ECE 2 - Année 2018-2019
Lycée français de Vienne
Mathématiques - F. Gaunard
http://frederic.gaunard.com

Concours Blanc n°4 - sujet type HEC

Lundi XX Avril 2019
Durée : 4 heures

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 et $\mathcal{B}_n(\mathbb{R})$ l'ensemble des matrices carrées 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 : $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
- (a) Calculer A^2.
- (b) Quelles sont les valeurs propres de A?
- (c) La matrice A est-elle diagonalisable?

(2) *Exemple 2.* Soit B la matrice de $\mathcal{B}(\mathbb{R})$ définie par : $B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

On considère les instructions et la sortie (`-->`) `SciLab` suivantes :

```
B=[0,1,0;1,0,0;0,0,1]
P=[1,1,0;1,-1,0;0,0,1]
inv(P)*B*P
```

```
-->
1.    0.    0.
0.   -1.    0.
0.    0.    1.
```

- (a) Déduire les valeurs propres de B de la séquence `SciLab` précédente.
- (b) Donner une base de chacun des sous-espaces propres de B.

(3) (a) Combien existe-t-il de matrices appartenant à $\mathcal{B}_n(\mathbb{R})$?
- (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?

(4) Dans cette question, $n \geq 2$. Soit E un espace vectoriel de dimension n et u un endomorphisme de E. On note :
- id l'endomorphisme identité de E;
- F le noyau de l'endomorphisme $(u + id)$ et G le noyau de l'endomorphisme $(u - id)$;
- p la dimension de F et q la dimension de G.

On suppose que $u \circ u = id$.
- (a) Justifier que l'image de $(u - id)$ est incluse dans F.
- (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.

(c) Justifier que $(f_1, f_2, \dots, f_p, g_1, g_2, \dots, g_q)$ est une base de E.
(d) Calculer $u(g_1 - f_1)$ et $u(g_1 + f_1)$;
(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 $\mathbb{R}^+$ 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 $]0, 1]$.
(b) Pour tout réel $y > 0$, résoudre l'équation d'inconnue $x \in \mathbb{R}$: $ax + \dfrac{b}{2}\, x^2 = y$.
(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 $\int_0^{+\infty} G_{a,b}(x)\mathrm{d}x$.
(b) Soit f la fonction définie sur $\mathbb{R}$ par

$$f(x) = \sqrt{\frac{b}{2\pi}} \times \exp\left(-\frac{1}{2} b \left(x + \frac{a}{b}\right)^2\right).$$

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

(c) Soit Φ la fonction de répartition de la loi normale centrée réduite. Déduire de la question 2) b, l'égalité

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

(3) Pour tout $a > 0$ et pour tout $b > 0$, on pose

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

(a) Justifier que $f_{a,b}$ 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.

(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 telle que
$$E(X) = \int_0^{+\infty} G_{a,b}(x)\mathrm{d}x.$$

(4) Soit Y une variable aléatoire suivant la loi exponentielle de paramètre 1. On pose
$$X = \frac{-a+\sqrt{a^2+2bY}}{b}.$$

(a) Justifier que pour tout réel $x \in \mathbb{R}^+$, on a $P([X \geq x]) = G_{a,b}(x)$.
(b) En déduire que X suit une loi $\mathcal{E}_\ell(a,b)$.
(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 `SciLab` suivante génère des simulations de la loi exponentielle linéaire.

```
(1)   function x=grandlinexp(a,b,n)
(2)       u=rand(n,1)
(3)       y= ....
(4)       x=(-a+sqrt(a^2+2*b*y))/b
(5)    endfunction
```

(a) Quelle est la signification de la ligne de code (2)?
(b) Compléter la ligne de code (3) 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 `SciLab` suivante fourniront des valeurs approchées de plus en plus précises et pourquoi ?

```
for k=1:6
   mean(grandlinexp(0,1,10^k))
end
```

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 X_1, X_2, ..., 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 :
$$M_n = \min(X_1, X_2, \ldots, X_n), \quad H_n = \min(h, X_1, X_2, \ldots, X_n) \quad \text{et} \quad U_n = nH_n.$$

