



PRÉPASIMPLE

prepasimple.fr

HEC

Mathématiques III — Option économique

2008

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

Exercice

Étant donné un entier $n \geq 2$, 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 $x_1, \dots, x_n$, respectivement $y_1, \dots, y_n$, ne sont pas tous égaux.

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

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k, \quad \sigma_x = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}.$$

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

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

$$\text{Cov}(x, y) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}), \quad r(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}.$$

Soit f la fonction définie sur l'ouvert $\mathbb{R}^2$, à valeurs réelles, qui à tout couple (a, b) associe :

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

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 .
2. 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 $\bar{x}$, $\bar{y}$, σ_x^2 et $\text{Cov}(x, y)$.

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

2. d) Établir la formule suivante :

$$f(\hat{a}, \hat{b}) = n\sigma_y^2(1 - r^2(x, y)).$$

3. a) Montrer que $|r(x, y)| \leq 1$.

3. 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 les individus sains, les individus qui viennent d'être contaminés et qui sont encore inoffensifs pour les autres, et les individus contaminés qui sont contagieux.

Les catégories évoluent selon le modèle suivant. Au début du jour n , les individus sont soit contagieux soit sains. Au cours de ce jour, chaque individu sain peut être contaminé par chacun des individus contagieux avec la même probabilité p , toutes ces transmissions éventuelles étant indépendantes les unes des autres. Un individu contaminé au cours du jour n devient contagieux au début du jour $n + 1$. Chaque individu contagieux le jour n redevient sain au début du jour $n + 1$.

On note 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 définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$, et $E(X_n)$ désigne, pour tout $n \in \mathbb{N}$, l'espérance de X_n .

Les transitions à partir d'un état donné sont celles du modèle, y compris lorsque cet état n'est pas atteint avec une probabilité positive. Une loi conditionnelle sachant $[X_n = i]$

n'est utilisée comme telle que lorsque $P(X_n = i) > 0$; sinon, on étudie la loi de transition à partir de l'état i .

Partie I. Un cas particulier

Dans cette partie uniquement, on suppose que $N = 3$ 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 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 .
2. b) Calculer le produit matriciel $R^{-1}SR$.
2. c) En déduire, pour tout $n \in \mathbb{N}$, l'expression de 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 $\mathbb{N}$.
3. a) Déterminer la loi de probabilité conditionnelle de X_{n+1} sachant $[X_n = 0]$.

3. b) Déterminer la loi de probabilité conditionnelle de X_{n+1} sachant $[X_n = 3]$.
3. c) Vérifier que la loi de probabilité conditionnelle de X_{n+1} sachant $[X_n = 1]$, respectivement $[X_n = 2]$, est la loi binomiale de paramètres $(2, 1/3)$, respectivement $(1, 5/9)$.
3. d) On note $E(X_{n+1}/X_n = i)$ l'espérance de la loi de transition à partir de l'état i , égale à l'espérance de la loi conditionnelle lorsque $P(X_n = i) > 0$. 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 $(3, 1/3)$.
4. a) Déterminer la loi de X_1 et calculer $E(X_1)$.
4. b) Vérifier la formule suivante :

$$E(X_1) = \sum_{i=0}^3 E(X_1/X_0 = i) P(X_0 = i).$$

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

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

Les lois sont ici écrites en colonnes. Si Q désigne la matrice des transitions dont les lignes correspondent aux états de départ, la convention en lignes serait ${}^t(U_{n+1}) = {}^t(U_n)Q$.

5. a) Déterminer une relation entre u_n , v_n , w_n et t_n .
5. b) À l'aide de la formule des probabilités totales appliquée au système complet d'événements $([X_n = i])_{0 \leq i \leq 3}$, déterminer une matrice M de $\mathcal{M}_4(\mathbb{R})$, indépendante de n , telle que $U_{n+1} = MU_n$. Les états de probabilité nulle ont une contribution nulle.
5. c) Exprimer M en fonction de S . En déduire les valeurs propres de M .
5. d) Donner l'expression des réels v_n et w_n en fonction de n , v_0 et w_0 .
6. On pose $F = \bigcup_{n=0}^{+\infty} [X_n = 0]$.
6. a) Que signifie l'événement F ?
6. 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

Pour tout couple $(i, j) \in \{0, \dots, N\}^2$, on note $q_{i,j}$ la probabilité de transition de l'état i vers l'état j donnée par le modèle, indépendante de n . Lorsque $P(X_n = i) > 0$, elle vérifie $q_{i,j} = P(X_{n+1} = j \mid X_n = i)$. On ne suppose pas que tous les états ont une probabilité strictement positive.

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

1. a) Déterminer, pour tout $j \in \{0, \dots, N\}$, les probabilités $q_{0,j}$ et $q_{N,j}$. De même, déterminer, pour tout $i \in \{0, \dots, N\}$, la probabilité $q_{i,N}$.
1. b) Justifier que si $j > N - i$, alors $q_{i,j} = 0$.

