

# Sommes de variables aléatoires

Une variable aléatoire décrit une expérience ; en pratique on en répète beaucoup : on lance deux dés et l'on additionne, on assure dix mille clients, on interroge mille personnes. Ce chapitre étudie ce qui se passe quand on **additionne** des variables aléatoires. L'espérance s'additionne toujours ; la variance s'additionne seulement pour des variables **indépendantes**. Appliqué à un **échantillon** de  $n$  répétitions, ce résultat explique pourquoi la moyenne observée se resserre autour de l'espérance quand  $n$  grandit : son écart-type est divisé par  $\sqrt{n}$ . Les **inégalités de Bienaymé-Tchebychev et de concentration** en donnent une version chiffrée, et la **loi des grands nombres** en est la conclusion : la fréquence d'un événement tend vers sa probabilité. C'est le fondement des sondages, des assurances et du contrôle de qualité.

**Prérequis :** Variables aléatoires, La loi binomiale. **Pour s'entraîner :** Exercices : sommes de variables aléatoires.

## Somme de variables aléatoires et linéarité de l'espérance (T<sup>le</sup>)

Soit  $X$  et  $Y$  deux variables aléatoires définies sur un même univers, et  $a, b$  deux réels.

**Définitions :**  $X + Y$  est la variable aléatoire qui à chaque issue associe la somme des valeurs de  $X$  et de  $Y$  ;  $aX + b$  associe  $a$  fois la valeur de  $X$ , plus  $b$ .

**Linéarité de l'espérance :**  $E(X + Y) = E(X) + E(Y)$  et  $E(aX + b) = a E(X) + b$ , **sans aucune hypothèse** d'indépendance.

**Variance d'une transformation affine :**  $V(aX + b) = a^2 V(X)$  et  $\sigma(aX + b) = |a| \sigma(X)$  : ajouter une constante ne disperse pas, multiplier par  $a$  multiplie l'écart-type par  $|a|$ .

**Exemple.** On lance deux dés équilibrés ;  $X$  et  $Y$  sont les résultats et  $S = X + Y$  la somme.  $E(X) = E(Y) = 3,5$ , donc  $E(S) = 7$  sans avoir à écrire la loi de  $S$  (onze valeurs). Le gain  $G = 2X - 3$  a pour espérance  $2 \times 3,5 - 3 = 4$ .

**Exercice 1** *Espérance et variance d'une somme, étape par étape (T<sup>le</sup>)* On lance deux dés équilibrés ;  $X$  et  $Y$  désignent les résultats.

1. Compléter :  $E(X) = \dots$ , donc  $E(X + Y) = \dots + \dots = \dots$
2. Montrer que  $V(X) = \frac{35}{12}$  (on pourra calculer  $E(X^2)$ ).
3. On pose  $G = 2X - 3$ . Calculer  $E(G)$  et  $V(G)$ .
4. On pose  $T = X + 10$ . Comparer  $V(T)$  et  $V(X)$ , puis  $E(T)$  et  $E(X)$ . Interpréter.

## Variance d'une somme de variables indépendantes (T<sup>le</sup>)

**Indépendance :**  $X$  et  $Y$  sont indépendantes si, pour toutes valeurs  $x$  et  $y$ ,  $P(X = x \text{ et } Y = y) = P(X = x) \times P(Y = y)$ . C'est le cas quand elles décrivent deux expériences sans influence l'une sur l'autre (deux dés, deux tirages avec remise).

**Théorème :** si  $X$  et  $Y$  sont **indépendantes**,  $V(X + Y) = V(X) + V(Y)$ , et donc  $\sigma(X + Y) = \sqrt{V(X) + V(Y)}$ .

**Attention :** sans indépendance, la formule est fautive en général :  $V(X + X) = V(2X) = 4V(X) \neq 2V(X)$ . Les écarts-types, eux, ne s'additionnent jamais.

