Convergence/approximation et estimation

I. Intervalles de confiance

Exercice 1 : HEC 2017

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

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

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

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

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 $x \in \mathbb{R}_+$, on a : $\mathbb{P}([X \geqslant x]) = G_{a,b}(x)$.

b) En déduire que X suit la 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 aléatoires $X_1, X_2, \ldots, X_n$.

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 $\mathbb{P}([M_n \geqslant x])$.
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 $x \in \mathbb{R}$, 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 $(U_n)_{n\in\mathbb{N}^*}$ 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 :

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

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 survivants et estimateur convergent de b

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

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

Pour tout $n \in \mathbb{N}^*$, on pose : $\overline{S}_n = \dfrac{1}{n} \sum\limits_{i=1}^{n} S_i$ et $\overline{D}_n = \dfrac{1}{n} \sum\limits_{i=1}^{n} D_i$.

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

b) Pour quels couples $(i, j) \in [\![1, n]\!]^2$, les variables aléatoires S_i et D_j sont-elles indépendantes ?

c) Déduire des questions précédentes l'expression de la covariance $\mathrm{Cov}(\overline{S}_n, \overline{D}_n)$ 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 $n \in \mathbb{N}^*$, 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 :

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

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

Exercice 2 : `adapté des oraux ESCP 2005`

Toutes les variables aléatoires considérées dans cet exercice sont définies sur le même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. On considère une suite de variables aléatoires réelles $X_1, X_2, \ldots, X_n, \ldots$ indépendantes, de même loi, à valeurs dans $[0, \theta]$, ayant pour fonction de densité :

$$f_\theta(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{2x}{\theta^2} & \text{si } 0 \leqslant x \leqslant \theta \\ 0 & \text{si } x > \theta \end{cases}$$

où θ est un paramètre réel strictement positif.
On note F_θ la fonction de répartition commune aux variables $X_1, X_2, \ldots, X_n, \ldots$.

1. **Étude d'un premier estimateur de θ**

a) Pour tout entier naturel non nul n, on considère la variable aléatoire $M_n = \sup(X_1, \dots, X_n)$. Déterminer une densité de M_n.

b) Calculer les moments d'ordre 1 et 2 de M_n.

c) Montrer que M_n est un estimateur biaisé de θ. Est-il asymptotiquement sans biais ?

d) Montrer que la suite $(M_n)_n$ est une suite d'estimateurs convergente de θ.

2. **Étude d'un second estimateur de θ**

a) Donner l'espérance et la variance de la variable $Y_n = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$.

b) À l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que la suite $(Y_n)_{n \in \mathbb{N}^*}$ est une suite d'estimateurs convergents de $\dfrac{2\theta}{3}$.

c) Déduire de la variable Y_n un estimateur sans biais Z_n de θ.

d) Calculer le risque quadratique de Z_n.

e) Montrer que la suite d'estimateurs $(Z_n)_{n \in \mathbb{N}^*}$ est convergente.

3. Comparer les risques quadratiques de M_n et de Z_n.
Quelle méthode choisiriez-vous pour estimer la valeur du paramètre θ ?

4. **Un premier intervalle de confiance asymptotique pour θ**

a) Montrer que la suite de variables aléatoires $\left(2\sqrt{2n}\ \dfrac{Z_n - \theta}{\theta}\right)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire de loi normale centrée réduite.

b) Dans la suite, on note $\alpha \in \,]0, 1[$ et on note Φ la fonction de répartition associée à loi normale centrée réduite. Enfin, on note $t_{1-\frac{\alpha}{2}}$ l'unique réel tel que $\Phi(t_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.
Justifier l'existence de $t_{1-\frac{\alpha}{2}}$.

c) Justifier que : $\lim\limits_{n \to +\infty} \mathbb{P}\left(\left[-t_{1-\frac{\alpha}{2}} \leqslant 2\sqrt{2n}\ \dfrac{Z_n - \theta}{\theta} \leqslant t_{1-\frac{\alpha}{2}}\right]\right) = 1 - \alpha.$
En déduire un intervalle de confiance asymptotique de θ au niveau de confiance 95%.
(on donne $\Phi(1,96) = 0,975$)

