réécrit en Python par Nicolas Damien

Maths hec sujet voie E 2017

Exercice

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

1) *Exemple 1.* Soit A la matrice de $\mathscr{B}_2(\mathbf{R})$ 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 $\mathscr{B}(\mathbf{R})$ définie par : $B = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

On exécute le programme Python ci-dessous :

```
import numpy as np
import numpy.linalg as al
B=np.array(([0,1,0],[1,0,0],[0,0,1]))
P=np.array(([1,1,0],[1,-1,0],[0,0,1]))
Q=np.dot(al.inv(P),B)
D=np.dot(Q,P)
print(D)
```

La console affiche :

```
array([[ 1.,  0.,  0.],
       [ 0., -1.,  0.],
       [ 0.,  0.,  1.]])
```

- **a:** Déduire les valeurs propres de B de la séquence `Python` précédente.
- **b:** Donner une base de chacun des sous-espaces propres de B.

3)
- **a:** Combien existe-t-il de matrices appartenant à $\mathscr{B}_n(\mathbf{R})$?
- **b:** Combien existe-t-il de matrices de $\mathscr{B}_n(\mathbf{R})$ dont chaque ligne et chaque colonne comporte exactement un coefficient égal à 1 ?

4) Dans cette question, n est un entier supérieur ou égal à 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 : $p + q \geqslant n$.
  On suppose désormais que $1 \leqslant p < q$. Soit $(f_1, f_2, \ldots, f_p)$ une base de F et $(g_1, g_2, \ldots, g_q)$ une base de G.
- **c:** Justifier que $(f_1, f_2, \ldots, f_p, g_1, g_2, \ldots, g_q)$ 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 $\mathscr{B}_n(\mathbf{R})$.

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 ;
- $(\Omega, \mathscr{A}, P)$ un espace probabilisé sur lequel sont définies toutes les variables aléatoires du problème ;
- $G_{a,b}$ la fonction définie sur $\mathbf{R}^+$ par : $G_{a,b}(x) = \exp\left(-ax - \dfrac{b}{2}\, x^2\right)$.

Partie I. Loi exponentielle linéaire

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

b: Pour tout réel $y > 0$, résoudre l'équation d'inconnue $x \in \mathbf{R}$: $ax + \dfrac{b}{2}\, x^2 = y$

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

b: Soit f la fonction définie sur $\mathbf{R}$ par : $f(x) = \sqrt{\dfrac{b}{2\pi}} \times \exp\left(-\dfrac{1}{2} b \left(x + \dfrac{a}{b}\right)^2\right)$.
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 : $f_{a,b}(x) = \begin{cases} (a+bx)\exp\left(-ax - \dfrac{b}{2}x^2\right) & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$

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 $\mathscr{E}_\ell(a,b)$ si elle admet $f_{a,b}$ pour densité.

b: Soit X une variable aléatoire suivant la loi $\mathscr{E}_\ell(a,b)$. À l'aide d'une intégration par parties, justifier que X admet une espérance telle que : $E(X) = \displaystyle\int_0^{+\infty} G_{a,b}(x)\, dx$.

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

a: Justifier que pour tout réel $x \in \mathbf{R}^+$, on a : $P([X \geqslant x]) = G_{a,b}(x)$.

b: En déduire que X suit une loi $\mathscr{E}_\ell(a,b)$.

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 `Python` suivante génère n simulations de la loi exponentielle linéaire.

```
import numpy as np
import numpy.random as rd
def linexp(a,b,n):
    u=[rd.random() for k in range(n)]
    y=[.............for k in range(n)]
    x=[(-a+np.sqrt(a**2+2*b*y[k]))/b for k in range(n)]
    return x
```

a: Quelle est la signification de la ligne 4 ?

b: Compléter la ligne 5 pour que la fonction `linexp` génère les simulations désirées.

6) De quel nombre réel peut-on penser que les six valeurs générées par la boucle `Python` ci-dessous fourniront des valeurs approchées de plus en plus précises et pourquoi ?

```
for k in range(6):
    print(np.mean(linexp(0,1,10**k)))
```

Dans la suite du problème, on note $(X_n)_{n\in\mathbf{N}^*}$ une suite de variables aléatoires indépendantes suivant chacune la loi exponentielle linéaire $\mathscr{E}_\ell(a,b)$ 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 $n\in\mathbf{N}^*$, 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 $x\in\mathbf{R}^+$, la probabilité $P([M_n\geqslant x])$. Reconnaître la loi de la variable aléatoire M_n.

8) Pour tout $n\in\mathbf{N}^*$, 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 : $F_{U_n}(x)=\begin{cases}0 & \text{si } x<0\\ 1-\exp\left(-ax-\dfrac{b}{2n}x^2\right) & \text{si } 0\leqslant x<nh\\ 1 & \text{si } x\geqslant nh\end{cases}$

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 $(U_n)_{n\in\mathbf{N}^*}$ 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 :

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

b: Montrer que $\left[\dfrac{c}{U_n},\dfrac{d}{U_n}\right]$ 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 $i\in\mathbf{N}^*$, soit S_i et D_i les variables aléatoires telles que :

$$S_i=\begin{cases}1 & \text{si } X_i\geqslant h\\ 0 & \text{sinon}\end{cases}\quad\text{et}\quad D_i=\begin{cases}1 & \text{si } X_i\leqslant 1\\ 0 & \text{sinon}\end{cases}$$

Pour tout $n\in\mathbf{N}^*$, on pose : $\overline{S_n}=\dfrac{1}{n}\displaystyle\sum_{i=1}^n S_i$ et $\overline{D_n}=\dfrac{1}{n}\displaystyle\sum_{i=1}^n D_i$.

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 $Cov(\overline{S_n},\overline{D_n})$ 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 : $z(a,b) = \ln(G_{a,b}(1)$ et $r(a,b) = \ln(G_{a,b}(h))$.

Pour tout $n \in \mathbf{N}^*$, on pose $Z_n = \ln\left(1 - \overline{D_n} + \dfrac{1}{n}\right)$ et $R_n = \ln\left(\overline{S_n} + \dfrac{1}{n}\right)$.

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 :

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

ii. En déduire l'inégalité suivante :

$$P(\ [\ |(\lambda Z_n - \mu R_n) - (\lambda z(a,b) - \mu r(a,b))| \geqslant \varepsilon]\ ) \leqslant P\left(\left[|Z_n - z(a,b)| \geqslant \frac{\varepsilon}{2\lambda}\right]\right) + P\left(\left[|R_n - r(a,b)| \geqslant \frac{\varepsilon}{2\mu}\right]\right)$$

b: Pour tout $n \in \mathbf{N}^*$, on pose : $B_n = \dfrac{2}{h-1} Z_n - \dfrac{2}{h(h-1)} R_n$.

Montrer que B_n est un estimateur convergent du paramètre b.