### Exercice 2 Additionner des variances (T<sup>le</sup>)

1. Pour la somme  $S$  de deux dés, calculer  $V(S)$  et  $\sigma(S)$  (arrondi au centième).
2. Soit  $D = X - Y$  la différence des deux dés. Calculer  $E(D)$  et  $V(D)$  (on pourra écrire  $D = X + (-Y)$ ).
3.  $X$  suit  $\mathcal{B}(10 ; 0,3)$  et  $Y$  suit  $\mathcal{B}(5 ; 0,3)$ , indépendantes. Quelle est la loi de  $X + Y$ ? Vérifier la cohérence avec les formules d'espérance et de variance.
4. Donner un exemple de deux variables  $X$  et  $Y$  de même variance non nulle telles que  $V(X + Y) = 0$ . Sont-elles indépendantes?

### Échantillon de taille $n$ , somme et moyenne (T<sup>le</sup>)

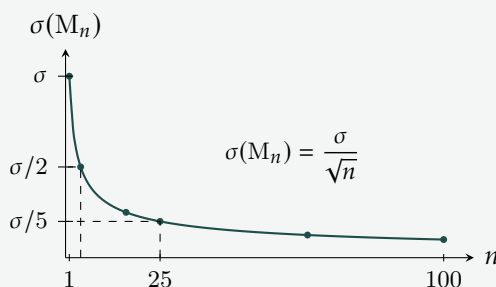
Un **échantillon de taille  $n$**  d'une variable  $X$  est une liste  $X_1, X_2, \dots, X_n$  de variables **indépendantes** suivant toutes la loi de  $X$  (répétition de la même expérience). On note  $\mu = E(X)$  et  $\sigma = \sigma(X)$ .

**Somme :**  $S_n = X_1 + \dots + X_n$  vérifie  $E(S_n) = n\mu$ ,  $V(S_n) = n\sigma^2$  et  $\sigma(S_n) = \sqrt{n}\sigma$ .

**Moyenne :**  $M_n = \frac{S_n}{n}$  vérifie  $E(M_n) = \mu$ ,  $V(M_n) = \frac{\sigma^2}{n}$  et  $\sigma(M_n) = \frac{\sigma}{\sqrt{n}}$ .

**Lecture :** la moyenne d'un échantillon vise juste (espérance  $\mu$ ) et se resserre autour de  $\mu$  quand  $n$  augmente, mais lentement : pour diviser  $\sigma(M_n)$  par 10, il faut multiplier  $n$  par 100.

**Cas de Bernoulli :** si  $X$  suit  $\mathcal{B}(p)$ , alors  $S_n$  suit  $\mathcal{B}(n ; p)$  : on retrouve  $E = np$  et  $V = np(1 - p)$ . La moyenne  $M_n$  est alors la **fréquence** de succès.



**Exercice 3 Somme et moyenne d'un échantillon, étape par étape (T<sup>le</sup>)** À un jeu, le gain  $X$  d'une partie (en euros) a pour espérance 0,5 et pour écart-type 2. Un joueur fait  $n = 100$  parties indépendantes ;  $S_{100}$  est son gain total et  $M_{100}$  son gain moyen par partie.

1. Compléter :  $E(S_{100}) = 100 \times \dots = \dots$  ;  $V(S_{100}) = 100 \times \dots = \dots$  ;  $\sigma(S_{100}) = \dots$
2. Calculer  $E(M_{100})$  et  $\sigma(M_{100})$ .
3. Combien de parties faudrait-il pour que  $\sigma(M_n) \leq 0,05$ ?
4. On note  $X_i$  la variable qui vaut 1 si la partie  $i$  est gagnée (probabilité  $p$ ) et 0 sinon. Quelle est la loi de  $X_1 + \dots + X_n$ ? Retrouver son espérance et sa variance à partir de celles de  $X_1$ .

### Inégalités de concentration et loi des grands nombres (T<sup>le</sup>)

**Inégalité de Bienaymé-Tchebychev :** pour toute variable  $X$  et tout réel  $\delta > 0$ ,

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

Elle majore la probabilité de s'écarter de l'espérance d'au moins  $\delta$ , quelle que soit la loi.

