Exercice Ex 3

Un générateur aléatoire donne des réels au hasard de l'intervalle $[0, a]$. En l'utilisant 100 fois on a obtenu 100 valeurs x_1, x_2, ..., x_{100}. On se propose de donner une estimation de a à partir de ces valeurs en construisant un estimateur adapté à la situation.

Q1. Proposer un ensemble Θ, une famille de loi $(\mu_\theta)_{\theta\in\Theta}$ et une application g adaptés à cette estimation.

Q2. $n \in [\![2, +\infty[\![$. Soit θ un réel strictement positif et $(X_1, X_2, \ldots, X_n)$ un n-échantillon indépendant identiquement distribué de la loi uniforme sur $[0, \theta]$.

a) On pose $\overline{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$. Montrer que $U_n = 2\,\overline{X}_n$ est un estimateur sans biais et convergent de θ.

Préciser son risque quadratique.

b) Montrer que $V_n = \text{Max}(X_1, X_2, \ldots, X_n)$ est un estimateur asymptotiquement sans biais et convergent de θ.

Préciser son son biais et son risque quadratique.

c) Déduire de V_n un estimateur W_n sans biais et convergent. Préciser son risque quadratique.

Q3. Quelle estimateur de θ choisissez-vous lorsque n est assez grand ?

Q4. Quelle estimation de a retenez vous ?

Q1 $\Theta =]0, +\infty[$. Pour tout $\theta \in]0,+\infty[$, μ_θ est la loi uniforme sur $[0, \theta]$.

On cherche à estimer θ donc on pose $\forall \theta \in]0,+\infty[$, $g(\theta) = \theta$ (pour estimer l'espérance (resp. la variance) on aurait choisi $g : \theta \mapsto \theta/2$ (resp. $g : \theta \mapsto \theta^2/12$).

Q2 a) $\forall k \in [\![1, n]\!]$, $E(X_k) = \dfrac{\theta}{2}$ et $V(X_k) = \dfrac{\theta^2}{12}$.

Alors $E(\overline{X}_n) = \dfrac{1}{n} \times n \times E(X_1) = \dfrac{\theta}{2}$ et $V(\overline{X}_n) = \dfrac{1}{n^2} \times n \times \dfrac{\theta^2}{12} = \dfrac{\theta^2}{12n}$.

Ainsi $E(U_n) = 2E(\overline{X}_n) = \theta$ et $V(U_n) = 4V(\overline{X}_n) = \dfrac{\theta^2}{3n}$.

$E(U_n) = \theta$ et $V(U_n) = \dfrac{\theta^2}{3n}$. U_n est un estimateur sans biais de θ.

$\forall \varepsilon \in \mathbb{R}_+^*$, $0 \leq P(|U_n - \theta| \geq \varepsilon) = P(|U_n - E(U_n)| \geq \varepsilon) \leq \dfrac{V(U_n)}{\varepsilon^2} = \dfrac{\theta^2}{3n\varepsilon^2}$ et $\lim\limits_{n\to+\infty} \dfrac{\theta^2}{3n\varepsilon^2} = 0$

Alors la suite de terme général U_n converge en loi vers la variable certaine θ.

U_n est un estimateur convergent de θ.

Son risque quadratique est sa variance car il est sans biais.

Le risque quadratique de U_n est $\dfrac{\theta^2}{3n}$.

b) Soit F_n la fonction de répartition de V_n. V_n prend ses valeurs dans $[0,\theta]$.

$\forall x \in]-\infty,0[,\ F_n(x)=0$ et $\forall x \in [\theta,+\infty[,\ F_n(x)=1$

$\forall x \in [0,\theta[,\ F_n(x) = P(\max(X_1,\dots,X_n) \leq x) \underset{\uparrow \text{indépendance}}{=} \prod_{k=1}^{n} P(X_k \leq x) \underset{\uparrow X_k \hookrightarrow \mathcal{U}([0,\theta])}{=} \prod_{k=1}^{n} \frac{x}{\theta} = \left(\frac{x}{\theta}\right)^n.$

$$\forall x \in \mathbb{R},\ F_n(x) = \begin{cases} 0 & \text{si } x \in]-\infty,0[\\ \left(\frac{x}{\theta}\right)^n & \text{si } x \in [0,\theta[\\ 1 & \text{si } x \in [\theta,+\infty[\end{cases}.$$

$x \mapsto 0$, $x \mapsto \left(\frac{x}{\theta}\right)^n$ et $x \mapsto 1$ sont de classe C^1 sur $\mathbb{R}$. Donc F_n est de classe C^1 sur $]-\infty,0[$, $[0,\theta[$ et $[\theta,+\infty[$. F_n est donc de classe C^1 sur $\mathbb{R}-\{0,\theta\}$, continue à droite en 0 et θ. $\lim\limits_{x\to 0^-} F_n(x) = 0 = \left(\frac{0}{\theta}\right)^n = F_n(0)$ et $\lim\limits_{x\to\theta^-} F_n(x) = 1 = F_n(\theta)$

Ainsi F_n est continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R}-\{0,\theta\}$. Donc V_n est une variable aléatoire à densité.

