BANQUE COMMUNE D'EPREUVES

**CODE EPREUVE :**
**289**
**HEC_M3_E**

**Concepteur : H.E.C.**

OPTION ECONOMIQUE

# MATHEMATIQUES III

Mercredi 30 avril 2008, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document : **l'utilisation de toute calculatrice et de tout matériel électronique est interdite.**
Seule l'utilisation d'une règle graduée est autorisée.

## EXERCICE

Étant donné un entier $n$ supérieur ou égal à 2, on considère un nuage de $n$ points du plan, c'est-à-dire un $n$-uplet $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ d'éléments de $\mathbb{R}^2$. On suppose que les réels $x_1, x_2, \dots, x_n$ (resp. $y_1, y_2, \dots, y_n$) ne sont pas tous égaux.

On appelle moyenne arithmétique $\overline{x}$ et écart-type $\sigma_x$ du $n$-uplet $x = (x_1, \dots, x_n)$, les réels suivants :

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \quad \text{et} \quad \sigma_x = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

On définit de même la moyenne arithmétique $\overline{y}$ et l'écart-type $\sigma_y$ du $n$-uplet $y = (y_1, \dots, y_n)$.

La covariance $\mathrm{cov}(x, y)$ et le coefficient de corrélation linéaire $r(x, y)$ du couple $(x, y)$ sont donnés par :

$$\mathrm{cov}(x, y) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) \quad \text{et} \quad r(x, y) = \frac{\mathrm{cov}(x, y)}{\sigma_x \times \sigma_y}$$

Soit $f$ la fonction définie sur $\mathbb{R}^2$ à valeurs réelles qui, à tout couple $(a, b)$ de $\mathbb{R}^2$, associe le réel $f(a, b)$ tel que :

$$f(a, b) = \sum_{k=1}^{n}(ax_k + b - y_k)^2$$

1. Justifier que $f$ est de classe $C^2$ sur $\mathbb{R}^2$.

2. a) Écrire le système d'équations $(S)$ permettant de déterminer les points critiques de $f$.

b) Résoudre le système $(S)$. En déduire que $f$ admet un unique point critique $(\hat{a}, \hat{b})$ que l'on exprimera en fonction de $\overline{x}, \overline{y}, \sigma_x^2$ et $\mathrm{cov}(x, y)$.

c) Montrer que ce point critique correspond à un minimum local de $f$.

d) Établir la formule suivante : $f(\hat{a}, \hat{b}) = n\sigma_y^2(1 - r^2(x, y))$

3. a) Montrer que l'on a : $|r(x, y)| \leqslant 1$.

b) Que peut-on dire du nuage de points lorsque $|r(x, y)| = 1$ ?

# PROBLÈME

Dans tout le problème, $N$ désigne un entier naturel fixé supérieur ou égal à 2, et $p$ un réel fixé de l'intervalle $]0,1[$. On pose : $q = 1 - p$. Soit $n$ un entier naturel quelconque.

Dans une population de $N$ individus, on s'intéresse à la propagation d'un certain virus. Chaque jour, on distingue dans cette population trois catégories d'individus : en premier lieu, les individus sains, c'est-à-dire ceux qui ne sont pas porteurs du virus, ensuite les individus qui viennent d'être contaminés et qui sont inoffensifs pour les autres, et enfin, les individus contaminés par le virus et qui sont contagieux.

Ces trois catégories évoluent jour après jour selon le modèle suivant :

- chaque jour $n$, chaque individu sain peut être contaminé par n'importe lequel des individus contagieux ce jour avec la même probabilité $p$, ces contaminations éventuelles étant indépendantes les unes des autres ;
- un individu contaminé le jour $n$ devient contagieux le jour $n+1$ ;
- chaque individu contagieux le jour $n$ redevient sain le jour $n+1$.