**Inégalité de concentration :** pour la moyenne  $M_n$  d'un échantillon de  $X$ ,

$$P(|M_n - \mu| \geq \delta) \leq \frac{V(X)}{n \delta^2}.$$

**Loi (faible) des grands nombres :** pour tout  $\delta > 0$ ,  $P(|M_n - \mu| \geq \delta)$  tend vers 0 quand  $n$  tend vers  $+\infty$  : la moyenne d'un grand échantillon est, avec une probabilité proche de 1, aussi proche de  $\mu$  que l'on veut. En particulier la fréquence d'un événement tend vers sa probabilité.

**Lecture :** ces inégalités donnent des **majorants**, souvent grossiers ; leur force est de ne rien supposer sur la loi de  $X$ .

#### Méthode : utiliser une inégalité de concentration

1. Identifier  $\mu = E(X)$ ,  $V(X)$  et l'écart  $\delta$  toléré (l'événement doit s'écrire  $|M_n - \mu| \geq \delta$  ou  $|X - E(X)| \geq \delta$ ).
2. Appliquer l'inégalité pour majorer la probabilité de l'écart, ou passer au contraire pour minorer  $P(|M_n - \mu| < \delta)$ .
3. Pour dimensionner un échantillon, résoudre  $\frac{V(X)}{n \delta^2} \leq (\text{seuil voulu})$  en  $n$ .

**Exemple.**  $X$  a pour espérance 10 et pour variance 4. Alors  $P(|X - 10| \geq 4) \leq \frac{4}{16} = 0,25$  : la probabilité que  $X$  sorte de  $]6 ; 14[$  est au plus 0,25, donc  $P(6 < X < 14) \geq 0,75$ .

**Exercice 4** *Appliquer l'inégalité de Bienaymé-Tchebychev ( $T^le$ )* Soit  $X$  une variable aléatoire suivant la loi  $\mathcal{B}(20 ; 0,5)$ .

1. Calculer  $E(X)$  et  $V(X)$ .
2. Majorer  $P(|X - 10| \geq 5)$  à l'aide de l'inégalité de Bienaymé-Tchebychev.
3. À l'aide de la calculatrice, on trouve  $P(X \leq 5) \approx 0,0207$ . En déduire la valeur exacte approchée de  $P(|X - 10| \geq 5)$  et commenter la qualité du majorant.
4. Que donne l'inégalité pour  $\delta = 2$ ? Que peut-on en conclure?

**Exercice 5** *Inégalité de concentration et loi des grands nombres ( $T^le$ )* On lance  $n$  fois une pièce équilibrée et l'on note  $F_n$  la fréquence de « pile » obtenue.

1. Justifier que  $F_n$  est la moyenne d'un échantillon d'une variable  $X$  dont on précisera l'espérance et la variance.
2. Montrer que  $P(|F_n - 0,5| \geq 0,1) \leq \frac{25}{n}$ . Que vaut ce majorant pour  $n = 100$ ? pour  $n = 1\,000$ ?
3. Quelle taille d'échantillon garantit  $P(|F_n - 0,5| \geq 0,1) \leq 0,01$ ?
4. On lance maintenant un dé et  $G_n$  est la fréquence du « 6 ». Quelle taille garantit  $P\left(\left|G_n - \frac{1}{6}\right| \geq 0,05\right) \leq 0,01$ ?
5. Énoncer ce que dit la loi des grands nombres dans ces deux situations.

**Exercice 6** *Problème en contexte : contrôle de qualité ( $T^le$ )* Une machine produit des pièces dont 2 % sont défectueuses, indépendamment les unes des autres. On prélève un lot de 500 pièces ;  $S$  est le nombre de pièces défectueuses et  $F = \frac{S}{500}$  la proportion observée.