5. **Un second intervalle de confiance asymptotique pour θ**

a) Montrer que la suite de variables aléatoires $\left(2n\left(1 - \dfrac{M_n}{\theta}\right)\right)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire W telle que W suit la loi exponentielle $\mathcal{E}(1)$.

b) Calculer $\mathbb{P}([-\ln(0,975) \leqslant W \leqslant -\ln(0,025)])$.

c) En déduire un nouvel intervalle de confiance asymptotique de θ au niveau de confiance 95%.

Exercice 3 : EDHEC 2018

On admet que toutes les variables aléatoires considérés dans cet exercice sont définies sur le même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ que l'on ne cherchera pas à déterminer.

Soit a un réel strictement positif et f la fonction définie par : $f(x) = \left\{ \begin{array}{cl} \frac{x}{a}\ \mathrm{e}^{-\frac{x^2}{2a}} & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{array} \right.$.

1. Montrer que la fonction f est une densité.

Dans la suite de l'exercice, on considère une variable aléatoire X de densité f.

2. Déterminer la fonction de répartition F_X de X.

3. On considère la variable aléatoire Y définie par : $Y = \dfrac{X^2}{2a}$.

a) Montrer que Y suit la loi exponentielle de paramètre 1.

b) On rappelle qu'en **Scilab** la commande `grand(1, 1, 'exp', 1/lambda)` simule une variable aléatoire suivant la loi exponentielle de paramètre λ. Écrire un script **Scilab** demandant la valeur de a à l'utilisateur et permettant de simuler la variable aléatoire X.

4. ***a)*** Vérifier que la fonction g qui à tout réel associe $x^2\ \mathrm{e}^{-\frac{x^2}{2a}}$, est paire.

b) Rappeler l'expression intégrale ainsi que la valeur du moment d'ordre 2 d'une variable aléatoire Z suivant la loi normale de paramètre 0 et a.

c) En déduire que X possède une espérance et la déterminer.

5. ***a)*** Rappeler l'espérance de Y puis montrer que X possède un moment d'ordre 2 et le calculer.

b) En déduire que la variance de X est donnée par :

$$\mathbb{V}(X) = \frac{(4-\pi)\ a}{2}$$

On suppose désormais que le paramètre a est inconnu et on souhaite l'estimer.

6. Soit n un entier naturel supérieur ou égal à 1. On considère un échantillon $(X_1, \ldots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X.
On note S_n la variable aléatoire définie par $S_n = \dfrac{1}{2n} \sum\limits_{k=1}^{n} {X_k}^2$.

a) Montrer que S_n est un estimateur sans biais de a.

b) Montrer que X^2 possède une variance et que $\mathbb{V}\big(X^2\big) = 4a^2$.

c) Déterminer le risque quadratique $r_a(S_n)$ de S_n en tant qu'estimateur de a.
En déduire que S_n est un estimateur convergent de a.

7. On suppose que a est inférieur ou égal à 1.

a) Écrire l'inégalité de Bienaymé-Tchebychev pour la variable aléatoire S_n et en déduire :

$$\forall \varepsilon > 0,\ \mathbb{P}\Big([|S_n - a| \leqslant \varepsilon] \Big)\ \geqslant\ 1 - \frac{1}{n\ \varepsilon^2}$$

b) Déterminer une valeur de n pour laquelle $\left[S_n - \dfrac{1}{10}, S_n + \dfrac{1}{10} \right]$ est un intervalle de confiance pour a avec niveau de confiance au moins égal à 95%.

II. Maximum de vraisemblance

Exercice 4 : EDHEC 2014

Dans cet exercice, θ désigne un réel strictement positif et n un entier naturel supérieur ou égal à 2.

