Plaquette 3 – Variance, covariance et corrélation : calculs et structure

Espérance et variance — Prépa MP / MPI (2e année)

Corrections détaillées

Boîte à outils

Variance. $\mathrm{V}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$; $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$; $\mathrm{V}(X) = 0 \Leftrightarrow X$ constante presque sûrement.

Covariance (bilinéaire, symétrique, $\mathrm{Cov}(X,X) = \mathrm{V}(X)$) :

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

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

— Indépendance ⇒ covariance nulle ; la **réciproque est fausse**.

— **Corrélation** : $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} \in [-1,1]$, et $|\rho| = 1$ si et seulement si $Y = aX+b$ presque sûrement.

— **Régression** : $\min_{a,b} \mathbb{E}\left[(Y - aX - b)^2\right] = \mathrm{V}(Y)(1-\rho^2)$, atteint pour $a = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$.

Variance totale. $\mathrm{V}(Y) = \mathbb{E}[\mathrm{V}(Y \mid N)] + \mathrm{V}[\mathbb{E}(Y \mid N)]$.

Correction de l’exercice 1 – Variance d’une loi uniforme discrète

a) Avec $\sum_{k=1}^{n} k = \dfrac{n(n+1)}{2}$ et $\sum_{k=1}^{n} k^2 = \dfrac{n(n+1)(2n+1)}{6}$:

— $\mathbb{E}(X) = \dfrac{n+1}{2}$ et $\mathbb{E}\left(X^2\right) = \dfrac{(n+1)(2n+1)}{6}$.

— **König-Huygens** :

$$\mathrm{V}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12}$$

b) Une telle variable s’écrit $Y = a + h(X-1)$, donc $\mathrm{V}(Y) = h^2\mathrm{V}(X) = h^2\dfrac{n^2-1}{12}$.

c) Les valeurs $10, 15, \ldots, 100$ sont au nombre de $n = 19$, avec $h = 5$:

— $\mathrm{V}(Y) = 25 \times \dfrac{360}{12} = 750$ et $\sigma(Y) = \sqrt{750} \approx 27{,}4$.

— **Contrôle** $n = 6$ (dé) : $\dfrac{35}{12}$ ✓.

Réponse. $\mathrm{V}(X) = \dfrac{n^2-1}{12}$; $h^2\dfrac{n^2-1}{12}$; $\sigma \approx 27{,}4$.

Correction de l’exercice 2 – Covariance nulle sans indépendance

a) Par symétrie de la loi, $\mathbb{E}(X) = 0$ et $\mathbb{E}\left(X^3\right) = 0$. Donc

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left(X^3\right) - \mathbb{E}(X)\,\mathbb{E}\left(X^2\right) = 0$$

b) $\mathbb{P}(X=1, Y=4) = 0$ alors que $\mathbb{P}(X=1)\,\mathbb{P}(Y=4) = \dfrac{1}{4} \times \dfrac{1}{2} = \dfrac{1}{8}$: **pas indépendantes** (et pour cause, Y est une fonction de X).

c) Si $X = 1_A$ et $Y = 1_B$, alors $XY = 1_{A \cap B}$ et

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

$$\mathrm{Cov}(X,Y) = \mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)$$

— Covariance nulle équivaut à l'indépendance de A et B, donc de leurs complémentaires, donc des variables X et Y (quatre égalités sur les quatre valeurs du couple).

— **Leçon** : la covariance ne mesure que la dépendance **affine** ; elle ne caractérise l'indépendance que pour des variables à deux valeurs.

Réponse. Cov = 0 mais dépendance ; pour des Bernoulli, Cov = 0 ⇔ indépendance.

Correction de l'exercice 3 – Covariances de la loi multinomiale

a) $N_1 = \sum_{i=1}^{n} 1_{A_i}$ et $N_6 = \sum_{j=1}^{n} 1_{B_j}$, où A_i (resp. B_j) est l'événement « le lancer i (resp. j) donne 1 (resp. 6) ». Par **bilinéarité** :

— si $i \neq j$, les lancers sont indépendants : $\mathrm{Cov}(1_{A_i}, 1_{B_j}) = 0$;

— si $i = j$, $A_i \cap B_i = \varnothing$: Cov $= 0 - \frac{1}{6}\times\frac{1}{6} = -\frac{1}{36}$.

$$\mathrm{Cov}(N_1, N_6) = -\frac{n}{36}$$

b) $\mathrm{V}(N_1) = \mathrm{V}(N_6) = n\times\frac{1}{6}\times\frac{5}{6} = \frac{5n}{36}$, donc $\rho = \dfrac{-n/36}{5n/36} = -\frac{1}{5}$, indépendamment de n.

c) $\sum_{k=1}^{6} N_k = n$ est constante, donc pour $u = (1,\dots,1)$: $u^{\top}Cu = \mathrm{V}\left(\sum_k N_k\right) = 0$. La matrice C, symétrique positive, a un vecteur isotrope, donc $u \in \ker C$: elle est singulière.