$\forall x \in]-\infty,0[\cup]\theta,+\infty[,\ F_n'(x)=0$ et $\forall x \in]0,\theta[,\ F_n'(x) = \frac{n}{\theta^n}x^{n-1}$.

Posons $\forall x \in \mathbb{R}$, $f_n(x) = \begin{cases} \frac{n}{\theta^n}x^{n-1} & \text{si } x \in [0,\theta] \\ 0 & \text{sinon} \end{cases}$. Alors f_n est positive sur $\mathbb{R}$ et f_n coïncide avec F_n' sur $\mathbb{R}-\{0,\theta\}$ donc f_n est une densité de V_n.

Soit $\ell \in \mathbb{N}$. $\int_{-\infty}^{0} t^\ell f_n(t)\,dt$ et $\int_{\theta}^{+\infty} t^\ell f_n(t)\,dt$ existent et valent 0.

De manière claire $\int_0^\theta t^\ell f_n(t)\,dt$ existe et vaut $\frac{n}{\theta^n}\left[\frac{t^{n+\ell}}{n+\ell}\right]_0^\theta$ ou $\frac{n}{n+\ell}\theta^\ell$.

Alors $E(V_n)$ existe et vaut $\frac{n}{n+1}\theta$ et $E(V_n^2)$ existe et vaut $\frac{n}{n+2}\theta^2$.

Alors $V(V_n)$ existe et $V(V_n) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$

$\lim\limits_{n\to+\infty} E(V_n) = \lim\limits_{n\to+\infty}\left(\frac{n}{n+1}\theta\right) = \theta$; V_n est un estimateur asymptotiquement sans biais de θ

V_n possède un moment d'ordre 2 donc $V_n - \theta$ également. Markov donne :

$$\forall \varepsilon \in \mathbb{R}_+^*,\quad 0 \leq P(|V_n-\theta| \geq \varepsilon) \leq \frac{E((V_n-\theta)^2)}{\varepsilon^2} = \frac{V(V_n-\theta)+(E(V_n-\theta))^2}{\varepsilon^2} = \frac{V(V_n)+(E(V_n)-\theta)^2}{\varepsilon^2}$$

Or $\lim\limits_{n\to+\infty} V(V_n) = \lim\limits_{n\to+\infty} \frac{n}{(n+2)(n+1)^2}\theta^2 = 0$ et $\lim\limits_{n\to+\infty} (E(V_n)-\theta)^2 = 0$.

Par encadrement il vient : $\forall \varepsilon \in \mathbb{R}_+^*,\ \lim\limits_{n\to+\infty} P(|V_n-\theta|\geq\varepsilon) = 0$.

V_n est un estimateur convergent de θ.

Le biais de V_n est $E(V_n)-\theta$ c'est à dire $-\frac{\theta}{n+1}$; $b_{V_n}(\theta) = -\frac{\theta}{n+1}$.

Le risque quadratique de V_n est $E((V_n-\theta)^2)$ ou $V(V_n)+(b_{V_n}(\theta))^2$.

$$V(V_n)+(b_{V_n}(\theta))^2 = \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+2)(n+1)}.$$

Le risque quadratique de V_n est $\frac{2\theta^2}{(n+2)(n+1)}$. $r_{V_n}(\theta) = \frac{2\theta^2}{(n+2)(n+1)}$.

c) Posons $W_n = \frac{n+1}{n}V_n$. $E(W_n) = \frac{n+1}{n}E(V_n) = \theta$ et

$$V(W_n) = \left(\frac{n+1}{n}\right)^2 \times \frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{\theta^2}{n(n+2)}.$$

W_n est un estimateur sans biais de θ.

