Plaquette 3 – Variance, covariance et corrélation

Espérance et variance — Prépa MPSI / MP2I (1re année)

Corrections détaillées

Boîte à outils

Variance et covariance.

— $\mathrm{V}(X) = \mathbb{E}((X-\mathbb{E}X)^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$; $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$; $\sigma(X) = \sqrt{\mathrm{V}(X)}$.

— $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$, bilinéaire, symétrique, $\mathrm{Cov}(X,X) = \mathrm{V}(X)$; nulle si X et Y sont indépendantes (réciproque fausse).

— **Variance d'une somme** :

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

— **Corrélation** : $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} \in [-1,1]$ (Cauchy-Schwarz).

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

— Variances usuelles : uniforme $[\![1,n]\!]$: $\dfrac{n^2-1}{12}$; $B(n,p)$: $np(1-p)$; $H(N,m,n)$: $n\dfrac{m}{N}\left(1-\dfrac{m}{N}\right)\dfrac{N-n}{N-1}$.

Correction de l'exercice 1 – Variance d'une loi uniforme

a) Transfert : $\mathbb{E}(X^2) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \dfrac{(n+1)(2n+1)}{6}$.

b) König-Huygens :

$$\mathrm{V}(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(2(2n+1)-3(n+1))}{12} = \frac{(n+1)(n-1)}{12}$$

c) Dé : $\mathrm{V} = \frac{35}{12}$, $\sigma \approx 1{,}71$. Nombre de 1 à 100 : $\mathrm{V} = \frac{9\,999}{12} \approx 833$, $\sigma \approx 28{,}9$ (environ $0{,}29 \times$ l'étendue).

Réponse. $\mathbb{E}(X^2) = \dfrac{(n+1)(2n+1)}{6}$; $\mathrm{V}(X) = \dfrac{n^2-1}{12}$; dé : $\sigma \approx 1{,}71$.

Correction de l'exercice 2 – Variance de la loi binomiale

a) $X = \sum_{i=1}^{n} X_i$, $X_i \sim B(p)$ indépendantes, donc sans covariance :

$$\mathrm{V}(X) = \sum_{i=1}^{n} \mathrm{V}(X_i) = np(1-p)$$

b) $G(t) = (q+pt)^n$ ($q = 1-p$), $G'(1) = np$, $G''(1) = n(n-1)p^2$. Donc $\mathrm{V}(X) = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$ (en utilisant $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X)$).

c) $\mathbb{E} = 500$, $\mathrm{V} = 250$, $\sigma \approx 15{,}8$: le nombre de Piles s'écarte typiquement de 500 d'une quinzaine d'unités.

Réponse. $\mathrm{V}(X) = np(1-p)$; 1 000 lancers : $500 \pm 15{,}8$.

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 3 – Variance de la loi hypergéométrique

a) Par symétrie, les n boules tirées forment une partie uniforme : $\mathbb{P}(A_i) = \dfrac{\binom{N-1}{n-1}}{\binom{N}{n}} = \dfrac{n}{N}$ et $\mathbb{P}(A_i \cap A_j) = \dfrac{\binom{N-2}{n-2}}{\binom{N}{n}} = \dfrac{n(n-1)}{N(N-1)}$.

b) Posons $f = \frac{n}{N}$. $\mathrm{Cov} = \frac{n(n-1)}{N(N-1)} - f^2 = f\left(\frac{n-1}{N-1} - \frac{n}{N}\right) = -f\frac{N-n}{N(N-1)}$ (négative : tirer la blanche i laisse moins de place aux autres). Alors

$$\mathrm{V}(X) = mf(1-f) - m(m-1)f\frac{N-n}{N(N-1)}$$

— En factorisant $mf = n\frac{m}{N}$ et en simplifiant $(1 - f = \frac{N-n}{N})$:
$\mathrm{V}(X) = n\frac{m}{N}\cdot\frac{N-n}{N}\left(1 - \frac{m-1}{N-1}\right) = n\frac{m}{N}\left(1-\frac{m}{N}\right)\frac{N-n}{N-1}$.

c) C’est la variance binomiale multipliée par $\frac{N-n}{N-1} \le 1$: tirer sans remise réduit la dispersion. Pour $n = N$, la variance est nulle : on tire toute l’urne, $X = m$ à coup sûr.

Réponse. $\mathrm{Cov} = -\frac{n}{N}\frac{N-n}{N(N-1)}$; $\mathrm{V}(X) = n\frac{m}{N}(1-\frac{m}{N})\frac{N-n}{N-1}$.