1. Quelle est la loi de  $S$ ? Calculer  $E(S)$ ,  $V(S)$  et  $\sigma(S)$  (arrondi au centième).



**Exercice 9** *Vrai ou faux?* Pour chaque affirmation, dire si elle est vraie ou fausse et justifier :

- A.  $E(X + Y) = E(X) + E(Y)$  même si  $X$  et  $Y$  ne sont pas indépendantes.
- B.  $V(X + Y) = V(X) + V(Y)$  pour toutes variables aléatoires  $X$  et  $Y$ .
- C.  $V(2X) = 2V(X)$ .
- D. L'écart-type de la moyenne  $M_n$  d'un échantillon tend vers 0 quand  $n$  tend vers  $+\infty$ .
- E. D'après la loi des grands nombres,  $M_n$  est égale à  $E(X)$  dès que  $n$  est assez grand.
- F. Une variable de loi  $\mathcal{B}(n ; p)$  est une somme de  $n$  variables de Bernoulli indépendantes de paramètre  $p$ .
- G. L'inégalité de Bienaymé-Tchebychev donne la valeur exacte de  $P(|X - E(X)| \geq \delta)$ .
- H. Si  $X$  et  $Y$  sont indépendantes, alors  $\sigma(X + Y) \leq \sigma(X) + \sigma(Y)$ .

*Tu bloques sur un exercice? Consulte la fiche **Que faire quand je bloque?***

*Tu veux retenir durablement ces méthodes? Consulte la fiche **Mémorisation et sciences cognitives.***

*Tu fais souvent les mêmes erreurs? Remplis ton **Carnet d'erreurs.***

## SOLUTIONS DES EXERCICES

### Corrigé de l'exercice 1.

1.  $E(X) = \frac{1+2+3+4+5+6}{6} = 3,5$ , donc  $E(X+Y) = 3,5 + 3,5 = 7$ .
2.  $E(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$ , donc  $V(X) = \frac{91}{6} - 3,5^2 = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12} \approx 2,92$ .
3.  $E(G) = 2 \times 3,5 - 3 = 4$  et  $V(G) = 2^2 \times \frac{35}{12} = \frac{35}{3} \approx 11,67$ .
4.  $V(T) = V(X)$  et  $E(T) = E(X) + 10 = 13,5$  : décaler toutes les valeurs de 10 déplace la moyenne sans changer la dispersion.

### Corrigé de l'exercice 2.

1. Les dés sont indépendants :  $V(S) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6} \approx 5,83$  et  $\sigma(S) \approx 2,42$ .
2.  $E(D) = E(X) - E(Y) = 0$ .  $-Y$  a pour variance  $(-1)^2 V(Y) = V(Y)$  et reste indépendante de  $X$  :  $V(D) = V(X) + V(Y) = \frac{35}{6}$ . La différence est aussi dispersée que la somme.
3.  $X$  compte les succès de 10 épreuves de Bernoulli indépendantes de paramètre 0,3,  $Y$  ceux de 5 autres épreuves indépendantes des premières :  $X+Y$  compte les succès de 15 épreuves indépendantes, donc suit  $\mathcal{B}(15; 0,3)$ . Vérification :  $E(X) + E(Y) = 3 + 1,5 = 4,5 = 15 \times 0,3$  et  $V(X) + V(Y) = 2,1 + 1,05 = 3,15 = 15 \times 0,3 \times 0,7$ . ✓
4.  $Y = -X$  :  $X+Y = 0$ , de variance nulle, alors que  $V(X) + V(Y) = 2V(X) \neq 0$ . Elles ne sont pas indépendantes (connaître  $X$  détermine  $Y$ ), ce qui explique l'échec de la formule.

### Corrigé de l'exercice 3.

1.  $E(S_{100}) = 100 \times 0,5 = 50$ ;  $V(S_{100}) = 100 \times 2^2 = 400$ ;  $\sigma(S_{100}) = 20$ .
2.  $E(M_{100}) = 0,5$  et  $\sigma(M_{100}) = \frac{2}{\sqrt{100}} = 0,2$ .
3.  $\frac{2}{\sqrt{n}} \leq 0,05 \iff \sqrt{n} \geq 40 \iff n \geq 1\,600$  parties.
4. Les  $X_i$  suivent  $\mathcal{B}(p)$ , indépendantes : la somme suit  $\mathcal{B}(n; p)$ . Comme  $E(X_1) = p$  et  $V(X_1) = p(1-p)$ , on obtient  $E = np$  et  $V = np(1-p)$  : les formules de la loi binomiale sont un cas particulier des formules d'échantillon.

