TD19 - estimation

Exercice 1 (inspiré d'edhec 2012) ★ ★ ☆☆

Dans tout l'exercice, λ est un réel strictement posistif.

1)Soit $(I_n)_{n\in\mathbf{N}}$ la suite d'intégrales définie par :

$$I_n = \int_0^{+\infty} x^n e^{-\lambda x} dx.$$

a)Montrer que I_n est une intégrale convergente, pour tout $n \in \mathbf{N}$.

b)Soit $A > 0$. A l'aide d'une intégration par parties, établir que

$$\int_0^A x^{n-1} e^{-\lambda x} dx = \frac{A^n}{n} e^{-\lambda A} + \frac{\lambda}{n} \int_0^A x^n e^{-\lambda x} dx.$$

c)En déduire que $\forall n \in \mathbf{N}^*,\ I_{n-1} = \dfrac{\lambda}{n} I_n$.

2)Soit $(f_n)_{n\in\mathbf{N}^*}$ la suite de fonctions définie par :

$$f_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

a)A l'aide de la question 1), montrer que $\displaystyle\int_0^{+\infty} f_n(x)$ converge pour tout $n \in \mathbf{N}^*$ et que $\forall n \in \mathbf{N}^*$, $\displaystyle\int_0^{+\infty} f_n(x)dx = \int_0^{+\infty} f_{n+1}(x)dx$.

b)En déduire que $\forall n \in \mathbf{N}^*$, $\displaystyle\int_0^{+\infty} f_n(x)dx = 1$.

c)Conclure que f_n est une densité de probabilité pour tout $n \in \mathbf{N}^*$.

3)Soit X une variable aléatoire telle que $X \hookrightarrow \mathscr{E}(\lambda)$.

Pour tout $n \in \mathbf{N}^*$, on considère $(X_1, ..., X_n)$ un n-échantillon de X et on admet que la variable aléatoire $\displaystyle\sum_{k=1}^{n} X_k$ est une variable aléatoire de densité f_n.

Soit $(Z_n)_{n\geq 2}$ la variable aléatoire définie par :

$$\forall n \geq 2,\ Z_n = \frac{n}{\displaystyle\sum_{k=1}^{n} X_k}.$$

a)Justifier que Z_n est un estimateur [1] de λ.

b)Montrer que Z_n admet une espérance et que $E(Z_n) = \dfrac{n\lambda}{n-1}$.

c)Déterminer un estimateur Z'_n de λ tel que $E\left(Z'_n\right) = \lambda$.

1. Z_n est appelé estimateur du maximum de vraisemblance de λ

Exercice 2 ★ ★ ☆☆

Soit $\theta \in \mathbf{R}$ et soit T_n un estimateur de θ et admettant une espérance.

On appelle *biais*[2] de T_n, le nombre réel, noté $b(T_n)$, défini par :

$$b(T_n) = E(T_n) - \theta = E(T_n - \theta).$$

On dit que T_n est un *estimateur sans biais* de θ si $b(T_n) = 0$ ou si $E(T_n) = \theta$.

On appelle *risque quadratique*[3] de T_n, le nombre réel, noté $r(T_n)$, défini sous réserve d'existence par :

$$r(T_n) = E\left((T_n - \theta)^2\right).$$

1)Montrer que $r(T_n) = b(T_n)^2 + V(T_n)$.

Dans toute la suite, on considère une variable aléatoire $X \hookrightarrow \mathscr{U}([0,\theta])$.
Pour tout $n \in \mathbf{N}^*$, on considère $(X_1, ..., X_n)$ un n-échantillon de X.

On note $\overline{X_n} = \dfrac{X_1 + \cdots + X_n}{n}$, la *moyenne empirique* de l'échantillon.

2)Soit $A_n = 2\overline{X_n}$.

a)Montrer que A_n est un estimateur sans biais de θ.

b)Montrer que $r(A_n) = \dfrac{\theta^2}{n}$.