Correction de l’exercice 4 – Variance du nombre de points fixes

a) $\mathbb{P}(A_i \cap A_j) = \dfrac{(n-2)!}{n!} = \dfrac{1}{n(n-1)}$, donc

$$\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

b) $\mathrm{V}(\mathbf{1}_{A_i}) = \frac{1}{n}\left(1-\frac{1}{n}\right)$; il y a $n(n-1)$ couples (i,j) ordonnés distincts :

— $\mathrm{V}(F) = n \times \frac{n-1}{n^2} + n(n-1) \times \frac{1}{n^2(n-1)} = \frac{n-1}{n} + \frac{1}{n} = 1$.

c) $\mathbb{E}(F^2) = \dfrac{0 + 8 + 4\times 6 + 0 + 16 \times 1}{24} = \dfrac{48}{24} = 2$ et $\mathbb{E}(F) = 1$: $\mathrm{V}(F) = 2 - 1 = 1$.

Réponse. $\mathrm{Cov} = \dfrac{1}{n^2(n-1)} > 0$; $\mathrm{V}(F) = 1$ pour tout $n \ge 2$.

Correction de l’exercice 5 – Covariance à partir d’un tableau

a) $\mathbb{E}(X) = 0\times 0{,}3 + 1\times 0{,}3 + 2\times 0{,}4 = 1{,}1$; $\mathbb{E}(Y) = 0{,}4$;
$\mathbb{E}(XY) = 1\times 1\times 0{,}1 + 2\times 1\times 0{,}1 = 0{,}3$.

$$\mathrm{Cov}(X,Y) = 0{,}3 - 1{,}1\times 0{,}4 = -0{,}14$$

b) $\mathbb{E}(X^2) = 0{,}3 + 4\times 0{,}4 = 1{,}9$, $\mathrm{V}(X) = 1{,}9 - 1{,}21 = 0{,}69$; $\mathrm{V}(Y) = 0{,}4\times 0{,}6 = 0{,}24$.
$\rho = \dfrac{-0{,}14}{\sqrt{0{,}69\times 0{,}24}} \approx -0{,}344$: corrélation négative modérée.

c) $\mathrm{V}(X+Y) = 0{,}69 + 0{,}24 + 2\times(-0{,}14) = 0{,}65$. Directement, avec la loi de $S = X+Y$ (0,1 ; 0,4 ; 0,4 ; 0,1) : $\mathbb{E}(S) = 1{,}5$, $\mathbb{E}(S^2) = 0{,}4 + 1{,}6 + 0{,}9 = 2{,}9$, $\mathrm{V}(S) = 2{,}9 - 2{,}25 = 0{,}65$.

Réponse. $\mathrm{Cov} = -0{,}14$; $\rho \approx -0{,}344$; $\mathrm{V}(X+Y) = 0{,}65$.

Correction de l’exercice 6 – Le coefficient de corrélation est entre −1 et 1

a) Par bilinéarité : $\mathrm{V}(X+tY) = \mathrm{V}(X) + 2t\mathrm{Cov}(X,Y) + t^2\mathrm{V}(Y)$, et une variance est positive.

b) Un trinôme de coefficient dominant > 0, positif sur $\mathbb{R}$, a un discriminant négatif ou

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).

nul :

$$4\mathrm{Cov}(X,Y)^2 - 4\mathrm{V}(X)\mathrm{V}(Y) \le 0$$

— d'où $|\rho| \le 1$.

c) $|\rho| = 1$ si et seulement si le discriminant est nul, c'est-à-dire s'il existe t_0 avec $\mathrm{V}(X + t_0 Y) = 0$: $X + t_0 Y$ est presque sûrement constante, donc (avec $t_0 \neq 0$ car $\mathrm{V}(X) \neq 0$) Y est une fonction affine de X.

d) $Y = 7 - X$: $\mathrm{Cov}(X,Y) = -\mathrm{V}(X)$ et $\mathrm{V}(Y) = \mathrm{V}(X)$, donc $\rho = -1$.

Réponse. $\mathrm{Cov}^2 \le \mathrm{V}(X)\mathrm{V}(Y)$; $|\rho| = 1$ ssi relation affine ; d) $\rho = -1$.

Correction de l'exercice 7 – Covariance nulle et indépendance

a) $\mathbb{E}(X) = 0$ et $XY = X^3 = X$, donc $\mathbb{E}(XY) = 0$:

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

b) Non : Y est une fonction de X. $\mathbb{P}(X = 0, Y = 1) = 0$ alors que $\mathbb{P}(X = 0)\mathbb{P}(Y = 1) = \frac{1}{3} \times \frac{2}{3} \neq 0$.

c) Pour $X = \mathbf{1}_A$, $Y = \mathbf{1}_B$: $\mathrm{Cov}(X,Y) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$. Covariance nulle équivaut à l'indépendance des événements A et B, qui entraîne celle des variables (cas binaire : une case suffit).

