



PRÉPASIMPLE

prepasimple.fr

HEC

Mathématiques — Option économique

2017

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Sujet

Exercice

Pour tout $n \in \mathbb{N}^*$, on note $\mathcal{M}_n(\mathbb{R})$ l'ensemble des matrices carrées à n lignes et n colonnes à coefficients réels et $\mathcal{B}_n(\mathbb{R})$ l'ensemble des matrices de $\mathcal{M}_n(\mathbb{R})$ dont tous les coefficients sont égaux à 0 ou à 1.

1. Exemple 1. Soit A la matrice de $\mathcal{B}_2(\mathbb{R})$ définie par :

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

1.a) Calculer la matrice A^2 .

1.b) Quelles sont les valeurs propres de A ?

1.c) La matrice A est-elle diagonalisable?

2. Exemple 2. Soit B la matrice de $\mathcal{B}_3(\mathbb{R})$ définie par :

$$B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

On considère les instructions Python et la sortie suivantes. La commande `np.array` construit une matrice à partir de la liste de ses lignes; `np.linalg.inv(P)` calcule l'inverse d'une matrice inversible; `np.dot(M, N)` effectue le produit matriciel de deux matrices de tailles compatibles; `print` affiche son argument. La matrice P ci-dessous est inversible.

Python

```
import numpy as np

B = np.array([[0, 1, 0], [1, 0, 0], [0, 0, 1]])
P = np.array([[1, 1, 0], [1, -1, 0], [0, 0, 1]])
print(np.dot(np.dot(np.linalg.inv(P), B), P))
```

La sortie est la matrice :

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

2.a) Dédurre les valeurs propres de B de la séquence Python précédente.

2.b) Déterminer une base de chacun des sous-espaces propres de B .

3.a) Combien existe-t-il de matrices appartenant à $\mathcal{B}_n(\mathbb{R})$? On pourra choisir successivement leurs coefficients.

3.b) Combien existe-t-il de matrices de $\mathcal{B}_n(\mathbb{R})$ dont chaque ligne et chaque colonne comporte exactement un coefficient égal à 1 ? On pourra choisir successivement la position du coefficient égal à 1 dans chaque ligne.

4. Dans cette question, n est un entier supérieur ou égal à 2. Soit $E = \mathbb{R}^n$ et soit u un endomorphisme de E . On note :

- id l'endomorphisme identité de E ;
- $F = \text{Ker}(u + \text{id})$ et $G = \text{Ker}(u - \text{id})$;
- $p = \dim(F)$ et $q = \dim(G)$.

On suppose que $u \circ u = \text{id}$. Les arguments peuvent être écrits à l'aide de la matrice de u dans la base canonique.

4.a) Justifier que l'image de $u - \text{id}$ est incluse dans F .

4.b) En déduire l'inégalité $p + q \geq n$.

On suppose désormais que $1 \leq p < q$. Soit $(f_1, f_2, \dots, f_p)$ une base de F et $(g_1, g_2, \dots, g_q)$ une base de G .

4.c) Justifier que $(f_1, f_2, \dots, f_p, g_1, g_2, \dots, g_q)$ est une base de E .

4.d) Calculer $u(g_1 - f_1)$ et $u(g_1 + f_1)$.

4.e) Trouver une base de E dans laquelle la matrice de u appartient à $\mathcal{B}_n(\mathbb{R})$.

Problème

Les tables de mortalité sont utilisées en démographie et en actuariat pour prévoir l'espérance de vie des individus d'une population. On s'intéresse dans ce problème à un modèle qui permet d'ajuster la durée de vie à des statistiques portant sur les décès observés au sein d'une génération.

Dans tout le problème, on note :

- a et b deux réels strictement positifs ;
- $(\Omega, \mathcal{A}, P)$ un espace probabilisé sur lequel sont définies toutes les variables aléatoires du problème ;
- $G_{a,b}$ la fonction définie sur $\mathbb{R}_+ = [0, +\infty[$ par $G_{a,b}(x) = \exp\left(-ax - \frac{b}{2}x^2\right)$.

Partie I. Loi exponentielle linéaire

1.a) Montrer que la fonction $G_{a,b}$ réalise une bijection de $\mathbb{R}_+$ sur l'intervalle $]0, 1]$.

1.b) Pour tout réel $y > 0$, résoudre l'équation d'inconnue $x \in \mathbb{R}$: $ax + \frac{b}{2}x^2 = y$.

1.c) On note $G_{a,b}^{-1}$ la bijection réciproque de $G_{a,b}$. Quelle est, pour tout $u \in [0, 1[$, l'expression de $G_{a,b}^{-1}(1 - u)$?

2.a) Justifier la convergence de l'intégrale $\int_0^{+\infty} G_{a,b}(x) dx$.