Réponse. $\mathrm{Cov}(N_1,N_6) = -\frac{n}{36}$, $\rho = -\frac{1}{5}$; $(1,\dots,1) \in \ker C$.

Correction de l'exercice 4 – Meilleure approximation affine

a) D'après l'écart quadratique minimal, $\mathbb{E}\left[(Z-b)^2\right]$ est minimal pour $b = \mathbb{E}(Z)$. Avec $Z = Y - aX$: $b = \mathbb{E}(Y) - a\mathbb{E}(X)$, et le minimum vaut $\mathrm{V}(Y-aX)$.

b) On développe :

$$\mathrm{V}(Y-aX) = \mathrm{V}(Y) - 2a\mathrm{Cov}(X,Y) + a^2\mathrm{V}(X)$$

— Trinôme en a, minimal pour $a = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$.

— Minimum : $\mathrm{V}(Y) - \dfrac{\mathrm{Cov}(X,Y)^2}{\mathrm{V}(X)} = \mathrm{V}(Y)\left(1-\rho^2\right)$.

c) $\mathrm{Cov}(D_1, Y) = \mathrm{V}(D_1) = \frac{35}{12}$, donc $a = 1$ et $b = 7 - \frac{7}{2} = \frac{7}{2}$: la meilleure prédiction est $D_1 + 3{,}5$.

— Erreur résiduelle : $\frac{35}{6}\left(1-\frac{1}{2}\right) = \frac{35}{12} = \mathrm{V}(D_2)$ ✓ — ce qu'on ne peut pas prévoir, c'est exactement le second dé.

Réponse. $a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$, $b = \mathbb{E}Y - a\mathbb{E}X$, minimum $\mathrm{V}(Y)(1-\rho^2)$; prédiction $D_1 + 3{,}5$.

Correction de l'exercice 5 – Corrélation du maximum et du minimum de deux dés

a) $Mm = D_1 D_2$ (le produit ne dépend pas de l'ordre), donc par indépendance $\mathbb{E}(Mm) = \left(\frac{7}{2}\right)^2 = \frac{49}{4}$.

b) $\mathbb{P}(M=k) = \frac{2k-1}{36}$, donc $\mathbb{E}\left(M^2\right) = \frac{1}{36}\sum_{k=1}^{6} k^2(2k-1) = \frac{791}{36}$ et

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

$$\mathrm{V}(M) = \frac{791}{36} - \left(\frac{161}{36}\right)^2 = \frac{28\,476 - 25\,921}{1\,296} = \frac{2\,555}{1\,296} \approx 1{,}971$$

— **Symétrie** : $(7 - D_1, 7 - D_2)$ a même loi que (D_1, D_2) et $7 - M$ est le minimum de ce couple, donc m a même loi que $7 - M$, et $\mathrm{V}(m) = \mathrm{V}(M)$.

c) $\mathrm{Cov}(M, m) = \frac{49}{4} - \frac{161 \times 91}{1\,296} = \frac{15\,876 - 14\,651}{1\,296} = \frac{1\,225}{1\,296} \approx 0{,}945$, d'où

$$\rho(M, m) = \frac{1\,225}{2\,555} = \frac{35}{73} \approx 0{,}48$$

— **Interprétation** : un grand minimum force un grand maximum ; la corrélation est positive, mais loin d'une relation affine.

Réponse. $\mathbb{E}(Mm) = \frac{49}{4}$; $\mathrm{V}(M) = \frac{2555}{1296}$; $\rho = \frac{35}{73} \approx 0{,}48$.