Pour tout k de $\mathbb{N}$, on pose : $u_k = \dfrac{1}{1+\theta}\left(\dfrac{\theta}{1+\theta}\right)^k$.

1. Montrer que la suite $(u_k)_{k\in\mathbb{N}}$ définit une loi de probabilité.

On considère maintenant une variable aléatoire X prenant ses valeurs dans $\mathbb{N}$ et dont la loi est donnée par :

$$\forall k \in \mathbb{N},\ \mathbb{P}([X = k]) = u_k.$$

2. *a)* On pose $Y = X + 1$. Reconnaître la loi de Y, puis en déduire l'espérance et la variance de X.

b) On rappelle que `grand(1, 1, 'geom', p)` renvoie une simulation d'une variable aléatoire géométrique de paramètre `p`. Compléter la fonction **Scilab** suivante pour qu'elle simule la variable aléatoire X :

```
function x = SimuX(theta)
    y = ................
    x = ................
endfunction
```

3. Dans cette question, on souhaite estimer le paramètre θ par la méthode du maximum de vraisemblance. Pour ce faire, on considère un échantillon $(X_1, X_2, \ldots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X et on introduit $\mathcal{L}$, de $\mathbb{R}_+^*$ dans $\mathbb{R}$, définie par :

$$\forall \theta \in \mathbb{R}_+^*,\ \mathcal{L}(\theta) = \prod_{k=1}^{n} \mathbb{P}([X_k = x_k])$$

où $x_1, x_2, \ldots, x_n$ désignent des entiers naturels éléments de $X(\Omega)$.
L'objectif est de chosir la valeur de θ qui rend $\mathcal{L}(\theta)$ maximale.

a) Écrire $\ln(\mathcal{L}(\theta))$ en fonction de θ et de $S_n = \sum\limits_{k=1}^{n} x_k$.

b) On considère la fonction φ définie par :

$$\forall \theta \in \left]0, +\infty\right[,\ \varphi(\theta) = S_n \ln(\theta) - (S_n + n)\ln(1+\theta)$$

Montrer que la fonction φ admet un maximum, atteint en un seul réel que l'on notera $\widehat{\theta}_n$ et que l'on exprimera en fonction de S_n. Que représente $\widehat{\theta}_n$ pour la fonction $\mathcal{L}$?

4. On pose dorénavant : $T_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.
La variable T_n est appelée estimateur du maximum de vraisemblance pour θ.

a) Vérifier que T_n est un estimateur sans biais de θ.

b) Calculer le risque quadratique $r_\theta(T_n)$ de T_n.

c) En déduire que T_n est un estimateur convergent de θ.

Exercice 5 : HEC 2012

Sous réserve d'existence, on note $\mathbb{E}(U)$ et $\mathbb{V}(U)$ respectivement, l'espérance mathématique et la variance d'une variable aléatoire U définie sur l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.
Pour p entier supérieur ou égal à 2, on dit que les variables aléatoires à densité $U_1, \dots, U_p$ sont indépendantes si pour tout p-uplet $(u_1, \dots, u_p)$ de réels, les événements $[U_1 \leqslant u_1], \dots, [U_p \leqslant u_p]$ sont indépendants.
L'objet du problème est l'étude de quelques propriétés d'une loi de probabilité utilisée notamment en fiabilité.
Les parties I et II sont largement indépendantes. La partie III est indépendante des parties I et II.

Partie I : Loi à 1 paramètre.

On note λ un paramètre réel strictement positif. On considère la fonction f_λ de $\mathbb{R}$ dans $\mathbb{R}$ définie par :

$$f_\lambda(x) = \begin{cases} \dfrac{\lambda}{2\sqrt{x}}\,\mathrm{e}^{-\lambda\sqrt{x}} & \text{si} \quad x > 0 \\ \\ 0 & \text{si} \quad x \leqslant 0 \end{cases}.$$

1. ***a)*** Montrer que la fonction f_λ est de classe $\mathcal{C}^2$ sur $\mathbb{R}_+^*$.

b) Dresser le tableau de variation de f_λ sur $\mathbb{R}_+^*$ et préciser les limites suivantes : $\lim\limits_{x \to 0^+} f_\lambda(x)$, $\lim\limits_{x \to +\infty} f_\lambda(x)$.

c) Établir la convexité de la fonction f_λ sur $\mathbb{R}_+^*$.

d) Tracer l'allure de la courbe représentative de f_λ dans le plan rapporté à un repère orthogonal.