2.b) Soit f la fonction définie sur $\mathbb{R}$ par :

$$f(x) = \sqrt{\frac{b}{2\pi}} \exp\left(-\frac{b}{2}\left(x + \frac{a}{b}\right)^2\right).$$

Montrer que f est une densité d'une variable aléatoire suivant une loi normale dont on précisera les paramètres : espérance et variance.

2.c) Soit Φ la fonction de répartition de la loi normale centrée réduite. Dédurre de la question 2.b) l'égalité :

$$\int_0^{+\infty} G_{a,b}(x) dx = \sqrt{\frac{2\pi}{b}} \exp\left(\frac{a^2}{2b}\right) \Phi\left(-\frac{a}{\sqrt{b}}\right).$$

3. Pour tout $a > 0$ et tout $b > 0$, on pose :

$$f_{a,b}(x) = \begin{cases} (a + bx) \exp\left(-ax - \frac{b}{2}x^2\right) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

3.a) Justifier que la fonction $f_{a,b}$ est une densité de probabilité.

On dit qu'une variable aléatoire suit la loi exponentielle linéaire de paramètres a et b , notée $\mathcal{E}_\ell(a, b)$, si elle admet $f_{a,b}$ pour densité.

3.b) Soit X une variable aléatoire suivant la loi $\mathcal{E}_\ell(a, b)$. À l'aide d'une intégration par parties, justifier que X admet une espérance $E(X)$ telle que :

$$E(X) = \int_0^{+\infty} G_{a,b}(x) dx.$$

4. Soit Y une variable aléatoire suivant la loi exponentielle de paramètre 1. On pose :

$$X = \frac{-a + \sqrt{a^2 + 2bY}}{b}.$$

4.a) Justifier que, pour tout réel $x \in \mathbb{R}_+$, on a $P([X \geq x]) = G_{a,b}(x)$.

4.b) En déduire que X suit la loi $\mathcal{E}_\ell(a, b)$.

4.c) On note U une variable aléatoire suivant la loi uniforme sur $[0, 1[$. Déterminer la loi de la variable aléatoire $G_{a,b}^{-1}(1 - U)$.

5. La fonction Python suivante génère des simulations de la loi exponentielle linéaire. Ses arguments vérifient $a > 0$, $b > 0$ et $n \in \mathbb{N}^*$. La commande `rd.random(n)` renvoie un tableau de n réalisations indépendantes de la loi uniforme sur $[0, 1[$. Les commandes `np.log` et `np.sqrt` appliquent respectivement le logarithme et la racine carrée à chaque élément d'un tableau. Les opérations arithmétiques sur ces tableaux sont effectuées élément par élément.

Python

```
import numpy as np

import numpy.random as rd

def grandlinexp(a, b, n):
    u = rd.random(n)
    y = ...
    x = (-a + np.sqrt(a**2 + 2*b*y))/b
    return x
```

5.a) Quelle est la signification de la ligne `u = rd.random(n)` ?

5.b) Compléter la ligne `y = ...` pour que la fonction `grandlinexp` génère les simulations désirées.

6. De quel nombre réel peut-on penser que les six valeurs générées par la boucle Python suivante fourniront des valeurs approchées de plus en plus précises et pourquoi ? La commande `np.mean` renvoie la moyenne des éléments d'un tableau ; `range(1, 7)` parcourt les entiers de 1 à 6 ; `print` affiche une valeur.

Python

```
import numpy as np

for k in range(1, 7):
    print(np.mean(grandlinexp(0, 1, 10**k)))
```

Pour cette question, on prolonge la définition de la fonction de simulation au cas $a = 0$, $b = 1$, avec la même formule ; la densité correspondante vaut $x \exp(-x^2/2)$ pour $x \geq 0$ et

0 pour $x < 0$.

Dans la suite du problème, on note $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant chacune la loi exponentielle linéaire $\mathcal{E}_\ell(a, b)$, dont les paramètres $a > 0$ et $b > 0$ sont inconnus.

Soit h un entier supérieur ou égal à 2. On suit pendant une période de h années une « cohorte » de n individus de même âge au début de l'étude et on modélise leurs durées de vie respectives à partir de cette date par les variables aléatoires $X_1, X_2, \dots, X_n$.

Partie II. Premier décès et intervalle de confiance de a

Pour tout $n \in \mathbb{N}^*$, on définit les variables aléatoires M_n , H_n et U_n par :

$$M_n = \min(X_1, X_2, \dots, X_n), \quad H_n = \min(h, X_1, X_2, \dots, X_n), \quad U_n = nH_n.$$

7. Calculer, pour tout $x \in \mathbb{R}_+$, la probabilité $P([M_n \geq x])$. On reconnaîtra la loi de la variable aléatoire M_n .

