Fiche 7 : préparation aux parisiennes

Exercice [HEC 2015]

Soit n un entier supérieur ou égal à 2 et $\mathcal{B} = (e_1, ..., e_n)$ la base canonique de $\mathbb{R}^n$. Soit v un vecteur donné de $\mathbb{R}^n$ de coordonnées $v_1, v_2, ..., v_n$ dans la base $\mathcal{B}$ et qui vérifie

$$\sum_{i=1}^{n} v_i = 1.$$

Soit f l'application définie sur $\mathbb{R}^n$ qui à tout vecteur $x = (x_1, ..., x_n) \in \mathbb{R}^n$ associe le vecteur $f(x)$ défini par :

$$f(x) = x - \left(\sum_{i=1}^{n} x_i \right) v.$$

1. (a) Montrer que f est un endomorphisme de $\mathbb{R}^n$.
   (b) Montrer que $f \circ f = f$.
2. Déterminer le spectre de f.
3. (a) Montrer que le vecteur y appartient à l'image de f, notée $\mathrm{Im}(f)$, si et seulement si $f(y) = y$.
   (b) Montrer que la dimension de $\mathrm{Im}(f)$ est inférieure ou égale à $n - 1$.
   (c) Montrer que pour tout $i \in [\![1, n-1]\!]$, on a $(e_i - e_{i+1}) \in \mathrm{Im}(f)$.
   (d) En déduire une base et la dimension de $\mathrm{Im}(f)$. Quel est le rang de f ?
4. (a) Déterminer une base du noyau de f.
   (b) Quels sont les sous-espaces propres de f ?
   (c) L'endomorphisme f est-il diagonalisable ?
5. Écrire la matrice de l'endomorphisme f dans la base canonique de $\mathbb{R}^n$ et la matrice M' de f dans une base de vecteurs propres.

Problème [HEC 2015]

Dans tout le problème :
• Toutes les variables aléatoires sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$.
• On considère une variable aléatoire X à valeurs strictement positives suivant la loi exponentielle de paramètre λ (avec $\lambda > 0$), de densité f_X définie par :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}.$$

Partie I. Comparaison de deux estimateurs de $\dfrac{1}{\lambda}$.

L'objectif de cette partie est de comparer deux estimateurs sans biais et convergents du paramètre inconnu $\dfrac{1}{\lambda}$.
Pour n entier de $\mathbb{N}^*$, soit $(X_1, X_2, ..., X_n)$ un n-échantillon de variables aléatoires à valeurs strictement positives, indépendantes et de même loi que X. On pose pour tout $n \in \mathbb{N}^*$:

$$Y_n = \sum_{i=1}^{n} X_i, \quad \overline{X_n} = \frac{1}{n} Y_n \text{ et } M_n = \max(X_1, X_2, ..., X_n).$$

1. Rappeler sans démonstration les valeurs de l'espérance $E(X)$, de la variance $V(X)$ ainsi que l'expression de la fonction de répartition F_X de la variable aléatoire X.
2. Calculer $E(\overline{X_n})$ et $V(\overline{X_n})$. En déduire que $\overline{X_n}$ est un estimateur sans biais et convergent du paramètre $\dfrac{1}{\lambda}$.
3. (a) Montrer qu'une densité f_{M_n} de M_n est donnée par :

$$f_{M_n}(x) = \begin{cases} n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}.$$

   (b) Etablir pour tout $n \in \mathbb{N}^*$, l'existence de l'espérance $E(M_n)$ de la variable aléatoire M_n.

   (c) En posant $z = 1 - e^{-\lambda x}$, justifier pour tout $a > 0$, l'égalité :

$$\int_0^a x e^{-\lambda x}(1 - e^{-\lambda x})^{n-1} dx = -\frac{1}{\lambda^2} \int_0^{1-e^{-\lambda a}} z^{n-1} \ln(1 - z) dz.$$

   (d) En déduire que l'on a :

$$E(M_n) = -\frac{n}{\lambda} \int_0^1 z^{n-1} \ln(1 - z) dz.$$

   (e) Montrer que la fonction $z \mapsto (1 - z)(1 - \ln(1 - z))$ définie sur l'intervalle $[0, 1[$, est une primitive de la fonction $z \mapsto \ln(1 - z)$.
   À l'aide d'une intégration par parties, en déduire pour tout $n \in \mathbb{N}^*$ une relation entre $E(M_{n+1})$ et $E(M_n)$.