2. ***a)*** Vérifier que la fonction $x \mapsto -\mathrm{e}^{-\lambda\sqrt{x}}$ est une primitive de f_λ sur $\mathbb{R}_+^*$.

b) Établir la convergence de l'intégrale $\displaystyle\int_0^{+\infty} f_\lambda(x)\, dx$ et calculer sa valeur.

c) En déduire que la fonction f_λ est une densité de probabilité sur $\mathbb{R}_+^*$.

3. Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, à valeurs strictement positives, ayant f_λ pour densité. On note F_λ la fonction de répartition de X et on pose : $Y = \lambda\sqrt{X}$.

a) Calculer pour tout x réel, $F_\lambda(x)$.

b) Montrer que Y suit la loi exponentielle de paramètre 1.

c) Établir pour tout r de $\mathbb{N}^*$, l'existence de $\mathbb{E}(Y^r)$.

d) Montrer que pour tout r de $\mathbb{N}^*$, on a : $\mathbb{E}\left(Y^{r+1}\right) = (r+1)\,\mathbb{E}(Y^r)$.

e) En déduire pour tout r de $\mathbb{N}^*$, $\mathbb{E}(Y^r)$ et $\mathbb{E}(X^r)$. En particulier, calculer $\mathbb{E}(X)$ et $\mathbb{V}(X)$.

4. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. définies sur $(\Omega, \mathscr{A}, \mathbb{P})$, indépendantes et de même loi que X. Soit $(a_n)_{n \in \mathbb{N}^*}$ et $(b_n)_{n \in \mathbb{N}^*}$ deux suites de réels strictement positifs vérifiant $\lim\limits_{n \to +\infty} n^2\, a_n = 1$ et $\lim\limits_{n \to +\infty} n^2\, b_n = 0$.

On pose pour tout n de $\mathbb{N}^*$, $M_n = \min\limits_{1 \leqslant k \leqslant n} (X_k)$ et $J_n = \dfrac{M_n - b_n}{a_n}$.

On admet que M_n et J_n sont des variables aléatoires à densité définies sur $(\Omega, \mathscr{A}, \mathbb{P})$.
Montrer que la suite $(J_n)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire dont on précisera la loi.

Partie II : Estimation ponctuelle de λ.

Pour n entier de $\mathbb{N}^*$, on note $(X_1, \ldots, X_n)$ un n-échantillon de variables aléatoires à valeurs strictement positives, indépendantes et de même loi que la variable aléatoire X définie dans la question *3.*. On rappelle que $Y = \lambda\sqrt{X}$, et on pose pour tout k de $[\![1, n]\!]$, $Y_k = \lambda\sqrt{X_k}$, $S_k = \sum\limits_{j=1}^{k} Y_j$ et g_k une densité de S_k.

On admet que pour tout entier n supérieur ou égal à 2, les variables aléatoires $Y_1, \ldots, Y_n$ sont indépendantes et que pour tout k de $[\![1, n]\!]$, les variables aléatoires S_k et Y_{k+1} sont indépendantes. On admet que si T et Z sont deux variables aléatoires à densité indépendantes définies sur le même espace probabilisé, de densités respectives f_T et f_Z telles que f_T et f_Z soient bornées, alors la variable aléatoire $T + Z$ admet une densité f_{T+Z} définie pour tout x réel par :

$$f_{T+Z}(x) = \int_{-\infty}^{+\infty} f_T(y)\, f_Z(x - y)\ dy$$

4. ***a)*** En utilisant les propriétés admises, montrer que : $g_2(x) = \begin{cases} x\mathrm{e}^{-x} & \text{si} \quad x > 0 \\ 0 & \text{si} \quad x \leqslant 0 \end{cases}$.