Correction de l'exercice 6 – Variance du nombre de descentes

a) $A_i \cap A_{i+1} = \{\sigma(i) > \sigma(i+1) > \sigma(i+2)\}$: un seul des $3! = 6$ ordres relatifs équiprobables, donc

$$\mathrm{Cov}(1_{A_i}, 1_{A_{i+1}}) = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}$$

b) Si $|i - j| \geq 2$, les paires de positions $\{i, i+1\}$ et $\{j, j+1\}$ sont disjointes ; les ordres relatifs dans chaque paire sont indépendants (les 4! ordres des quatre valeurs se répartissent également), donc $\mathbb{P}(A_i \cap A_j) = \frac{1}{4}$ et la covariance est nulle.

c) **Variance d'une somme** : $n - 1$ variances égales à $\frac{1}{4}$ et $n - 2$ paires adjacentes :

$$\mathrm{V}(D) = \frac{n-1}{4} + 2(n-2)\left(-\frac{1}{12}\right) = \frac{n+1}{12}$$

— **Contrôle** $n = 3$: D vaut $0, 1, 2$ avec probabilités $\frac{1}{6}, \frac{4}{6}, \frac{1}{6}$; $\mathbb{E}(D) = 1$, $\mathbb{E}\left(D^2\right) = \frac{8}{6}$, $\mathrm{V}(D) = \frac{1}{3} = \frac{4}{12}$ ✓.

Réponse. $-\frac{1}{12}$ pour des descentes voisines, 0 sinon ; $\mathrm{V}(D) = \frac{n+1}{12}$.

Correction de l'exercice 7 – Variance du nombre de cycles

a) **Linéarité** : $\mathbb{E}(C_n) = \sum_{k=1}^{n} \frac{1}{k} = H_n$. Les B_k étant **indépendantes**, les covariances sont nulles :

$$\mathrm{V}(C_n) = \sum_{k=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) = H_n - \sum_{k=1}^{n} \frac{1}{k^2}$$

b) Pour $n = 4$: $H_4 = \frac{25}{12}$ et $\sum_{k=1}^{4} \frac{1}{k^2} = \frac{205}{144}$, donc $\mathrm{V}(C_4) = \frac{300 - 205}{144} = \frac{95}{144}$.

— **Contrôle direct** : $\mathbb{E}\left(C_4^2\right) = \frac{6 + 44 + 54 + 16}{24} = 5$, et $5 - \left(\frac{25}{12}\right)^2 = \frac{720 - 625}{144} = \frac{95}{144}$ ✓.

c) $H_n \sim \ln n$ et $\sum \frac{1}{k^2} \to \frac{\pi^2}{6}$, donc $\mathrm{V}(C_n) \sim \ln n$.

— **Conséquence** (Tchebychev) : $\frac{C_n}{\ln n} \to 1$ en probabilité — une grande permutation a « environ $\ln n$ » cycles.

Réponse. $\mathrm{V}(C_n) = H_n - \sum_{k \leq n} \frac{1}{k^2} \sim \ln n$; $\frac{95}{144}$ pour $n = 4$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l'exercice 8 – Portefeuille à variance minimale

a) Posons $c = \mathrm{Cov}(X, Y) = \rho\sigma_X\sigma_Y = -0{,}03$. Par bilinéarité :

$$v(t) = t^2\sigma_X^2 + 2t(1-t)c + (1-t)^2\sigma_Y^2$$

b) $v'(t) = 2t\sigma_X^2 + 2(1-2t)c - 2(1-t)\sigma_Y^2 = 0$ donne

$$t^* = \frac{\sigma_Y^2 - c}{\sigma_X^2 + \sigma_Y^2 - 2c} = \frac{0{,}09 + 0{,}03}{0{,}04 + 0{,}09 + 0{,}06} = \frac{0{,}12}{0{,}19} \approx 0{,}632$$

— Le dénominateur vaut $\mathrm{V}(X - Y) > 0$: c'est bien un minimum, et $t^* \in [0, 1]$.

— Valeur : $v(t^*) = \dfrac{\sigma_X^2\sigma_Y^2 - c^2}{\sigma_X^2 + \sigma_Y^2 - 2c} = \dfrac{0{,}0036 - 0{,}0009}{0{,}19} \approx 0{,}0142$, soit un écart type $\approx 0{,}119$.

c) L'actif le moins risqué seul a un écart type de 0,2 : la diversification réduit le risque de 40 % environ, grâce à la corrélation négative.

Réponse. $t^* = \frac{12}{19} \approx 0{,}63$; écart type minimal $\approx 0{,}119$ (contre 0,2 et 0,3).

Correction de l'exercice 9 – Somme d'un nombre aléatoire de termes

a) Sachant $N = n$, S a la loi de $X_1 + \cdots + X_n$ (indépendance de N et des X_i) :
$\mathbb{E}(S \mid N = n) = n\mu$ et $\mathrm{V}(S \mid N = n) = ns^2$.

b) **Espérance totale** : $\mathbb{E}(S) = \mathbb{E}(N\mu) = \lambda\mu$. **Variance totale** :

$$\mathrm{V}(S) = \mathbb{E}\left(Ns^2\right) + \mathrm{V}(N\mu) = \mathbb{E}(N)s^2 + \mu^2\mathrm{V}(N)$$

— Pour une loi de Poisson, $\mathbb{E}(N) = \mathrm{V}(N) = \lambda$: $\mathrm{V}(S) = \lambda\left(s^2 + \mu^2\right) = \lambda\,\mathbb{E}\left(X_1^2\right)$.

c) $\mathbb{E}(S) = 100\,000$ € ; $\mathrm{V}(S) = 50 \times (2\,250\,000 + 4\,000\,000) = 312\,500\,000$, soit un écart type $\approx 17\,678$ €.