8. Pour tout $n \in \mathbb{N}^*$, on note F_{U_n} la fonction de répartition de la variable aléatoire U_n .

8.a) Montrer que, pour tout $x \in \mathbb{R}$, on a :

$$F_{U_n}(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - \exp\left(-ax - \frac{b}{2n}x^2\right) & \text{si } 0 \leq x < nh, \\ 1 & \text{si } x \geq nh. \end{cases}$$

8.b) Étudier la continuité de la fonction F_{U_n} .

8.c) La variable aléatoire U_n admet-elle une densité ?

8.d) Montrer que la suite de variables aléatoires $(U_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire dont on précisera la loi.

9. Soit $\alpha \in]0, 1[$.

9.a) Soit Y une variable aléatoire qui suit la loi exponentielle de paramètre 1. Trouver deux réels c et d strictement positifs tels que :

$$P([c \leq Y \leq d]) = 1 - \alpha \quad \text{et} \quad P([Y \leq c]) = \frac{\alpha}{2}.$$

9.b) Montrer que $\left[\frac{c}{U_n}, \frac{d}{U_n} \right]$ est un intervalle de confiance asymptotique de a , de niveau de confiance $1 - \alpha$.

Dans cette question, cette expression signifie que $P\left(\frac{c}{U_n} \leq a \leq \frac{d}{U_n}\right) \rightarrow 1 - \alpha$.

L'intervalle est défini sur l'événement $U_n > 0$, de probabilité 1 ; on peut lui donner une valeur arbitraire sur son complément.

Partie III. Nombre de survivants et estimateur convergent de b

Pour tout $i \in \mathbb{N}^*$, soient S_i et D_i les variables aléatoires telles que :

$$S_i = \begin{cases} 1 & \text{si } X_i \geq h, \\ 0 & \text{sinon,} \end{cases} \quad D_i = \begin{cases} 1 & \text{si } X_i \leq 1, \\ 0 & \text{sinon.} \end{cases}$$

Pour tout $n \in \mathbb{N}^*$, on pose :

$$\bar{S}_n = \frac{1}{n} \sum_{i=1}^n S_i, \quad \bar{D}_n = \frac{1}{n} \sum_{i=1}^n D_i.$$

On dit ici qu'un estimateur T_n de θ est sans biais si $E(T_n) = \theta$. On dit qu'une suite d'estimateurs (T_n) est convergente vers θ si, pour tout $\varepsilon > 0$, $P(|T_n - \theta| \geq \varepsilon) \rightarrow 0$. La valeur obtenue à partir des observations est l'estimation associée à l'estimateur.

10.a) Justifier que, pour tout $i \in \{1, \dots, n\}$, on a $E(S_i) = G_{a,b}(h)$ et calculer $E(S_i D_i)$.

10.b) Pour quels couples $(i, j) \in \{1, \dots, n\}^2$ les variables aléatoires S_i et D_j sont-elles indépendantes ?

10.c) Dédurre des questions précédentes l'expression de la covariance $\text{Cov}(\bar{S}_n, \bar{D}_n)$ de $\bar{S}_n$ et $\bar{D}_n$ en fonction de $G_{a,b}(h)$ et $G_{a,b}(1)$. Le signe de cette covariance était-il prévisible ?

11.a) Montrer que $\bar{S}_n$ est un estimateur sans biais et convergent du paramètre $G_{a,b}(h)$.

11.b) De quel paramètre $\bar{D}_n$ est-il un estimateur sans biais et convergent ?

12. On pose $z(a, b) = \ln(G_{a,b}(1))$ et $r(a, b) = \ln(G_{a,b}(h))$.

Pour tout $n \in \mathbb{N}^*$, on pose :

$$Z_n = \ln \left(1 - \bar{D}_n + \frac{1}{n} \right), \quad R_n = \ln \left(\bar{S}_n + \frac{1}{n} \right).$$

On admet que Z_n et R_n sont des estimateurs convergents de $z(a, b)$ et $r(a, b)$ respectivement, au sens de la définition précédente.

12.a) Soient ε , λ et μ des réels strictement positifs.

12.a.i) Justifier l'inclusion suivante :

$$[|(\lambda Z_n - \mu R_n) - (\lambda z(a, b) - \mu r(a, b))| \geq \varepsilon] \subset [\lambda |Z_n - z(a, b)| + \mu |R_n - r(a, b)| \geq \varepsilon].$$

12.a.ii) En déduire l'inégalité suivante :

$$\begin{aligned} P(|(\lambda Z_n - \mu R_n) - (\lambda z(a, b) - \mu r(a, b))| \geq \varepsilon) &\leq P\left(|Z_n - z(a, b)| \geq \frac{\varepsilon}{2\lambda}\right) \\ &+ P\left(|R_n - r(a, b)| \geq \frac{\varepsilon}{2\mu}\right). \end{aligned}$$

12.b) Pour tout $n \in \mathbb{N}^*$, on pose :

$$B_n = \frac{2}{h-1} Z_n - \frac{2}{h(h-1)} R_n.$$

Montrer que B_n est un estimateur convergent du paramètre b .