3)Soit $B_n = max(X_1, ..., X_n)$.

a)Déterminer la fonction de répartition F de X, puis la fonction de répartition F_n de B_n.

b)En déduire une densité f_n de B_n.

c)Montrer que $E(B_n) = \dfrac{n\theta}{n+1}$, puis que $b(B_n) = -\dfrac{\theta}{n+1}$.

d)Montrer que $r(B_n) = \dfrac{2\theta^2}{(n+1)(n+2)}$.

4)Soit $C_n = \dfrac{n+1}{n} B_n$.

a)Montrer que C_n est un estimateur sans biais de θ.

b)Montrer que $r(C_n) = \dfrac{\theta^2}{(n+1)(n+2)}$.

5)Un bon estimateur de θ est un estimateur dont le biais est nul et dont le risque quadratique est le plus petit possible.

Des trois estimateurs A_n, B_n et C_n de θ, lequel est le meilleur ?

2. le biais représente l'écart moyen entre la valeur aléatoire prise par T_n et θ
3. le risque quadratique est l'écart au carré moyen entre la valeur aléatoire prise par T_n et θ

Exercice 3 (d'après edhec 2019) ★ ★ ☆☆

Soit $\theta > 0$ un réel inconnu et $Y \hookrightarrow \mathscr{E}\left(\dfrac{1}{\theta}\right)$.

Le but de l'exercice est de trouver une estimation ponctuelle de θ, puis de construire un intervalle de confiance de θ.

On considère n variables aléatoires $Y_1, \ldots, Y_n$ mutuellement indépendantes, et suivant toutes la même loi que Y.

On pose $T_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} Y_k$.

1)Justifier que T_n est un estimateur sans biais de θ.

2)Calculer le risque quadratique de T_n.

On pourra utiliser la formule vue dans la question 1) de l'exercice 2.

3)a)En écrivant l'inégalité de Bienaymé-Tchebychev pour T_n, établir l'inégalité :

$$\forall \epsilon > 0, \quad P\Big(\theta \in [T_n - \epsilon, T_n + \epsilon]\Big) \geqslant 1 - \frac{\theta^2}{n\epsilon^2}$$

b)On suppose que $\theta \leq \dfrac{1}{2}$ et que $n = 1000$.

Déterminer un intervalle de confiance de θ au niveau de confiance 90%.

Exercice 4 (d'après eml 2020) ★ ★ ★☆

Dans tout l'exercice, a est un réel strictement positif.

On note f la fonction définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < 1 \\ \dfrac{a}{x^{a+1}} & \text{si } x \geq 1 \end{cases}$$

1)Montrer que f est une densité de probabilité.

Par la suite, on considère une variable aléatoire[4] X de densité f.

On considère également une suite $(X_n)_{n\in\mathbf{N}^*}$ de variables aléatoires indépendantes, et toutes de même loi que X.

Pour tout $n \in \mathbf{N}^*$, on pose $W_n = \ln(X_n)$.

2)Déterminer la fonction de répartition F_n de X_n.

3)Montrer que la variable aléatoire W_n suit une loi exponentielle dont on précisera le paramètre. En déduire l'espérance et la variance de W_n.

4)On définit pour tout $n \in \mathbf{N}^*$:

$$M_n = \frac{W_1 + \cdots + W_n}{n} \qquad \text{et} \qquad T_n = \sqrt{n}\left(a\,M_n - 1\right)$$

a)Justifier que la suite de variables aléatoires $(T_n)_{n\in\mathbf{N}^*}$ converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.

b)En déduire que l'intervalle $\left[\dfrac{\sqrt{n}-2}{\sqrt{n}\,M_n}\ ,\ \dfrac{\sqrt{n}+2}{\sqrt{n}\,M_n}\right]$ est un intervalle de confiance asymptotique pour a au niveau de confiance 95%.

On admettra que $\Phi(2) \geq 0,975$, où Φ désigne la fonction de répartition d'une variable aléatoire suivant la loi normale centrée réduite.