b) À l'aide d'un raisonnement par récurrence, montrer que pour tout n de $\mathbb{N}^*$, on a :

$$g_n(x) = \begin{cases} \dfrac{1}{(n-1)!}\, x^{n-1}\mathrm{e}^{-x} & \text{si} \quad x > 0 \\ 0 & \text{si} \quad x \leqslant 0 \end{cases}$$

c) On admet que pour tout n de $\mathbb{N}^*$, $\dfrac{1}{S_n}$ est une variable aléatoire à densité.

Pour quelles valeurs de n, l'espérance $\mathbb{E}\left(\dfrac{1}{S_n}\right)$ et la variance $\mathbb{V}\left(\dfrac{1}{S_n}\right)$ existent-elles ?

Calculer alors leurs valeurs respectives.

5. On note $(x_1, \ldots, x_n)$ un n-uplet de $(\mathbb{R}_+^*)^n$ constituant une réalisation du n-échantillon $(X_1, \cdots, X_n)$. On suppose que le paramètre λ est inconnu. Soit H la fonction de $\mathbb{R}_+^*$ dans $\mathbb{R}$ définie par :

$$H(\lambda) = \ln\left(\prod_{k=1}^{n} f_\lambda(x_k)\right)$$

Montrer que la fonction H admet un maximum atteint en un unique point λ_0 dont on donnera la valeur.

6. On pose pour tout entier n supérieur ou égal à 3 : $\lambda_n^* = \dfrac{n}{\sum\limits_{k=1}^{n} \sqrt{X_k}}$.

a) Que représente λ_0 pour λ_n^* ?

b) Déterminer une suite réelle $(a_n)_{n\in\mathbb{N}^*}$ telle que $\mathbb{E}\left(a_n\lambda_n^*\right) = \lambda$. On note alors : $\hat{\lambda}_n = a_n\lambda_n^*$.

c) Calculer $\lim\limits_{n\to+\infty} \mathbb{V}\left(\hat{\lambda}_n\right)$.

d) Construire à partir de λ_n^* un estimateur sans biais $\hat{\lambda}_n$ de λ et calculer le risque quadratique $\rho\left(\hat{\lambda}_n\right)$ de $\hat{\lambda}_n$.

e) Calculer $\lim\limits_{n\to+\infty} \rho\left(\hat{\lambda}_n\right)$. Commenter.

Partie III : Loi à 2 paramètres.

7. Soit λ et α deux paramètres réels strictement positifs et $f_{(\lambda,\alpha)}$ la fonction définie sur $\mathbb{R}$ par

$$f_{(\lambda,\alpha)}(x) = \begin{cases} \lambda\alpha\, x^{\alpha-1}\,\mathrm{e}^{-\lambda x^{\alpha}} & \text{si} \quad x > 0 \\ 0 & \text{si} \quad x \leqslant 0 \end{cases}$$

a) Montrer que $f_{(\lambda,\alpha)}$ est une densité de probabilité sur $\mathbb{R}$.
Soit W une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, à valeurs strictement positives, de densité $f_{(\lambda,\alpha)}$. On dit que W suit la loi $\mathcal{WB}(\lambda, \alpha)$.

b) On note $F_{(\lambda,\alpha)}$ la fonction de répartition de W. Calculer pour tout x réel, $F_{(\lambda,\alpha)}(x)$.

c) Montrer que la variable aléatoire $F_{(\lambda,\alpha)}(x)$ suit la loi uniforme sur $[0, 1]$.

d) Écrire une fonction **Scilab** permettant de simuler W.