On note alors $X_n$ le nombre aléatoire d'individus contagieux le jour $n$.
On remarquera que si, pour un certain entier naturel $i$, on a $X_i = 0$, alors on a aussi $X_{i+1} = 0$.
Les variables aléatoires $X_0, X_1, \dots, X_n, \dots$ sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, et $E(X_n)$ désigne, pour tout $n$ de $\mathbb{N}$, l'espérance de $X_n$.

**Partie I. Un cas particulier**

*Dans cette partie uniquement, on suppose que l'on a : $N = 3$ et $p = 1/3$.*
On considère les matrices $S$ et $R$ suivantes :

$$S = \begin{pmatrix} 9 & 4 & 4 & 9 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad R = \begin{pmatrix} 0 & 1 & -6 & 1 \\ 1 & 0 & 5 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$

L'ensemble $\mathcal{M}_{4,1}(\mathbb{R})$ des matrices colonnes à quatre lignes est confondu avec l'espace vectoriel $\mathbb{R}^4$.

1. Montrer que la matrice $R$ est inversible et calculer son inverse $R^{-1}$.

2. a) Montrer que les réels $-1, 0, 5$ et $9$ sont des valeurs propres de $S$.

b) Calculer le produit matriciel $R^{-1}SR$.

c) En déduire, pour tout $n$ de $\mathbb{N}$, l'expression de la matrice $S^n$ en fonction de $n$ (on pose $S^0 = I$, où $I$ désigne la matrice identité de $\mathcal{M}_4(\mathbb{R})$).

3. Soit $n$ un entier fixé de $\mathbb{N}$.

a) Déterminer la loi de probabilité conditionnelle de $X_{n+1}$ sachant l'événement $[X_n = 0]$.

b) Déterminer la loi de probabilité conditionnelle de $X_{n+1}$ sachant l'événement $[X_n = 3]$.

c) Vérifier que la loi de probabilité conditionnelle de $X_{n+1}$ sachant l'événement $[X_n = 1]$ (resp. $[X_n = 2]$) est la loi binomiale de paramètres $(2, \frac{1}{3})$ (resp. $(1, \frac{5}{9})$).

d) On note $E(X_{n+1}/X_n = i)$ l'espérance de la loi de probabilité conditionnelle de $X_{n+1}$ sachant l'événement $[X_n = i]$. Déterminer les valeurs respectives de $E(X_{n+1}/X_n = 1)$ et $E(X_{n+1}/X_n = 2)$.

4. On suppose, *uniquement dans cette question*, que $X_0$ suit la loi binomiale de paramètres $(3, \frac{1}{3})$.

a) Déterminer la loi de $X_1$ et calculer $E(X_1)$.

b) Vérifier la formule suivante : $E(X_1) = \sum_{i=0}^{3} E(X_1/X_0 = i) \times P([X_0 = i])$.

5. Pour tout entier naturel $n$, on considère le vecteur $U_n$ de $\mathbb{R}^4$ défini par :

$$U_n = \begin{pmatrix} u_n \\ v_n \\ w_n \\ t_n \end{pmatrix} = \begin{pmatrix} P([X_n = 0]) \\ P([X_n = 1]) \\ P([X_n = 2]) \\ P([X_n = 3]) \end{pmatrix}$$

a) Déterminer une relation entre $u_n, v_n, w_n$, et $t_n$.

b) À l'aide de la formule des probabilités totales appliquée au système complet d'événements $([X_n = i])_{0 \leqslant i \leqslant 3}$, déterminer une matrice $M$ de $\mathcal{M}_4(\mathbb{R})$ indépendante de $n$, telle que : $U_{n+1} = MU_n$.

c) Exprimer $M$ en fonction de $S$. En déduire les valeurs propres de $M$.

d) Donner l'expression des réels $u_n$ et $v_n$ en fonction de $n$, $v_0$ et $w_0$.

