Devoir surveillé n° 6

5 avril 2025

Durée de l'épreuve : 3h

Le devoir comporte deux problèmes indépendants.
La calculatrice n'est pas autorisée. Il sera tenu compte de la rigueur, du soin et de la rédaction dans la notation. Les résultats seront soulignés ou encadrés.

Problème 1

Ce problème débute par l'étude probabiliste de la position d'un virus informatique. Cette étude menée en partie A amène à calculer une moyenne et une variance empirique en partie B. Un changement de modélisation aboutit à la résolution d'une équation matricielle (c'est-à-dire dont l'inconnue est une matrice) en partie C.

Dans tout le problème p désigne un réel appartenant à $]0,1[$ et on note $q = 1 - p$.

Partie A : Une matrice diagonalisable

Un réseau informatique est constitué de deux serveurs notés A et B. À une date initiale, un virus s'introduit dans le serveur A. Au bout de deux semaines, ce virus reste en A avec une probabilité de p ou quitte A pour aller en B avec une probabilité de q. De même, s'il est en B, au bout de deux semaines, il peut y rester avec une probabilité de p ou revenir en A avec une probabilité de q. On admet qu'à chaque nouvelle quinzaine, le virus peut rester sur le même serveur ou le quitter avec les probabilités p et q.

Pour tout $n \in \mathbb{N}$, on note u_n la probabilité de l'évènement A_{2n} : « le virus se trouve en A au bout de $2n$ semaines » et v_n la probabilité de l'évènement B_{2n} : « le virus se trouve en B au bout de $2n$ semaines ».

1. Déterminer une expression de u_{n+1} en fonction de u_n et v_n pour tout $n \in \mathbb{N}$. Justifier avec soin votre réponse.
2. Pour tout $n \in \mathbb{N}$, on note C_n la matrice colonne de $\mathscr{M}_{2,1}(\mathbb{R})$ de coefficients u_n et v_n.
   a) Préciser C_0 et déterminer $M \in \mathscr{M}_2(\mathbb{R})$ telle que :
   $$\forall n \in \mathbb{N}, \quad C_{n+1} = MC_n.$$
   b) Justifier sans calcul que M est diagonalisable puis déterminer les espaces propres et les valeurs propres de M. En déduire qu'il existe une matrice $P \in \mathscr{M}_2(\mathbb{R})$ inversible et une matrice $D \in \mathscr{M}_2(\mathbb{R})$ diagonale dont les coefficients diagonaux sont classés dans l'ordre croissant telles que :
   $$M = PDP^{-1}.$$
   c) Calculer $PP^\top$ (produit de P par sa transposée) et en déduire P^{-1}.
3. a) Déduire de ce qui précède, une expression de $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ en fonction de $n \in \mathbb{N}$.
   b) Ces suites sont-elles convergentes ? Si oui, préciser leurs limites.
4. Quels résultats obtiendrait-on si le virus avait été initialement positionné sur le serveur B ?

Partie B : Moyenne et variance empiriques

Pour tout $i \in \mathbb{N}$, on note X_i la variable aléatoire valant 1 si au bout de $2i$ semaines le virus est sur le serveur A et -1, s'il est sur le serveur B. Par ailleurs, pour tout $n \in \mathbb{N}$, on note M_n la moyenne empirique de $(X_0, \ldots, X_n)$.