$\forall \varepsilon \in \mathbb{R}_+^*,\ 0 \leq P(|W_n-\theta|\geq\varepsilon) = P(|W_n-E(W_n)|\geq\varepsilon) \leq \frac{V(W_n)}{\varepsilon^2} = \frac{\theta^2}{n(n+2)\varepsilon^2}$ et $\lim\limits_{n\to+\infty}\frac{\theta^2}{n(n+2)\varepsilon^2} = 0$

Par encadrement : $\forall \varepsilon \in \mathbb{R}_+^*,\ \lim\limits_{n\to+\infty} P(|W_n-\theta|\geq\varepsilon) = 0$

W_n est un estimateur convergent de θ. Son risque quadratique est $V(W_n) = \frac{\theta^2}{n(n+2)}$.

Q3 $r_\theta(U_n) = \frac{\theta^2}{3n} \geq \frac{\theta^2}{n(n+2)} = r_\theta(W_n)$ et $r_\theta(V_n) = \frac{2\theta^2}{(n+2)(n+1)} \geq \frac{\theta^2}{n(n+2)} = r_\theta(W_n)$.

Nous choisirons donc comme estimateur de θ : $W_n = \frac{n+1}{n}\operatorname{Max}(X_1,X_2,\dots,X_n)$.

Q4 Nous proposerons comme estimation de a : $\frac{101}{100}\operatorname{Max}(x_1,x_2,\dots,x_{100})$.

Exercice θ est un réel strictement positif. X est une variable aléatoire de densité f définie par :

Ex 5

$$\forall t \in]-\infty, \theta[,\ f(t) = 0 \quad \text{et} \quad \forall t \in [\theta, +\infty[,\ f(t) = e^{-(t-\theta)}.$$

Q1. Vérifier rapidement que f est une densité de probabilité. Que dire de : $Z = X - \theta$? Calculer $E(X)$ et $V(X)$.

Q2. $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, P)$ de même loi que X.

Pour tout n élément de $\mathbb{N}^*$ on pose $\overline{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ et $Y_n = \text{Min}(X_1, X_2, \ldots, X_n)$.

a) Construire à partir de $\overline{X}_n$ un estimateur T_n sans biais et convergent de θ.

b) Montrer que Y_n est un estimateur asymptotiquement sans biais et convergent de θ (on pourra comencer par étudier $Y_n - \theta$).

Préciser le risque quadratique.

Construire à partir de Y_n un estimateur U_n sans biais et convergent de θ.

c) Comparer T_n et U_n.

Q1 * → $\forall t \in]-\infty, \theta[$, $f(t) = 0 \geqslant 0$ et $\forall t \in [\theta, +\infty[$, $f(t) = e^{-(t-\theta)} \geqslant 0$.

f est positive sur son domaine de définition qui est $\mathbb{R}$.

→ $t \to 0$ et $t \mapsto e^{-(t-\theta)}$ sont continues sur $\mathbb{R}$ donc f est continue sur $]-\infty, \theta[$ et sur $[\theta, +\infty[$. Ainsi f est au moins continue sur $\mathbb{R} - \{\theta\}$ donc sur $\mathbb{R}$ privé d'un ~~ensemble~~ fini de points.

→ $\int_{-\infty}^{\theta} f(t)dt$ existe et vaut 0. Soit $A \in [\theta, +\infty[$.

$\int_{\theta}^{A} f(t)dt = \int_{\theta}^{A} e^{-(t-\theta)}dt = \left[-e^{-(t-\theta)}\right]_{\theta}^{A} = 1 - e^{-(A-\theta)}$ et $\lim_{A\to+\infty} (1 - e^{-(A-\theta)}) = 1$

Alors $\int_{\theta}^{+\infty} f(t)dt$ existe et vaut 1.

Finalement $\int_{-\infty}^{+\infty} f(t)dt$ converge et vaut 1. Ceci achève de montrer que

<u>f est une densité de probabilité.</u>

X est une variable aléatoire à densité de densité f. Alors $Z = X - \theta$ est une variable aléatoire à densité de densité $g : x \mapsto \frac{1}{|1|} f\left(\frac{x-(-\theta)}{1}\right)$.

$\forall x \in \mathbb{R}$, $g(x) = f(x+\theta)$.

Alors $\forall x \in]-\infty, 0[$, $x+\theta \in]-\infty, \theta[$ et $g(x) = 0$.

$\forall x \in [0, +\infty[$, $x+\theta \in [\theta, +\infty[$ et $g(x) = e^{-((x+\theta)-\theta)} = e^{-x}$.

$\forall x \in \mathbb{R}$, $g(x) = \begin{cases} e^{-x} & \text{si } x \in [0, +\infty[\\ 0 & \text{sinon} \end{cases}$

Plus de doute $Z = X - \theta$ suit la loi exponentielle de paramètre 1.