6. On pose : $F = \bigcup_{n=0}^{+\infty} [X_n = 0]$.

a) Que signifie l'événement $F$ ?

b) Montrer que le virus finit par disparaître presque sûrement, quelle que soit la loi de la variable aléatoire initiale $X_0$.

**Partie II. Le cas général**

On suppose que pour tout entier naturel $n$ et pour tout entier $i$ de $[\![0, N]\!]$, on a : $P([X_n = i]) > 0$. On suppose également que pour tout couple $(i, j)$ de $[\![0, N]\!]^2$, le réel $q_{i,j}$ défini par : $q_{i,j} = P_{[X_n=i]}([X_{n+1} = j])$, est indépendant de $n$.

Soit $Q$ la matrice de $\mathcal{M}_{N+1}(\mathbb{R})$ définie par : $Q = (q_{i,j})_{0 \leqslant i,j \leqslant N}$.

1. a) Déterminer, pour tout $j$ de $[\![0, N]\!]$, les probabilités $q_{0,j}$ et $q_{N,j}$. De même, déterminer pour tout $i$ de $[\![0, N]\!]$, la probabilité $q_{i,N}$.

b) Justifier que si l'on a $j > N - i$, alors $q_{i,j} = 0$.

c) Montrer que pour tout $i$ de $[\![1, N-1]\!]$, la loi de probabilité conditionnelle de $X_{n+1}$ sachant $[X_n = i]$, est une loi binomiale dont on déterminera les paramètres.

2. a) Montrer que 1 est valeur propre de la matrice $Q$.

b) Soit $\lambda$ une valeur propre de $Q$, et $V = \begin{pmatrix} V(0) \\ V(1) \\ \vdots \\ V(N) \end{pmatrix}$ un vecteur propre associé à $\lambda$.

On pose : $|V(i)| = \max_{0 \leqslant j \leqslant N} |V(j)|$. Justifier que la composante $V(i)$ n'est pas nulle, puis, en examinant la ligne $i$ du système $QV = \lambda V$, montrer que l'on a : $|\lambda| \leqslant 1$.

3. On pose pour tout entier naturel $n$ : $U_n = \begin{pmatrix} P([X_n = 0]) \\ P([X_n = 1]) \\ \vdots \\ P([X_n = N]) \end{pmatrix}$.

À l'aide de la formule des probabilités totales, déterminer en fonction de $Q$ une matrice $M$ de $\mathcal{M}_{N+1}(\mathbb{R})$ indépendante de $n$ et vérifiant, pour tout $n$ de $\mathbb{N}$, la relation : $U_{n+1} = MU_n$.

*On suppose jusqu'à la fin de la partie II que la matrice $M$ est diagonalisable, et que $\mathcal{B} = (V_0, \ldots, V_N)$ est une base de vecteurs propres de $M$ telle que, pour tout $k$ de $[\![0, N]\!]$, le vecteur propre $V_k$ est associé à une valeur propre $\lambda_k$.*

*De plus, on suppose que : $\lambda_0 = 1$, $V_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, et que pour tout $k$ de $[\![1, N]\!]$, on a : $|\lambda_k| < 1$.*

4. On décompose alors le vecteur $U_0$ sur la base $\mathcal{B}$ : $U_0 = \sum_{k=0}^{N} \alpha_k V_k$.

a) Déterminer, pour tout $n$ de $\mathbb{N}$, la décomposition du vecteur $U_n$ sur la base $\mathcal{B}$.

b) On note, pour tout couple $(k, i)$ de $[\![0, N]\!]^2$, $V_k(i)$ la $(i+1)$-ième composante du vecteur $V_k$. Exprimer, pour tout $n$ de $\mathbb{N}$ et pour tout $i$ de $[\![0, N]\!]$, la probabilité de l'événement $[X_n = i]$ en fonction des réels $\alpha_k$, $\lambda_k$ et $V_k(i)$ $(k \in [\![0, N]\!])$.

c) Montrer que, pour tout $i$ de $[\![1, N]\!]$, on a : $\lim_{n \to +\infty} P([X_n = i]) = 0$.

d) En déduire que le virus finit par disparaître presque sûrement, quelle que soit la loi de la variable aléatoire initiale $X_0$.