5. Calculer l'espérance et la variance de X_i pour tout $i \in \mathbb{N}$.
6. Soit $(i,j) \in \mathbb{N}^2$ tel que $i < j$.
   a) Justifier que :
   $$\mathbb{P}_{[X_i=1]}(X_j = 1) = \mathbb{P}(X_{j-i} = 1).$$
   b) En déduire une expression de $\mathbb{P}([X_j = 1] \cap [X_i = 1])$.
   c) Déterminer de même une expression de $\mathbb{P}([X_j = -1] \cap [X_i = 1])$, $\mathbb{P}([X_j = 1] \cap [X_i = -1])$ et $\mathbb{P}([X_j = -1] \cap [X_i = -1])$.
   d) En déduire que $\mathbb{P}(X_iX_j = 1) = \mathbb{P}(X_{j-i} = 1)$ puis calculer $\mathbb{E}(X_iX_j)$.
7. a) Déterminer l'espérance de M_n.
   b) Démontrer par récurrence que :
   $$\forall n \in \mathbb{N}, \quad \mathbb{E}\left(M_n^2\right) = \frac{1}{n+1} + \frac{2}{(n+1)^2} \sum_{0 \leqslant i < j \leqslant n} \mathbb{E}(X_iX_j).$$
   c) Déduire de ce qui précède une expression de l'espérance de la variance empirique de $(X_0, \ldots, X_n)$.

Partie C : Équation matricielle

On souhaite modéliser la position du virus toutes les semaines plutôt que toutes les quinzaines. Pour tout $n \in \mathbb{N}$, on note donc w_n la probabilité de l'évènement A_n : « le virus se trouve en A au bout de n semaines » et x_n la probabilité de l'évènement B_n : « le virus se trouve en B au bout de n semaines ». Enfin on note D_n la matrice colonne de $\mathscr{M}_{2,1}(\mathbb{R})$ de coefficients w_n et x_n.

8. Quelle relation lie D_{2n} et C_n ?
9. On suppose qu'il existe au moins une matrice $N \in \mathscr{M}_2(\mathbb{R})$ à coefficients positifs ou nuls telle que :
   $$\forall n \in \mathbb{N}, \quad D_{n+1} = ND_n.$$
   a) Démontrer que :
   $$\forall n \in \mathbb{N}, \quad \left(N^2 - M\right) C_n = 0.$$
   b) En déduire que $N^2 = M$.
   c) On pose $\Delta = P^{-1}NP$. Démontrer que $\Delta^2 = D$ et en déduire que Δ est une matrice diagonale et que $p - q$ vérifie une égalité à préciser. Donner alors toutes les matrices Δ solutions de $\Delta^2 = D$.
   d) En déduire enfin qu'il existe au plus deux matrices N solutions du problème.

Problème 2

Ce problème est relatif à l'étude de variables aléatoires réelles qui suivent une loi en lien avec les lois géométriques, binomiales et normales.

Dans la partie A, on étudie deux premières variables aléatoires réelles. Cette étude, généralisée en partie B, permet de montrer que les variables aléatoires étudiées suivent toutes la même loi. Enfin, en partie C, on établit des liens entre cette loi, la loi binomiale et la loi normale.

Ces trois parties ne sont pas indépendantes. Dans tout le problème k désigne un entier naturel, n un entier naturel non nul et p un réel appartenant à $]0,1[$. On note enfin $q = 1 - p$.

Partie A : Autour de la loi géométrique

Une équipe de géologues travaille en zone sismique et installe un sismomètre au sommet d'un volcan. Chaque jour, si ce sismomètre détecte une onde sismique, ce dernier envoie une alerte par satellite à un camp de base situé au pied du volcan.

On note X_1 la variable aléatoire réelle correspondant au nombre de jours sans alerte avant une première alerte, par exemple si cette première alerte survient au troisième jour après l'installation du sismomètre alors X_1 prend la valeur 2.

On admet que ce sismomètre envoie au plus une alerte par jour, et que la probabilité que survienne une alerte est constante et égale à p, indépendante des résultats obtenus les jours précédents.