   (f) On pose pour tout $n \in \mathbb{N}^*$:

$$u_n = \sum_{j=1}^{n} \frac{1}{j}.$$

   Déduire de la question précédente que

$$E(M_n) = \frac{1}{\lambda} u_n.$$

4. On pose pour tout $n \in \mathbb{N}^*$: $M'_n = \dfrac{M_n}{u_n}$ et $v_n = \displaystyle\sum_{j=1}^{n} \frac{1}{j^2}$. *On admet que* $V(M_n) = \dfrac{1}{\lambda^2} v_n$.

   (a) Calculer $E(M'_n)$ et $V(M'_n)$.

   (b) Justifier la convergence de (v_n). Déterminer $\lim\limits_{n \to +\infty} u_n$.

   (c) Déduire des questions 4a et 4b que M'_n est un estimateur sans biais et convergent du paramètre $\dfrac{1}{\lambda}$.

   (d) Soit Q la fonction polynomiale définie sur $\mathbb{R}$ par

$$Q(x) = \sum_{j=1}^{n} \left(x - \frac{1}{j} \right)^2.$$

   À l'aide de l'étude de Q, établir l'inégalité :

$$u_n^2 \leq n v_n.$$

   (e) Comparer alors $V(M'_n)$ et $V(\overline{X_n})$. Conclure.

Partie II. Un exemple.

Les notations et le contexte sont ceux de la partie I.
Dans cette partie, on suppose que la durée de vie (en heures) d'un composant électronique est une variable aléatoire X à valeurs strictement positives suivant la loi exponentielle de paramètre λ (avec $\lambda > 0$).
On suppose qu'en cas de panne, le composant électronique est immédiatement remplacé par un composant neuf dont la durée de vie est indépendante et de même loi que celle des composants précédents.
Pour $j \in \mathbb{N}^*$, on note X_j la variable aléatoire égale à la durée de vie du j-ème composant.
Pour n entier de $\mathbb{N}^*$, on constitue ainsi un échantillon $(X_1, X_n, ...X_n)$ de variables aléatoires à valeurs strictement positives, indépendantes et de même loi que X.
On note $(x_1, x_2, ..., x_n)$ la réalisation du n-échantillon $(X_1, X_2, ..., X_n)$ et on pose

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

5. (a) Donner une interprétation de $\frac{1}{\lambda}$. Dans quelle unité s'exprime $\frac{1}{\lambda}$?
   (b) Soit $p \in]0, 1[$. Déterminer en fonction de p et λ, l'unique réel h_p pour lequel on a $P(X > h_p) = p$.
   (c) Proposer un estimateur H_n sans biais et convergent pour le paramètre h_p.
   (d) On suppose que sur un échantillon de 100 composants, on a obtenu :

   $$\sum_{i=1}^{100} x_i = 10^5 \text{ heures.}$$

   Donner une estimation de $h_{\frac{1}{2}}$ (on donne $\ln(2) \simeq 0,7$).

6. *On admet sans démonstration* que pour tout $n \in \mathbb{N}^*$, la fonction de répartition F_{Y_n} de la variable aléatoire Y_n est donnée par :

$$F_{Y_n}(x) = \begin{cases} 1 - e^{-\lambda x}\displaystyle\sum_{k=0}^{n-1}\frac{(\lambda x)^k}{k!} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}.$$

   Soit $t > 0$ un réel fixé et $N(t)$ la variable aléatoire égale au nombre de pannes dans l'intervalle $[0, t[$.
   (a) Déterminer la loi de la variable aléatoire $N(t)$.
   (b) Donner une estimation "naturelle" du nombre moyen de pannes dans l'intervalle $[0, t[$.