— **Lecture** : l'incertitude sur le **nombre** de sinistres ($\mu^2\lambda$) pèse plus que celle sur leur **montant** ($s^2\lambda$).

Réponse. $\mathbb{E}(S) = \lambda\mu$, $\mathrm{V}(S) = \lambda(s^2 + \mu^2)$; 100 000 € et écart type $\approx 17\,678$ €.

Correction de l'exercice 10 – Variance totale : un dé puis des pièces

a) Sachant $N = n$, $X \sim B\left(n, \frac{1}{2}\right)$, donc $\mathbb{E}(X \mid N) = \frac{N}{2}$ et $\mathbb{E}(X) = \frac{\mathbb{E}(N)}{2} = \frac{7}{4} = 1{,}75$.

b) $\mathrm{V}(X \mid N) = \frac{N}{4}$. **Variance totale** :

$$\mathrm{V}(X) = \mathbb{E}\left(\frac{N}{4}\right) + \mathrm{V}\left(\frac{N}{2}\right) = \frac{7/2}{4} + \frac{35/12}{4} = \frac{7}{8} + \frac{35}{48} = \frac{77}{48} \approx 1{,}604$$

c) Une binomiale $B(6, p)$ d'espérance 1,75 a $p = \frac{7}{24}$ et une variance
$6 \times \frac{7}{24} \times \frac{17}{24} = \frac{119}{96} \approx 1{,}240$.

— **Interprétation** : le tirage préalable de N ajoute le terme $\mathrm{V}(\mathbb{E}(X \mid N))$: un mélange est toujours plus dispersé que la moyenne de ses composantes.

Réponse. $\mathbb{E}(X) = \frac{7}{4}$; $\mathrm{V}(X) = \frac{77}{48} \approx 1{,}60$ (contre $\approx 1{,}24$).

Correction de l'exercice 11 – Variance d'un produit de variables indépendantes

a) X^2 et Y^2 sont indépendantes, donc $\mathbb{E}\left(X^2Y^2\right) = \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)$, et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. En notant μ, ν les espérances et a, b les variances :

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

$$\mathrm{V}(XY) = (a+\mu^2)(b+\nu^2) - \mu^2\nu^2 = ab + a\nu^2 + b\mu^2$$

b) $a = b = \frac{35}{12}$ et $\mu = \nu = \frac{7}{2}$:

— $\mathrm{V}(D_1 D_2) = \frac{1\,225}{144} + 2 \times \frac{35}{12} \times \frac{49}{4} = \frac{1\,225 + 10\,290}{144} = \frac{11\,515}{144} \approx 79{,}97$.

— **Contrôle direct** : $\mathbb{E}\left(D^2\right) = \frac{91}{6}$, donc $\mathbb{E}\left(D_1^2 D_2^2\right) = \frac{8\,281}{36}$ et $\mathrm{V}(D_1 D_2) = \frac{8\,281}{36} - \frac{2\,401}{16} = \frac{33\,124 - 21\,609}{144} = \frac{11\,515}{144}$ ✓.

— **Piège** : $\mathrm{V}(XY) \neq \mathrm{V}(X)\mathrm{V}(Y)$, même pour des variables indépendantes ; l'égalité n'a lieu que si les deux espérances sont nulles.

Réponse. $\mathrm{V}(XY) = ab + a\nu^2 + b\mu^2$; $\frac{11\,515}{144} \approx 79{,}97$ pour deux dés.

Correction de l'exercice 12 – Variables échangeables : une covariance bornée

a) **Variance d'une somme** avec $\frac{n(n-1)}{2}$ paires :

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2 + n(n-1)c \geq 0$$

— D'où $c \geq -\dfrac{\sigma^2}{n-1}$: des variables nombreuses ne peuvent pas être toutes fortement corrélées négativement.

b) On tire toutes les boules : $\sum_i A_i = m$ est constante, de variance nulle, donc il y a **égalité** dans a).