### Corrigé de l'exercice 4.

1.  $E(X) = 20 \times 0,5 = 10$  et  $V(X) = 20 \times 0,5 \times 0,5 = 5$ .
2.  $P(|X-10| \geq 5) \leq \frac{5}{25} = 0,2$ .
3.  $|X-10| \geq 5$  signifie  $X \leq 5$  ou  $X \geq 15$  ; par symétrie de la loi  $\mathcal{B}(20; 0,5)$ , ces deux événements ont la même probabilité :  $P \approx 2 \times 0,0207 \approx 0,041$ . Le majorant 0,2 est cinq fois trop grand : l'inégalité est vraie mais grossière, car elle ne tient pas compte de la forme de la loi.
4.  $\frac{5}{4} = 1,25 > 1$  : l'inégalité ne dit rien d'utile (une probabilité est toujours inférieure à 1). Elle n'a d'intérêt que pour  $\delta > \sigma(X)$ .

### Corrigé de l'exercice 5.

1.  $F_n = \frac{X_1 + \dots + X_n}{n}$  où  $X_i$  vaut 1 si le  $i$ -ième lancer donne pile, 0 sinon : les  $X_i$  sont indépendantes de loi  $\mathcal{B}(0,5)$ , avec  $E(X) = 0,5$  et  $V(X) = 0,5 \times 0,5 = 0,25$ .
2. Inégalité de concentration avec  $\delta = 0,1$  :  $P(|F_n - 0,5| \geq 0,1) \leq \frac{0,25}{n \times 0,01} = \frac{25}{n}$ . Pour  $n = 100$  : 0,25 ; pour  $n = 1\,000$  : 0,025.
3.  $\frac{25}{n} \leq 0,01 \iff n \geq 2\,500$  lancers.
4.  $E(X) = \frac{1}{6}$ ,  $V(X) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$ . Il faut  $\frac{5/36}{n \times 0,0025} \leq 0,01$ , soit  $n \geq \frac{5}{36 \times 0,0025 \times 0,01} \approx 5\,556$  lancers.
5. Quand  $n$  devient grand, la probabilité que la fréquence de pile s'écarte de 0,5 de plus de 0,1 (ou de tout écart fixé) tend vers 0 ; de même la fréquence du 6 se rapproche de  $\frac{1}{6}$ . La fréquence observée tend vers la probabilité : c'est ce qui justifie d'estimer une probabilité inconnue par une fréquence sur un grand échantillon.

### Corrigé de l'exercice 6.

1.  $S$  suit  $\mathcal{B}(500 ; 0,02)$  :  $E(S) = 10$ ,  $V(S) = 500 \times 0,02 \times 0,98 = 9,8$ ,  $\sigma(S) \approx 3,13$ .
2.  $S \geq 20$  entraîne  $|S - 10| \geq 10$ , donc  $P(S \geq 20) \leq P(|S - 10| \geq 10) \leq \frac{9,8}{100} = 0,098$ .
3.  $F$  est la moyenne d'un échantillon de taille 500 d'une variable de Bernoulli de variance  $0,02 \times 0,98 = 0,0196$  :  $P(|F - 0,02| \geq 0,01) \leq \frac{0,0196}{500 \times 0,0001} = 0,392$ .
4.  $\frac{0,0196}{n \times 0,0001} \leq 0,05 \iff n \geq \frac{0,0196}{0,05 \times 0,0001} = 3\,920$  pièces.