7. L'objectif de cette question est de déterminer un intervalle de confiance asymptotique du paramètre inconnu $\frac{1}{\lambda}$ au niveau de confiance $1 - \alpha$ $(0 < \alpha < 1)$.
   (a) Soit Φ la fonction de répartition de la loi normale centrée réduite et soit t_1, t_2 deux réels vérifiant

   $$\Phi(t_1) = 1 - \frac{\alpha}{2} \quad \text{et} \quad \Phi(t_2) = \frac{\alpha}{2}.$$

   Montrer que $t_2 = -t_1$.
   (b) On pose :

   $$R_n = \sqrt{n}\,\left(\lambda\overline{X_n} - 1\right).$$

   Justifier que la suite de variables aléatoires $(R_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.
   (c) En déduire que $\lim\limits_{n\to+\infty} P([-t_1 \leq R_n \leq t_1]) = 1 - \alpha$.
   (d) Déterminer un intervalle de confiance asymptotique de $\frac{1}{\lambda}$ au niveau de confiance $1 - \alpha$.
   (e) Que se passe-t-il lorsque α est proche de 0 ou lorsque α est proche de 1 ?
   (f) Pour $\alpha = 0,05$, on donne $t_1 \simeq 2$. Montrer qu'avec l'échantillon de la question 5d, la réalisation de l'intervalle de confiance asymptotique de $\frac{1}{\lambda}$ au niveau de confiance 0,95 est : $[833, 1250]$.

Partie III. Un résultat asymptotique.

Les notations et le contexte sont ceux des parties précédentes.
Pour tout $n \in \mathbb{N}^*$, on pose

$$T_n = M_n - \frac{1}{\lambda}\ln n.$$

On note F_{T_n} la fonction de répartition de T_n.

8. (a) Montrer que pour tout x réel, on a :

$$F_{T_n}(x) = \left(F_X\left(x + \frac{1}{\lambda}\ln n\right)\right)^n.$$

   (b) Pour chaque réel x fixé, déterminer un entier naturel N_x (qui dépend de x) tel que pour tout entier $n \geq N_x$, on a :

$$x + \frac{1}{\lambda}\ln n > 0.$$

   En déduire que pour tout entier $n \geq N_x$, on a :

$$F_X\left(x + \frac{1}{\lambda}\ln n\right) = 1 - \frac{e^{-\lambda x}}{n}.$$

   (c) Montrer que pour tout x réel, on a :

$$\lim_{n\to+\infty} F_{T_n}(x) = \exp(-e^{-\lambda x}) = e^{-e^{-\lambda x}}.$$

9. Soit F la fonction définie sur $\mathbb{R}$ à valeurs réelles telle que :

$$F(x) = \exp(-e^{-\lambda x}).$$

   (a) Justifier que F est de classe C^∞ sur $\mathbb{R}$ et montrer que F réalise une bijection de $\mathbb{R}$ sur l'intervalle $]0, 1[$.
   (b) En déduire que F est la fonction de répartition d'une variable aléatoire T admettant une densité f_T continue sur $\mathbb{R}$ que l'on déterminera : on dit que T *suit la loi de Gumbel de paramètre* λ.
   (c) Justifier la convergence en loi de la suite de variables aléatoires $(T_n)_{n\in\mathbb{N}^*}$ vers la variable aléatoire T.

10. (a) Etablir l'existence de l'espérance $E(T)$ de la variable aléatoire T.
   (b) On pose :

$$Z = -\frac{1}{\lambda}\ln(\lambda X).$$

   Montrer que les variables aléatoires Z et T ont la même loi.
   (c) Justifier l'égalité :

$$E(T) = -\frac{1}{\lambda}\int_0^{+\infty} e^{-t}\ln t\,dt.$$

   (d) A l'aide de la concavité de la fonction ln sur $\mathbb{R}_+^*$, établir l'inégalité :

$$E(T) \geq 0.$$

11. On suppose dans cette question que $\lambda = 1$.
   (a) Expliciter la bijection réciproque G de la fonction F.
   (b) On considère le programme *Scilab* suivant :
   `x=linspace(-2,2,400); y=(exp(-exp(-x))); plot(x,y), plot(y,x).`
      i. Le réel 0 fait-il partie des nombres renvoyés par la commande `x=linspace(-2,2,400)` ?
      ii. Quel sera le résultat de l'exécution de ce programme ?
   (c) Soit U une variable suivant la loi uniforme sur l'intervalle $]0, 1[$.
   Quelle est la loi de la variable aléatoire $G(U)$?
   (d) Établir l'inégalité :

$$E(T) \leq 1.$$

   (e) Par une méthode de votre choix, écrire en *Scilab* les commandes qui permettent de simuler la loi de T.
   (f) Écrire en *Scilab* les commandes qui permettent de renvoyer une valeur numérique approchée de $E(T)$ en utilisant la méthode de Monte-Carlo.