $p$  à  $10^{-2}$  près est-elle supérieure à 0,9 ?

**En effet**, notant  $X$  le nombre de « Pile » obtenus, on a  $F = \frac{X}{n}$  et, par indépendance des lancers,  $X \sim \mathcal{B}(n, p)$ . On a donc  $\mathbf{E}(F) = \frac{1}{n} \mathbf{E}(X) = p$ , et  $F$  va nous permettre d'estimer  $p$ . En outre, d'après l'inégalité de Bienaymé-Tchebychev,

$$P(|F - p| \geq 10^{-2}) = P(|X - \mathbf{E}(X)| \geq 10^{-2}n) \leq \frac{\mathbf{V}(X)}{(10^{-2}n)^2} = \frac{np(1-p)}{10^{-4}n^2} = \frac{10^4}{n} \underbrace{p(1-p)}_{\leq 1/4} \leq \frac{2500}{n}.$$

Or on a l'équivalence

$$P(|F - p| < 10^{-2}) \geq 0,9 \iff P(|F - p| \geq 10^{-2}) \leq 0,1$$

<sup>‡</sup> Irénée-Jules Bienaymé (1796 à Paris – 1878 à Paris) est un mathématicien français qui contribua à la théorie des probabilités et des statistiques et leurs applications aux calculs financiers et démographiques.

Pafnouti Lvovitch Tchebychev (1821 à Okatovo – 1894 à Saint-Pétersbourg) est un mathématicien russe qui contribua à la théorie des probabilité et à la théorie de nombres.

et il nous suffit donc de choisir  $n$  tel que  $\frac{2500}{n} \leq 0,1$  pour garantir que  $F$  soit une approximation de  $p$  à  $10^{-2}$  près, avec une probabilité de 90%. En pratique, cela exige tout de même de réaliser 25 000 lancers! Ou d'obtenir une majoration de  $P(|F - p| \geq 10^{-2})$  plus fine que celle proposée par l'inégalité de Bienaymé-Tchebychev...

**Exemple 33** Un institut de sondage cherche à anticiper le résultat d'un référendum en interrogeant un échantillon de  $n$  personnes. Au moment du sondage, la proportion des « oui » dans la population complète est  $p$  (valeur inconnue que l'institut essaie justement d'estimer).

Si l'on note  $\bar{p}$  la proportion de « oui » dans l'échantillon des  $n$  personnes sondées, comment faut-il choisir  $n$  afin d'assurer que la probabilité de l'événement  $(|p - \bar{p}| \geq 0,05)$  soit inférieure ou égale à 0,1 ?

**En effet**, notons  $N$  le nombre total d'individus de la population étudiée. Le nombre  $n\bar{p}$  de « oui » collectés lors du sondage suit la loi hypergéométrique  $\mathcal{H}(N, n, p)$  (ou loi des tirage sans remise). Toutefois, lors d'un sondage, l'échantillon  $n$  est très petit devant la population totale  $N$  et de ce fait la distinction « avec » et « sans » remise perd tout intérêt. On peut alors considérer que la loi du nombre  $n\bar{p}$  des « oui » est binomiale  $\mathcal{B}(n, p)$  (approximation d'une loi hypergéométrique par une loi binomiale). L'inégalité de Bienaymé-Tchebychev donne aussitôt :

$$P(|\bar{p} - p| \geq 0,05) = P(|n\bar{p} - np| \geq 0,05n) \leq \frac{\mathbf{V}(n\bar{p})}{(0,05n)^2} = \frac{np(1-p)}{0,0025n^2} = \frac{400}{n} \underbrace{p(1-p)}_{\leq 1/4} = \frac{100}{n}.$$

Une condition suffisante sur  $n$  est alors  $\frac{100}{n} \leq 0,1$ , soit  $n \geq 1000$ .

## 3.2 Loi faible des grands nombres

L'inégalité de Bienaymé-Tchebychev va s'avérer suffisante pour établir la loi faible des grands nombres pour les variables aléatoires réelles définies sur un espace probabilisé fini.

### Théorème 34 – Loi faible des grands nombres

Si  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires réelles définies sur  $\Omega$  indépendantes de même loi et d'espérance  $\mu$ , alors

$$\forall \varepsilon > 0, \quad P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow +\infty} 0.^\dagger$$

*Démonstration.* ...

La loi faible des grands nombres permet notamment de légitimer a posteriori notre interprétation *fréquentiste* des probabilités.

Pour déterminer empiriquement la probabilité d'un événement  $A$  lié à une expérience aléatoire, une idée courante consiste à répéter l'expérience aléatoire un grand nombre de fois et à observer le nombre de réalisations de l'événement  $A$ . Lorsque le nombre d'expériences augmente, il est légitime d'espérer que la fréquence d'apparition de l'événement  $A$  se rapprochera de sa probabilité.

Précisément, notons  $X_k$  la variable aléatoire égale à 1 lorsque l'événement  $A$  se réalise lors de la  $k^e$  répétition de l'expérience et 0 sinon. La suite  $(X_k)_{k \geq 1}$  est donc une suite de variables aléatoires de Bernoulli indépendantes de paramètre  $p$  et  $\frac{1}{n} \sum_{k=1}^n X_k$  est la fréquence empirique d'apparition de l'événement  $A$  lors des  $n$  répétitions de l'expérience.

Or, d'après la loi faible des grands nombres,

$$\forall \varepsilon > 0, \quad P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - p\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow +\infty} 0.$$

Ainsi la fréquence d'un événement lorsque l'on répète indéfiniment une expérience aléatoire tend vers la probabilité a priori de cet événement.

†. On dit que la variable aléatoire  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$  converge en probabilité vers  $\mu$  (terminologie hors programme).

## Compétences à acquérir

- Calculer et manipuler l'espérance d'une variable aléatoire : exercices 1 à 14, 16, 18, 22 à 25, 27, 30, 34, 40, 41 et 47.
- Utiliser la formule de transfert : exercices 28, 29 et 43.
- Calculer et manipuler la variance d'une variable aléatoire : exercices 1, 2, 6, 8, 14, 18, 25, 29, 30, 37, 40 et 42.
- Calculer la covariance d'un couple de variables aléatoires : exercices 14, 15 et 32.
- Utiliser les inégalités de Markov ou de Bienaymé-Tchebychev : exercices 38 à 40, 45 et 47.

### Quelques résultats classiques :

- Une preuve de la formule du crible (exemple 9).
- Intervalles de confiance (exemples 32 et 33, exercices 38 et 39).
- Espérance et variance d'une loi uniforme sur  $[[m, n]]$  (exercice 25).
- Moments factoriels des lois uniforme et binomiale (exercice 26).
- Premier succès d'une loi binomiale (exercice 27).
- Minimum et maximum de deux variables uniformes indépendantes (exercice 36).
- Inégalité de Jensen pour l'espérance (exercice 43).
- Marche aléatoire dans  $\mathbb{Z}$  (exercice 10).
- Quelques résultats autour des permutations aléatoires (exercices 18 à 21).