1. c) Montrer que pour tout $i \in \{1, \dots, N-1\}$, la loi de transition à partir de l'état i est une loi binomiale dont on déterminera les paramètres ; c'est la loi conditionnelle de X_{n+1} sachant $[X_n = i]$ lorsque cet événement a une probabilité strictement positive.
2. a) Montrer que 1 est valeur propre de Q .
2. b) Soit λ une valeur propre réelle de Q , et $V = {}^t(V(0), \dots, V(N))$ un vecteur propre associé à λ . On pose $|V(i)| = \max_{0 \leq j \leq N} |V(j)|$. Justifier que $V(i) \neq 0$, puis, en examinant la ligne i du système $QV = \lambda V$, montrer que $|\lambda| \leq 1$.
3. On pose, pour tout entier naturel n , $U_n = {}^t(P(X_n = 0), \dots, P(X_n = N))$. À l'aide de la formule des probabilités totales, déterminer en fonction de Q une matrice $M \in \mathcal{M}_{N+1}(\mathbb{R})$, indépendante de n , vérifiant $U_{n+1} = MU_n$ pour tout $n \in \mathbb{N}$.

On suppose jusqu'à la fin de la partie II que M est diagonalisable sur $\mathbb{R}$ et que $\mathcal{B} = (V_0, \dots, V_N)$ est une base de vecteurs propres de M , où V_k est associé à la valeur propre λ_k . On suppose de plus que $\lambda_0 = 1$, que $V_0 = {}^t(1, 0, \dots, 0)$ et que, pour tout $k \in \{1, \dots, N\}$, $|\lambda_k| < 1$.

4. On décompose U_0 dans la base $\mathcal{B}$:

$$U_0 = \sum_{k=0}^N \alpha_k V_k.$$

4. a) Déterminer, pour tout $n \in \mathbb{N}$, la décomposition du vecteur U_n dans la base $\mathcal{B}$.
4. b) On note, pour tout $(k, i) \in \{0, \dots, N\}^2$, $V_k(i)$ la $(i+1)$ -ième composante de V_k . Exprimer, pour tout $n \in \mathbb{N}$ et tout $i \in \{0, \dots, N\}$, la probabilité de $[X_n = i]$ en fonction des réels α_k , λ_k et $V_k(i)$.
4. c) Montrer que, pour tout $i \in \{1, \dots, N\}$, $\lim_{n \rightarrow +\infty} P(X_n = i) = 0$.

4. 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 un entier $m \geq 1$, on considère un m -échantillon $(Y_1, \dots, Y_m)$ de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p , définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

On pose

$$\bar{Y}_m = \frac{1}{m} \sum_{i=1}^m Y_i.$$

Un estimateur T de p est dit sans biais si $E(T) = p$. Son risque quadratique est $E((T - p)^2)$. Un intervalle aléatoire I est un intervalle de confiance de p au niveau 0,95 si, pour tout $p \in]0, 1[$, $P(p \in I) \geq 0,95$.

Dans toute la suite de cette partie, ε est un réel strictement positif quelconque.

1. a) Montrer que $\bar{Y}_m$ est un estimateur sans biais de p ; déterminer son risque quadratique.
1. b) À l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que l'intervalle $[\bar{Y}_m - \sqrt{5/m}, \bar{Y}_m + \sqrt{5/m}]$ est un intervalle de confiance de p au niveau de confiance 0,95.
2. Soit θ un réel strictement positif.

2. a) Établir l'égalité suivante :

$$P(\bar{Y}_m - p \geq \varepsilon) = P(e^{m\theta\bar{Y}_m} \geq e^{m\theta(p+\varepsilon)}).$$

2. 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 $P(T \geq a) \leq E(T)/a$.

2. c) Soit g la fonction définie sur $\mathbb{R}_+$ par $g(x) = \ln(pe^x + q)$. Dédurre des questions précédentes l'inégalité :

$$P(\bar{Y}_m - p \geq \varepsilon) \leq e^{m(g(\theta) - \theta(p+\varepsilon))}.$$

2. d) Montrer que g est de classe C^2 sur $\mathbb{R}_+$ et vérifie, pour tout $x \in \mathbb{R}_+$, $|g''(x)| \leq \frac{1}{4}$.

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

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

$$P(\bar{Y}_m - p \geq \varepsilon) \leq e^{-2m\varepsilon^2}.$$

3. On pose

$$\bar{W}_m = \frac{1}{m} \sum_{i=1}^m (1 - Y_i).$$

Établir l'inégalité $P(\bar{W}_m - q \geq \varepsilon) \leq e^{-2m\varepsilon^2}$.

4. a) Dédurre des questions 2. f) et 3 l'inégalité $P(|\bar{Y}_m - p| \geq \varepsilon) \leq 2e^{-2m\varepsilon^2}$.

4. b) Sachant que $\ln(0,025) \simeq -3,688$, calculer $2e^{-2m\varepsilon^2}$ pour

$\varepsilon = \sqrt{\frac{-\ln(0,025)}{2m}} \simeq \sqrt{\frac{1,844}{m}}$. En déduire un nouvel intervalle de confiance de p au niveau de confiance 0,95. Comparer cet intervalle à celui obtenu à la question 1. b). Conclure.