### Corrigé de l'exercice 7.

1.  $E(S_n) = 200n$ ,  $\sigma(S_n) = 1\,000\sqrt{n}$  et  $\sigma(M_n) = \frac{1\,000}{\sqrt{n}}$  : 100 € pour  $n = 100$ , 10 € pour  $n = 10\,000$ , 1 € pour  $n = 1\,000\,000$ . Le coût moyen par assuré devient presque certain.
2.  $S_n > 230n \iff S_n - 200n > 30n$ , ce qui entraîne  $|S_n - 200n| \geq 30n$ . Par Bienaymé-Tchebychev avec  $\delta = 30n$  :  $P(\text{perte}) \leq \frac{V(S_n)}{(30n)^2} = \frac{10^6 n}{900 n^2} = \frac{10^6}{900 n} \approx \frac{1\,111}{n}$ .
3.  $n = 10\,000$  : au plus 0,11 ;  $n = 100\,000$  : au plus 0,011. Plus le nombre d'assurés est grand, plus le coût moyen se concentre autour de 200 € et plus la prime de 230 € couvre les sinistres avec une probabilité proche de 1 : c'est la mutualisation, application directe de la loi des grands nombres.

### Corrigé de l'exercice 8.

1. **Réponse D.** La linéarité de l'espérance ne demande aucune hypothèse.
  - A :  $10 = 2 \times 5$  serait un produit, sans rapport.
  - B : 3,5 est la moyenne des deux espérances, pas leur somme.
  - C : l'indépendance n'est nécessaire que pour la variance.
2. **Réponse B.**  $X - Y = X + (-Y)$  avec  $V(-Y) = V(Y)$  :  $V(X - Y) = 3 + 1 = 4$ .
  - A : on ne soustrait jamais des variances ; une variance est toujours positive.
  - C : la variance de  $Y$  a été oubliée.
  - D :  $16 = 4^2$  confond variance et carré.

3. **Réponse C.**  $V(M_n) = \frac{\sigma^2}{n}$ , donc  $\sigma(M_n) = \frac{\sigma}{\sqrt{n}}$ .

- A : c'est la variance qui est divisée par  $n$ , pas l'écart-type.
- B : c'est l'écart-type de la somme  $S_n$ .
- D : la moyenne est moins dispersée que  $X$  ; c'est tout l'intérêt d'un échantillon.

4. **Réponse A.**  $\frac{V(X)}{\delta^2} = \frac{2}{4} = 0,5$ .

- B : le majorant 1 est toujours vrai mais n'est pas celui donné par l'inégalité.
- C :  $0,25 = \frac{1}{\delta^2}$  oublie la variance.
- D : une probabilité ne dépasse pas 1 ;  $2 = \frac{V(X)}{\delta}$  oublie le carré.

### Corrigé de l'exercice 9.

**A. Vrai.** C'est la linéarité de l'espérance, valable sans hypothèse.

**B. Faux.** Contre-exemple  $Y = X$  :  $V(2X) = 4V(X) \neq 2V(X)$ . La formule exige l'indépendance.

**C. Faux.**  $V(aX) = a^2V(X)$ , donc  $V(2X) = 4V(X)$ .

**D. Vrai.**  $\sigma(M_n) = \frac{\sigma}{\sqrt{n}} \rightarrow 0$ .

**E. Faux.** La loi des grands nombres affirme que la **probabilité** d'un écart donné tend vers 0, pas que l'écart devient nul :  $M_n$  reste aléatoire.

**F. Vrai.** Chaque épreuve est codée par une Bernoulli, et le nombre de succès en est la somme.

**G. Faux.** Elle fournit un majorant, souvent très supérieur à la vraie valeur (voir la loi  $\mathcal{B}(20 ; 0,5)$  : 0,2 contre 0,04).

**H. Vrai.**  $\sigma(X + Y) = \sqrt{V(X) + V(Y)} \leq \sqrt{V(X)} + \sqrt{V(Y)}$  car  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$  pour  $a, b \geq 0$  (élever au carré :  $a+b \leq a+b+2\sqrt{ab}$ ).