Réponse. $\mathrm{Cov}(X, X^2) = 0$ mais dépendantes ; pour des Bernoulli, Cov = 0 ⇔ indépendance.

Correction de l'exercice 8 – La moyenne minimise l'écart quadratique

a) On écrit $X - a = (X - m) + (m - a)$ avec $m = \mathbb{E}(X)$ et l'on développe :

$$\mathbb{E}((X-a)^2) = \mathbb{E}((X-m)^2) + 2(m-a)\mathbb{E}(X-m) + (m-a)^2 = \mathrm{V}(X) + (m-a)^2$$

— car $\mathbb{E}(X - m) = 0$.

b) Le second terme est positif, nul si et seulement si $a = m$.

c) $a = 3{,}5$; erreur quadratique moyenne $\frac{35}{12}$, soit une erreur typique $\sigma \approx 1{,}71$. (C'est la version probabiliste de la projection orthogonale sur les constantes.)

Réponse. $\mathbb{E}((X-a)^2) = \mathrm{V}(X) + (\mathbb{E}X - a)^2$; minimum en $\mathbb{E}(X)$; dé : 3,5, erreur $\frac{35}{12}$.

Correction de l'exercice 9 – Diversifier un portefeuille

a) Bilinéarité :

$$v(t) = t^2\sigma_1^2 + (1-t)^2\sigma_2^2 + 2t(1-t)c$$

b) $v'(t) = 2t\sigma_1^2 - 2(1-t)\sigma_2^2 + 2(1-2t)c = 0$ donne

— $t^* = \dfrac{\sigma_2^2 - c}{\sigma_1^2 + \sigma_2^2 - 2c} = \dfrac{0{,}078}{0{,}106} \approx 0{,}736$ (dans $[0,1]$), et
$v(t^*) = \dfrac{\sigma_1^2\sigma_2^2 - c^2}{\sigma_1^2 + \sigma_2^2 - 2c} \approx \dfrac{0{,}003456}{0{,}106} \approx 0{,}0326$.

c) Écart type $\sqrt{0{,}0326} \approx 0{,}181$, inférieur à celui du placement 1 seul (0,2). Mélanger deux placements imparfaitement corrélés ($\rho = \frac{0{,}012}{0{,}06} = 0{,}2$) réduit le risque en dessous de celui du moins risqué : c'est l'effet de **diversification**.

Réponse. $t^* \approx 0{,}736$; $v(t^*) \approx 0{,}0326$; écart type $0{,}181 < 0{,}2$.

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).

Correction de l'exercice 10 – Droite de régression

a) Pour a fixé, la meilleure constante b est l'espérance de $Y - aX$ (exercice précédent), et l'erreur vaut alors $\mathrm{V}(Y-aX) = \mathrm{V}(Y) - 2a\mathrm{Cov}(X,Y) + a^2\mathrm{V}(X)$, trinôme en a minimal pour

$$a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}, \qquad b = \mathbb{E}(Y) - a\mathbb{E}(X)$$

b) $a = \frac{-0{,}14}{0{,}69} \approx -0{,}203$, $b \approx 0{,}4 + 0{,}203 \times 1{,}1 \approx 0{,}623$.

c) En remplaçant : $\mathrm{V}(Y) - \frac{\mathrm{Cov}^2}{\mathrm{V}(X)} = \mathrm{V}(Y)\left(1-\rho^2\right)$. Ici $0{,}24 - \frac{0{,}0196}{0{,}69} \approx 0{,}24 - 0{,}0284 = 0{,}2116$: la régression ne réduit l'erreur que de 12 % ($\rho^2 \approx 0{,}118$).

Réponse. $a \approx -0{,}203$, $b \approx 0{,}623$; erreur $\mathrm{V}(Y)(1-\rho^2) \approx 0{,}212$.

Correction de l'exercice 11 – Variance nulle

a) $\mathrm{V}(X) = \sum_x (x-m)^2 \mathbb{P}(X=x)$ est une somme de termes positifs : elle est nulle si et seulement si $\mathbb{P}(X=x) = 0$ pour tout $x \neq m$, soit $\mathbb{P}(X=m) = 1$.

b) Par indépendance, $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) = 0$: les deux variances, positives, sont nulles.

c) X résultat d'un dé et $Y = 7 - X$: $X + Y = 7$, de variance nulle, alors que $\mathrm{V}(X) = \mathrm{V}(Y) = \frac{35}{12}$ (et $\mathrm{Cov}(X,Y) = -\frac{35}{12}$).