**Partie III. Estimations ponctuelle et par intervalle de confiance de $p$**

On suppose que le paramètre $p$, qui exprime la probabilité qu'un individu contagieux transmette le virus à un individu sain, est inconnu, et on cherche à l'estimer. On rappelle que : $q = 1 - p$.

Pour $m$ entier supérieur ou égal à 1, on considère un $m$-échantillon $(Y_1, Y_2, \ldots, Y_m)$ de variables aléatoires indépendantes, de même loi de Bernoulli de paramètre $p$, définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

On pose : $\overline{Y_m} = \frac{1}{m} \sum_{i=1}^{m} Y_i$.

*Dans toute la suite de cette partie, on note $\varepsilon$ un réel strictement positif quelconque.*

1. a) Montrer que $\overline{Y_m}$ est un estimateur sans biais de $p$ ; déterminer son risque quadratique.

b) À l'aide de l'inégalité de Bienaymé-Tchebycheff, montrer que l'intervalle $\left[\overline{Y_m} - \sqrt{\frac{5}{m}}, \overline{Y_m} + \sqrt{\frac{5}{m}}\right]$ est un intervalle de confiance de $p$ au niveau de confiance 0.95.

2. Soit $\theta$ un réel positif.

a) Établir l'égalité suivante : $P([\overline{Y_m} - p \geqslant \varepsilon]) = P([e^{m\theta \overline{Y_m}} \geqslant e^{m\theta(p+\varepsilon)}])$.

b) Montrer que si $T$ est une variable aléatoire discrète finie à valeurs positives d'espérance $E(T)$, et $a$ un réel strictement positif, on a l'inégalité : $P([T \geqslant a]) \leqslant \frac{E(T)}{a}$.

c) Soit $g$ la fonction définie sur $\mathbb{R}^+$ par : $g(x) = \ln(pe^x + q)$. Déduire des questions précédentes, l'inégalité suivante : $P([\overline{Y_m} - p \geqslant \varepsilon]) \leqslant e^{m(g(\theta) - \theta(p+\varepsilon))}$.

d) Montrer que la fonction $g$ est de classe $C^2$ sur $\mathbb{R}^+$ et vérifie, pour tout $x$ de $\mathbb{R}^+$, l'inégalité : $|g''(x)| \leqslant \frac{1}{4}$.

e) En déduire l'inégalité suivante : $g(\theta) \leqslant \theta p + \frac{\theta^2}{8}$.

f) Étudier les variations de la fonction $h$ définie sur $\mathbb{R}^+$ par : $h(x) = \frac{x^2}{8} - \varepsilon x$.

En déduire l'inégalité : $P([\overline{Y_m} - p \geqslant \varepsilon]) \leqslant e^{-2m\varepsilon^2}$.

3. On pose : $\overline{W_m} = \frac{1}{m} \sum_{i=1}^{m} (1 - Y_i)$. Établir l'inégalité : $P([\overline{W_m} - q \geqslant \varepsilon]) \leqslant e^{-2m\varepsilon^2}$.

4. a) Déduire des questions 2 f) et 3, l'inégalité suivante : $P([|\overline{Y_m} - p| \geqslant \varepsilon]) \leqslant 2e^{-2m\varepsilon^2}$.

b) Sachant que $\ln(0.025) \approx -3.688$, calculer $2e^{-2m\varepsilon^2}$ pour $\varepsilon = \sqrt{\frac{1.844}{m}}$. En déduire un nouvel intervalle de confiance de $p$ au niveau de confiance 0.95. Comparer cet intervalle de confiance à celui obtenu à la question 1.b. Conclure.