8. Soit K une variable aléatoire à densité définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, à valeurs strictement positives, de densité f_K nulle sur $\mathbb{R}_-$, continue sur $\mathbb{R}$, de classe $\mathcal{C}^1$ sur $\mathbb{R}_+^*$ et strictement positive sur $\mathbb{R}_+^*$. On note F_K la fonction de répartition de K.
On pose pour tout x réel $R(x) = -\ln(1 - F_K(x))$ et $r(x) = R'(x)$, où R' est la dérivée de R.

a) On suppose dans cette question que K suit la loi $\mathcal{WB}(\lambda, 2)$ avec $\lambda > 0$.
Établir les propriétés ***(i)*** et ***(ii)*** suivantes :

(i) la fonction r est continue et strictement croissante sur $\mathbb{R}_+$, et $r(0) = 0$.

(ii) la variable aléatoire $r(K)$ suit la loi $\mathcal{WB}\left(\dfrac{1}{4\lambda}, 2\right)$.

b) Réciproquement, on suppose dans cette question que les propriétés ***(i)*** et ***(ii)*** sont vérifiées.
Montrer que K suit la loi $\mathcal{WB}(\lambda, 2)$. Conclusion ?

Dans les questions ***9.*** et ***10.***, l'entier n est supérieur ou égal à 2. On note $w_1, \dots, w_n$ des réels strictement positifs et non tous égaux.

9. Soit φ la fonction de $\mathbb{R}_+^*$ dans $\mathbb{R}$ définie par : $\varphi(x) = \dfrac{\sum\limits_{k=1}^{n} (w_k)^x \,\ln(w_k)}{\sum\limits_{k=1}^{n} (w_k)^x} - \dfrac{1}{x}$.

a) Soit $y_1, \dots, y_n$ des réels non tous nuls et $z_1, \dots, z_n$ des réels quelconques.
En étudiant la fonction polynomiale du second degré Q définie sur $\mathbb{R}$ par $Q(t) = \sum\limits_{k=1}^{n} (z_k - t\, y_k)^2$, établir l'inégalité :

$$\left(\sum_{k=1}^{n} y_k\, z_k\right)^2 \leqslant \left(\sum_{k=1}^{n} y_k^2\right)\left(\sum_{k=1}^{n} z_k^2\right)$$

b) Montrer que la fonction φ est strictement croissante sur $\mathbb{R}_+^*$.

c) On note n_0 le nombre d'entiers k_0 de $[\![1, n]\!]$ vérifiant $w_{k_0} = \max\limits_{1 \leqslant k \leqslant n} (w_k)$.
Montrer que $1 \leqslant n_0 \leqslant n - 1$.

d) Donner un équivalent de $\sum\limits_{k=1}^{n} (w_k)^x$ en fonction de n_0 et w_{k_0}, lorsque x tend vers $+\infty$.

e) Calculer en fonction de w_{k_0}, la limite de $\varphi(x)$ lorsque x tend vers $+\infty$.
(on distinguera les deux cas $w_{k_0} = 1$ et $w_{k_0} \neq 1$).

f) En déduire que sur $\mathbb{R}_+^*$ l'équation $\varphi(x) = \dfrac{1}{n} \sum\limits_{k=1}^{n} \ln(w_k)$ admet une unique solution.

10. On note $(W_1, \dots, W_n)$ un n-échantillon de variables aléatoires à valeurs strictement positives, indépendantes et de même loi $\mathcal{WB}(\lambda, \alpha)$ définie dans la question ***7.*** dont une réalisation est le n-uplet $(w_1, \dots, w_n)$.
On suppose que les paramètres λ et α sont inconnus.
Soit G la fonction de $(\mathbb{R}_+^*)^2$ dans $\mathbb{R}$ définie par $G(\lambda, \alpha) = \ln\left(\prod\limits_{k=1}^{n} f_{(\lambda,\alpha)}(w_k)\right)$.

a) Montrer que la fonction G admet un unique point critique $(\hat{\lambda}, \hat{\alpha})$ sur $(\mathbb{R}_+^*)^2$.

b) Montrer que la fonction G admet un maximum local au point $(\hat{\lambda}, \hat{\alpha})$.