Réponse. $\mathrm{V}(X) = 0 \Leftrightarrow X$ constante p.s. ; contre-exemple dépendant : $Y = 7 - X$.

Correction de l'exercice 12 – Effectifs de deux faces : covariance négative

a) Pour $i \neq j$: dés indépendants, covariance nulle. Pour $i = j$: un dé ne peut pas donner 1 et 6 à la fois,

$$\mathrm{Cov} = 0 - \frac{1}{6} \times \frac{1}{6} = -\frac{1}{36}$$

b) Par bilinéarité, seuls les n termes $i = j$ contribuent : $\mathrm{Cov}(N_1, N_6) = -\frac{n}{36}$. Comme $N_1, N_6 \sim B\left(n, \frac{1}{6}\right)$ ont pour variance $\frac{5n}{36}$:

— $\rho = \frac{-n/36}{5n/36} = -\frac{1}{5}$, indépendamment de n.

c) Plus il y a de 1, moins il reste de dés pour faire des 6 : les effectifs d'un tirage multinomial sont négativement corrélés.

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

Correction de l'exercice 13 – Variance du nombre de boîtes vides

a) $A_i \cap A_j$: les trois boules vont dans la troisième boîte, $\left(\frac{1}{3}\right)^3 = \frac{1}{27}$. Donc

$$\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \frac{1}{27} - \frac{64}{729} = \frac{27-64}{729} = -\frac{37}{729}$$

b) $\mathrm{V}(\mathbf{1}_{A_j}) = \frac{8}{27} \times \frac{19}{27} = \frac{152}{729}$. Avec les 6 couples ordonnés (i,j), $i \neq j$:

— $\mathrm{V}(V) = 3 \times \frac{152}{729} + 6 \times \left(-\frac{37}{729}\right) = \frac{456-222}{729} = \frac{234}{729} = \frac{26}{81}$.

c) $\mathbb{E}(V) = \frac{24}{27}$, $\mathbb{E}(V^2) = \frac{18 + 4 \times 3}{27} = \frac{30}{27}$, $\mathrm{V}(V) = \frac{30}{27} - \frac{576}{729} = \frac{810-576}{729} = \frac{234}{729}$.

Réponse. $\mathrm{Cov} = -\frac{37}{729}$; $\mathrm{V}(V) = \frac{26}{81}$.

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 – Variable centrée réduite

a) Linéarité et $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$:

$$\mathbb{E}(Z) = \frac{\mathbb{E}(X) - m}{\sigma} = 0, \qquad \mathrm{V}(Z) = \frac{\mathrm{V}(X)}{\sigma^2} = 1$$

b) $m = 50$, $\sigma = \sqrt{25} = 5$: $Z = \dfrac{60-50}{5} = 2$. Soixante Piles, c’est « deux écarts types » au-dessus de la moyenne.

c) Scores réduits : $\dfrac{14-11}{2} = 1{,}5$ et $\dfrac{15-12}{4} = 0{,}75$. Le premier élève se distingue davantage de sa classe, bien que sa note brute soit plus faible.

Réponse. $\mathbb{E}(Z) = 0$, $\mathrm{V}(Z) = 1$; $Z = 2$; le premier élève (1,5 contre 0,75).

Correction de l’exercice 15 – Synthèse : variance du nombre de records

a) Indépendance (deux à deux suffit) : les covariances sont nulles, et $\mathrm{V}(I_i) = \frac{1}{i}\left(1 - \frac{1}{i}\right)$. D’où

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

b) $n = 4$: $0 + \frac{1}{4} + \frac{2}{9} + \frac{3}{16} = \frac{36+32+27}{144} = \frac{95}{144}$. Avec la loi : $\mathbb{E}(R^2) = \frac{6+44+54+16}{24} = 5$ et $\mathbb{E}(R) = \frac{25}{12}$, donc $\mathrm{V}(R) = 5 - \frac{625}{144} = \frac{95}{144}$.

c) $\sum_{i \le n} \frac{1}{i^2} \le \frac{\pi^2}{6}$ reste borné tandis que $H_n \sim \ln n \to +\infty$: $\mathrm{V}(R) \sim \ln n$. Espérance et variance sont toutes deux de l’ordre de $\ln n$, et l’écart type $\sqrt{\ln n}$ est petit devant l’espérance : R se concentre autour de $\ln n$.

Réponse. $\mathrm{V}(R) = H_n - \sum \frac{1}{i^2}$; $\frac{95}{144}$ pour $n = 4$; $\mathrm{V}(R) \sim \ln n$.

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).