Covariance et corrélation

Exercices

Exercice 1. On considère n clients se répartissant aléatoirement dans 3 hôtels H_1, H_2, H_3 de manière équiprobable. Les choix des clients sont indépendants. On note X_i le nombre de personnes ayant choisi l'hôtel i.

1. Déterminer les lois de X_1, X_2, X_3 ; leurs espérances et variances.
2. Déterminer loi, espérance et variance de $X_1 + X_2$.
3. Calculer la covariance de X_1, X_2, puis leur coefficient de corrélation linéaire. Le signe de ce coefficient vous paraît-il raisonnable ?

Exercice 2. On considère trois vad A,B,C indépendantes telles que $A \hookrightarrow \mathscr{P}(\lambda)$, $B \hookrightarrow \mathscr{P}(\mu)$, $C \hookrightarrow \mathscr{P}(\lambda)$. On pose $X = A + B$ et $Y = B + C$.

1. Rappeler les lois de X et Y.
2. Montrer que X et Y admettent une covariance, et la calculer.
3. En déduire le coefficient de corrélation linéaire de X et Y.

Exercice 3. Soient $n \geqslant 3$, et $X_1, \ldots, X_n$ n vad mutuellement indépendantes suivant $\mathscr{B}(p)$.
Pour tout $i \in [\![1, n-1]\!]$, on pose $Y_i = X_i X_{i+1}$.

1. Donner la loi de Y_i. Montrer que si $|i - j| > 1$, Y_i et Y_j sont indépendantes.
2. Pour $i \in [\![1, n-2]\!]$, calculer la covariance et le coefficient de corrélation linéaire de (Y_i, Y_{i+1}).

Exercice 4. On considère n joueurs qui tirent à la carabine sur une cible. Chaque joueur dispose de 2 coups, et touche la cible avec une probabilité p.
Les différents joueurs et les différents tirs sont supposés indépendants.

1. On note X_1 le nombre de tireurs qui touchent la cible au premier tir, et X_2 le nombre de tireurs qui touchent au second tir. Donner les lois de X_1, X_2, $X_1 + X_2$.
2. On note A le nombre de tireurs touchant la cible lors de leurs deux essais ; et B le nombre de tireurs touchant la cible sur un seul des deux essais.
   Donner les lois de A et B. Montrer que le couple (A, B) admet une covariance.
3. Exprimer $X_1 + X_2$ en fonction de A et B ; en déduire $\mathrm{Cov}(A, B)$. Interpréter son signe.

Solutions

1

2

3 1. Les X_i valant 0 ou 1, il en est de même pour les Y_i : $Y_i(\Omega) = \{0,1\}$.
De plus $P(Y_i = 1) = P(X_iX_{i+1} = 1) = P((X_i = 1) \cap (X_{i+1} = 1)) = P(X_i = 1)P(X_{i+1} = 1)$ par indépendance des X_i ; donc $P(Y_i) = p^2$ et $Y_i \hookrightarrow \mathscr{B}(p^2)$.
On a $Y_i = X_iX_{i+1}$ et $Y_j = X_jX_{j+1}$. Si $|i-j| > 1$ on a $j < i-1$ ou $j > i+1$: dans les deux cas les indices i, $i+1$, j, $j+1$ sont deux à deux distincts, donc les variables X_i, X_{i+1}, X_j, X_{j+1} sont indépendantes. Par lemme des coalitions, Y_i et Y_j sont indépendantes.

2. Cette fois $Y_i = X_iX_{i+1}$ et $Y_{i+1} = X_{i+1}X_{i+2}$ donc on n'a plus forcément l'indépendance (X_{i+1} apparaît dans les deux expressions).
Par définition de la covariance : $\mathrm{Cov}(Y_i, Y_{i+1}) = E(Y_iY_{i+1}) - E(Y_i)E(Y_{i+1})$.
Or $E(Y_iY_{i+1}) = E(X_i(X_{i+1})^2X_{i+2}) = E(X_iX_{i+1}X_{i+2}) = E(X_i)E(X_{i+1})E(X_{i+2})$; car X_{i+1} étant à valeurs dans $\{0,1\}$ on a $X_{i+1}^2 = X_{i+1}$; puis par indépendance des X_i.
Donc $E(Y_iY_{i+1}) = p^3$ et $\mathrm{Cov}(Y_i, Y_{i+1}) = E(Y_iY_{i+1}) - E(Y_i)E(Y_{i+1}) = p^3 - p^4 = p^3(1-p)$. Avec $V(Y_i) = p^2(1-p^2)$ on trouve finalement

$$\rho(Y_i, Y_{i+1}) = \frac{p^3(1-p)}{\sqrt{p^2(1-p^2)}\sqrt{p^2(1-p^2)}} = \frac{p^3(1-p)}{p^2(1-p)(1+p)} = \frac{p}{1+p}$$

4 1. Avec les arguments habituels, $X_1 \hookrightarrow \mathscr{B}(n,p)$, $X_2 \hookrightarrow \mathscr{B}(n,p)$. Les tirs étant indépendants, X_1 et X_2 sont indépendantes, donc par stabilité $X_1 + X_2 \hookrightarrow \mathscr{B}(2n,p)$.

2. Une expérience est cette fois : un tireur tire 2 coups. A compte le nombre de fois où les 2 coups atteignent la cible : $A \hookrightarrow \mathscr{B}(n,p^2)$; B compte le nombre de fois où un seul des deux coups atteint la cible (événement de probabilité $pq + qp = 2pq$ car on peut avoir succès-échec ou échec-succès) donc $B \hookrightarrow \mathscr{B}(n, 2pq)$.
A et B admettent des variances (lois binomiales) donc (A,B) admet une covariance.

3. On peut compter le nombre de tirs atteignant la cible, qui vaut $X_1 + X_2$ (nb de tirs réussis à la première série + à la seconde série) et qui vaut aussi $2A + B$ (A tireurs touchent 2 fois, et B touchent une fois).
Donc $X_1 + X_2 = 2A + B$.

En prenant les variances : $V(X_1 + X_2) = V(X_1) + V(X_2) = 2npq$ par indépendance de X_1 et X_2 ; donc $V(2A + B) = 2npq$.
Mais $V(2A+B) = 4V(A) + 4\mathrm{Cov}(A,B) + V(B)$. $V(A) = np^2(1-p^2)$ et $V(B) = n(2pq)(1-2pq)$. Finalement

$$\begin{aligned}
\mathrm{Cov}(A,B) &= \frac{1}{4}\left(V(2A+B) - 4V(A) - V(B)\right) \\
&= \frac{1}{4}\left(2npq - 4np^2(1-p^2) - 2npq(1-2pq)\right) \qquad (1-p^2) = (1-p)(1+p) = q(1+p) \\
&= \frac{1}{4}\left(\underbrace{2npq} - 4np^2q(1+p)\underbrace{-2npq} + 4np^2q^2\right) \\
&= \frac{1}{4}4np^2q(-(1+p)+q) \\
&= np^2q(-(1+p)+1-p) \\
\mathrm{Cov}(A,B) &= -2np^3q
\end{aligned}$$