(7) Calculer pour tout $x \in \mathbb{R}^+$, la probabilité $P([M_n \geq x])$. Reconnaître 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

(a) Montrer que pour tout $x \in R$, on a
$$F_{U_n}(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1-\exp\left(-ax - \dfrac{b}{2n}x^2\right), & \text{si } 0 \leq x < nh \\ 1, & \text{si } x \geq nh \end{cases}.$$

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

(c) La variable aléatoire U_n admet-elle une densité ?
(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 $\alpha \in]0,1[$.
(a) Soit Y une variable aléatoire qui suit la loi exponentielle de paramètre 1.
Trouver deux réels c et d strictement positifs tels que :
$$P([c \leq Y \leq d]) = 1-\alpha \quad \text{et} \quad P([Y \leq c]) = \frac{\alpha}{2}.$$
(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 $1-\alpha$.

Partie III. Nombre de survivant et estimateur convergent de b

Pour tout $i \in \mathbb{N}^*$, soit S_i et D_i les variables aléatoires telles que :
$$S_i = \begin{cases} 1, & \text{si } X_i \geq h \\ 0, & \text{sinon} \end{cases} \quad \text{et} \quad D_i = \begin{cases} 1, & \text{si } X_i \leq 1 \\ 0, & \text{sinon} \end{cases}.$$
Pour tout $n \in \mathbb{N}^*$, on pose :
$$\overline{S_n} = \frac{1}{n}\sum_{i=1}^{n} S_i \quad \text{et} \quad \overline{D_n} = \frac{1}{n}\sum_{i=1}^{n} D_i.$$

(10) (a) Justifier que pour tout $i \in [\![1;n]\!]$, on a $E(S_i) = G_{a,b}(h)$ et calculer $E(S_iD_i)$.
(b) Pour quels couples $(i,j) \in [\![1;n]\!]^2$, les variables S_i et D_j sont-elles indépendantes ?
(c) Déduire des questions précédentes l'expression de la covariance $\text{Cov}(\overline{S_n}, \overline{D_n})$ de $\overline{S_n}$ et $\overline{D_n}$ en fonction de n, $G_{a,b}(h)$ et $G_{a,b}(1)$. Le signe de cette covariance était-il prévisible ?

(11) (a) Montrer que $\overline{S_n}$ est un estimateur sans biais et convergent du paramètre $G_{a,b}(h)$.
(b) De quel paramètre, $\overline{D_n}$ est-il un estimateur sans biais et convergent ?

(12) On pose
$$z(a,b) = \ln(G_{a,b}(1) \qquad \text{et} \qquad r(a,b) = \ln(G_{a,b}(h)).$$
Pour tout $n \in \mathbb{N}^*$, on pose
$$Z_n = \ln\left(1 - \overline{D_n} + \frac{1}{n}\right) \quad \text{et} \quad {}_n = \ln\left(\overline{S_n} + \frac{1}{n}\right).$$
On admet que Z_n et R_n sont des estimateurs convergents de $z(a,b)$ et $r(a,b)$ respectivement.

(a) Soit ε, λ et μ des réels strictement positifs.
(i) Justifier l'inclusion suivante :
$$[\ |(\lambda Z_n - \mu R_n) - (\lambda z(a,b) - \mu r(a,b))| \geq \varepsilon] \subset [\lambda\,|Z_n - z(a,b)| + \mu\,|R_n - r(a,b)| \geq \varepsilon]\,.$$
(ii) En déduire l'inégalité suivante :
$$P\,(\ [\ |(\lambda Z_n - \mu R_n) - (\lambda z(a,b) - \mu r(a,b))| \geq \varepsilon]\ ) \leq P\left(\left[|Z_n - z(a,b)| \geq \frac{\varepsilon}{2\lambda}\right]\right) + P\left(\left[|R_n - r(a,b)| \geq \frac{\varepsilon}{2\mu}\right]\right).$$
(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.