— Vérification : $\sigma^2 = \frac{m}{n}\left(1 - \frac{m}{n}\right)$ et

$$c = \frac{m(m-1)}{n(n-1)} - \frac{m^2}{n^2} = -\frac{m(n-m)}{n^2(n-1)} = -\frac{\sigma^2}{n-1}$$

— **Lecture** : c'est la covariance la plus négative possible pour n variables échangeables ; elle explique le facteur de correction $\frac{N-n}{N-1}$ de la loi hypergéométrique.

Réponse. $c \geq -\frac{\sigma^2}{n-1}$, avec égalité pour l'urne vidée sans remise.

Correction de l'exercice 13 – Corrélation égale à 1

a) $\rho = \dfrac{6}{2 \times 3} = 1$.

b) **Bilinéarité** :

$$\mathrm{V}\left(Y - \frac{3}{2}X\right) = \mathrm{V}(Y) - 3\mathrm{Cov}(X,Y) + \frac{9}{4}\mathrm{V}(X) = 9 - 18 + 9 = 0$$

— Une variable de variance nulle est **presque sûrement constante**, égale à son espérance $5 - \frac{3}{2} = \frac{7}{2}$. Donc $Y = \frac{3}{2}X + \frac{7}{2}$ presque sûrement.

— Le coefficient $\frac{3}{2}$ n'est pas deviné : c'est $\frac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}$, la pente de régression.

c) Non : Cauchy-Schwarz impose $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y) = 6$. Avec 7, on trouverait $\mathrm{V}\left(Y - \frac{7}{4}X\right) = 9 - 24{,}5 + 12{,}25 = -3{,}25 < 0$, absurde.

Réponse. $\rho = 1$; $Y = \frac{3}{2}X + \frac{7}{2}$ p.s. ; Cov = 7 impossible (variance négative).

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

Correction de l'exercice 14 – Covariance et indépendance de fonctions

a) $f(X)$ et $g(Y)$ sont indépendantes et bornées, donc $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.

b) On choisit des **indicatrices** $f = 1_{\{x\}}$ et $g = 1_{\{y\}}$:

$$0 = \mathrm{Cov}\left(1_{\{X=x\}}, 1_{\{Y=y\}}\right) = \mathbb{P}(X=x, Y=y) - \mathbb{P}(X=x)\,\mathbb{P}(Y=y)$$

— C'est la définition de l'indépendance de X et Y.

c) $\mathrm{Cov}(X, Y) = \mathbb{E}(X|X|) - 0 = 0$, mais avec $f(x) = x^2$ et $g(y) = y$:
$\mathrm{Cov}\left(X^2, |X|\right) = \mathrm{V}(|X|) = \frac{2}{3} - \frac{4}{9} = \frac{2}{9} \neq 0$.

— **Leçon** : l'indépendance, c'est la covariance nulle de **toutes** les transformées, pas seulement de X et Y.

Réponse. Indépendance $\Leftrightarrow \mathrm{Cov}(f(X), g(Y)) = 0$ pour toutes f, g bornées ; ici $\mathrm{Cov}(X^2, |X|) = \frac{2}{9}$.

Correction de l'exercice 15 – Synthèse : estimer la variance sans biais

a) **König-Huygens** : $\mathbb{E}\left(X_i^2\right) = \sigma^2 + \mu^2$ et, comme $\mathrm{V}\left(\overline{X}\right) = \frac{\sigma^2}{n}$, $\mathbb{E}\left(\overline{X}^2\right) = \frac{\sigma^2}{n} + \mu^2$.

b) On développe : $\sum_i \left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right) = \sum_i X_i^2 - 2n\overline{X}^2 + n\overline{X}^2$.

c) Par linéarité :

$$\mathbb{E}\left[\sum_i \left(X_i - \overline{X}\right)^2\right] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2$$

— Donc $S^2 = \frac{1}{n-1}\sum_i \left(X_i - \overline{X}\right)^2$ vérifie $\mathbb{E}\left(S^2\right) = \sigma^2$: diviser par n **sous-estime** la variance.

d) $n = 2$: $S^2 = \frac{(X_1 - X_2)^2}{2}$ et $\mathbb{E}\left[(X_1 - X_2)^2\right] = \mathrm{V}(X_1 - X_2) = 2\sigma^2$, d'où $\mathbb{E}\left(S^2\right) = \sigma^2 = \frac{35}{12}$ ✓.

— **Pourquoi** $n - 1$ **?** On mesure les écarts à $\overline{X}$, qui est lui-même ajusté aux données : on « perd » un degré de liberté.

Réponse. $\mathbb{E}\left[\sum (X_i - \overline{X})^2\right] = (n-1)\sigma^2$; $S^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).