4. on dit que X suit une loi de Pareto

Indications / Réponses

Exercice 1

1)a)Montrer d'abord que $x^n e^{-\lambda x} \underset{+\infty}{=} o\left(\dfrac{1}{x^2}\right)$.

c)Faire tendre A vers $+\infty$ dans 1)b).

2)b)Calculer $\displaystyle\int_0^{+\infty} f_1(x)dx$ et utiliser 2)a).

3)b)Appliquer le théorème de transfert.

c)Prendre $Z_n' = \dfrac{n-1}{n} Z_n$.

Exercice 2

1)Développer le carré, utiliser la linéarité de l'espérance, puis la formule de Koënig.

3)a)$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x}{\theta} & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } x > \theta \end{cases}$

$F_n(x) = \cdots = F(x)^n = \begin{cases} 0 & \text{si } x < 0 \\ \left(\dfrac{x}{\theta}\right)^n & \text{si } 0 \leq x \leq \theta \\ 1 & \text{si } x > \theta \end{cases}$

b)$f_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{n}{\theta}\left(\dfrac{x}{\theta}\right)^{n-1} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{si } x > \theta \end{cases}$

5)C'est C_n the best !

Exercice 3

1)Vérifier, entre autres, que $E(T_n) = \theta$. Classique !

2)$r(T_n) = b(T_n)^2 + V(T_n) = 0^2 + \dfrac{\theta^2}{n} = \dfrac{\theta^2}{n}$.

3)a)L'inégalité de Bienaymé-Tchébychev (deuxième forme) s'écrit :

$\forall \epsilon > 0,\ P\left(\left|T_n - E(T_n)\right| < \epsilon\right) \geq 1 - \dfrac{V(T_n)}{\epsilon^2}$.

Concluez en utilisant les questions 1) et 2).

b)Il s'agit de choisir $\epsilon > 0$ de sorte que $P\Big(\theta \in [T_n - \epsilon, T_n + \epsilon]\Big) \geq 0,90$.

Comme $\theta \leq \dfrac{1}{2}$, on a : $\theta^2 \leq \dfrac{1}{4}$, puis $1 - \dfrac{\theta^2}{n\epsilon^2} \geq 1 - \dfrac{1}{4n\epsilon^2}$.

Donc pour que $P\Big(\theta \in [T_n - \epsilon, T_n + \epsilon]\Big) \geq 0,90$, il suffit de choisir ϵ de sorte que

$1-\dfrac{1}{4n\epsilon^2}=0,9$, ce qui mène à $\epsilon=\dfrac{1}{20}$.

L'intervalle de confiance demandé est donc $\left[T_n-\dfrac{1}{20},T_n+\dfrac{1}{20}\right]$.

<u>Exercice 4</u>

2)On trouve $F_n(x)=\begin{cases}0 & \text{si } x<1\\ 1-\dfrac{1}{x^a} & \text{si } x\geq 1\end{cases}$

3)Notons G_n la fonction de répartition de W_n.

$\forall x\in\mathbf{R},\ G_n(x)=P(W_n\leq x)=P(\ln(X_n)\leq x)=P(X_n\leq e^x)=F_n(e^x)$.

D'où $G_n(x)=\begin{cases}0 & \text{si } x<0\\ 1-e^{-ax} & \text{si } x\geq 0\end{cases}$

Donc $W_n\hookrightarrow\mathscr{E}(a)$. Le cours donne : $E(W_n)=\dfrac{1}{a}$ et $V(W_n)=\dfrac{1}{a^2}$.

4)a)Vérifier que $T_n=M_n^*$, puis appliquer le théorème de la limite centrée.

b)Il faut établir que $\lim\limits_{n\to+\infty}P\left(\dfrac{\sqrt{n}-2}{\sqrt{n}M_n}\leq a\leq\dfrac{\sqrt{n}+2}{\sqrt{n}M_n}\right)\geq 0,95$.

En utilisant la question 4)a), voir que la limite ci-dessus est égale à $2\Phi(2)-1$ et conclure.