1. a) Déterminer l'ensemble des valeurs prises par X_1 noté $X_1(\Omega)$, et pour tout $k \in X_1(\Omega)$, calculer $\mathbb{P}(X_1 = k)$.
   b) En déduire que $1 + X_1$ suit la loi géométrique de paramètre p, ce qu'on notera dorénavant $X_1 \hookrightarrow \mathscr{G}^{-1}(p)$.
2. Ce sismomètre étant extrêmement sensible, les géologues décident de mener des analyses complémentaires à partir d'une seconde alerte. Aussi on note X_2 la variable aléatoire réelle correspondant au nombre de jours sans alerte avant cette seconde alerte, depuis l'installation du sismomètre. Par exemple, si une première alerte survient au troisième jour après l'installation de ce sismomètre et une seconde au septième jour alors X_2 prend la valeur 5.
   a) Pour tout $(j, k) \in \mathbb{N}^2$, calculer $\mathbb{P}_{[X_1=j]}(X_2 = k)$
   b) En déduire que : $\forall k \in \mathbb{N}, \quad \mathbb{P}(X_2 = k) = (k+1)q^k p^2$.

Partie B : Généralisation de la situation précédente

On généralise l'étude précédente et pour tout $n \in \mathbb{N}^*$, on note X_n la variable aléatoire réelle correspondant au nombre de jours sans alerte avant une n-ième alerte.

3. a) Démontrer par récurrence sur $k \in \mathbb{N}$ que :

$$\forall k \in \mathbb{N}, \qquad \binom{n+k}{k} = \sum_{j=0}^{k} \binom{n+j-1}{j}.$$

   b) Démontrer par récurrence sur $n \in \mathbb{N}^*$ que :

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(X_n = k) = \binom{n+k-1}{k} q^k p^n.$$

   On le notera dorénavant $X_n \hookrightarrow \mathscr{B}^{-1}(n, p)$.

4. Soit $(X_1^{(i)})_{i \in [\![1,n]\!]}$ une famille de n variables aléatoires réelles mutuellement indépendantes telle que :

$$\forall i \in [\![1,n]\!], \qquad X_1^{(i)} \hookrightarrow \mathscr{G}^{-1}(p).$$

   On cherche à montrer par récurrence sur n que :

$$\sum_{i=1}^{n} X_1^{(i)} \hookrightarrow \mathscr{B}^{-1}(n, p).$$

   a) Vérifier que le résultat est vrai pour $n = 1$ et après avoir posé l'hypothèse de récurrence, justifier que $\displaystyle\sum_{k=1}^{n} X_1^{(k)}$ et $X_1^{(n+1)}$ sont indépendantes.
   b) Conclure.
   c) Si X_n et X_m (avec $m \in \mathbb{N}^*$) sont deux variables aléatoires réelles indépendantes telles que $X_n \hookrightarrow \mathscr{B}^{-1}(n,p)$ et $X_m \hookrightarrow \mathscr{B}^{-1}(m,p)$, que dire (sans démonstration) de $X_n + X_m$?

Partie C : Une loi binomiale et une loi normale

5. Pour tout $m \in \mathbb{N}^*$, on note Y_m la variable aléatoire réelle correspondant au nombre de jours avec alerte pendant les m premiers jours.
   a) Justifier que Y_{n+k} suit une loi binomiale dont on précisera les paramètres.
   b) Démontrer que les événements $(X_n > k)$ et $(Y_{n+k} < n)$ sont égaux.
6. Soient $(a, b) \in \mathbb{R}^2$ tel que $a < b$ et Z_n une variable aléatoire réelle suivant un loi normale de paramètres la moyenne de X_n et la variance de X_n.
   a) Justifier que $Z_n \hookrightarrow \mathscr{N}\left(\dfrac{nq}{p}, \dfrac{nq}{p^2}\right)$.
   b) Énoncer précisément le théorème central limite et en déduire que les deux suites ci-dessous convergent vers une même limite ℓ qu'on donnera sous forme d'une intégrale qu'on ne cherchera pas à calculer :

$$\left(\mathbb{P}\left(a < \frac{\frac{X_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right)\right)_{n \in \mathbb{N}^*}, \qquad \left(\mathbb{P}\left(a < \frac{\frac{Z_n}{n} - \frac{q}{p}}{\frac{\sqrt{q}}{p\sqrt{n}}} < b\right)\right)_{n \in \mathbb{N}^*}.$$