Alors $E(Z)$ existe et vaut 1. Même chose pour $V(Z)$!

$X = Z + \theta$ donc 1) $E(X)$ existe et vaut $1 + \theta$.
2) $V(X)$ existe et vaut 1.

$E(X) = \theta + 1$ et $V(X) = 1$.

Q2 a) Soit $n \in \mathbb{N}^*$. Pour tout $i \in [\![1, n]\!]$, $E(X_i)$ existe et vaut $1 + \theta$.

Ainsi $E(\overline{X}_n)$ existe et vaut $\frac{1}{n} \sum_{i=1}^{n} E(X_i)$ donc $\frac{1}{n} \times n(1+\theta)$ c'est-à-dire $1 + \theta$.

$\forall n \in \mathbb{N}^*$, $E(\overline{X}_n) = 1 + \theta$

Posons $\forall n \in \mathbb{N}^*$, $T_n = \overline{X}_n - 1$.

Alors pour tout $n \in \mathbb{N}^*$, $E(T_n)$ existe et vaut $1 + \theta - 1$ donc θ.

$T_n = \overline{X}_n - 1$ est un estimateur sans biais de θ.

Soit $n \in \mathbb{N}^*$. $V(X_i)$ existe pour tout $i \in [\![1, n]\!]$ et vaut 1.

Comme $X_1, X_2, \dots, X_n$ sont indépendantes : $V(X_1 + X_2 + \dots + X_n)$ existe et vaut $V(X_1) + V(X_2) + \dots + V(X_n)$ donc n.

Alors $V(\overline{X}_n)$ existe et vaut $\frac{1}{n^2} \times n$ donc $\frac{1}{n}$.

Alors $V(T_n)$ existe et vaut $\frac{1}{n}$. $\frac{1}{n}$ est donc le risque quadratique de T_n car T_n est sans biais. L'inégalité de Bienaymé Tchebychev donne alors :

$\forall \varepsilon \in \mathbb{R}_+^*$, $0 \leqslant P(|T_n - \theta| \geqslant \varepsilon) = P(|T_n - E(T_n)| \geqslant \varepsilon) \leqslant \frac{V(T_n)}{\varepsilon^2} = \frac{1}{n\varepsilon^2}$ et $\lim_{n \to +\infty} \frac{1}{n\varepsilon^2} = 0$

Par encadrement $\lim_{n \to +\infty} P(|T_n - \theta| \geqslant \varepsilon) = 0$ pour tout ε dans $\mathbb{R}_+^*$.

$(T_n)_{n\geq 1}$ converge en probabilité vers θ.

Ainsi T_n est un estimateur sans biais et convergent de θ de risque quadratique $\frac{1}{n}$.

Remarque.- Retrouvons la convergence. Posons $\forall i \in \mathbb{N}^*$, $X'_i = X_i - 1$.

- $\forall n \in \mathbb{N}^*$, $T_n = \overline{X}_n - 1 = \frac{1}{n}\sum_{i=1}^{n} X_i - 1 = \frac{1}{n}\left(\sum_{i=1}^{n}(X_i - 1)\right) = \frac{1}{n}\sum_{i=1}^{n} X'_i$.
- De plus les variables aléatoires $(X'_i)_{i\in\mathbb{N}^*}$ sont indépendantes
- Notons encore que pour tout $i \in \mathbb{N}^*$, $E(X'_i)$ existe et vaut θ et $V(X'_i)$ existe et vaut 1.

La loi faible des grands nombres montre alors que $(T_n)_{n\geq 1}$ converge en probabilité vers la variable certaine égale à θ.

b) Posons $\forall i \in \mathbb{N}^*$, $Z_i = X_i - \theta$ et $\forall n \in \mathbb{N}^*$, $V_n = Y_n - \theta$.

$\forall n \in \mathbb{N}^*$, $V_n = \min(X_1, X_2, \dots, X_n) - \theta = \min(X_1 - \theta, X_2 - \theta, \dots, X_n - \theta) = \min(Z_1, Z_2, \dots, Z_n)$

Soit $n \in \mathbb{N}^*$ et soit F_{V_n} la fonction de répartition de V_n.

Rappelons que $\forall i \in \mathbb{N}^*$, $Z_i \hookrightarrow \mathcal{E}(1)$. Posons $\forall x \in \mathbb{R}$, $F(x) = \begin{cases} 1 - e^{-x} & \text{si } x \in [0, +\infty[\\ 0 & \text{sinon} \end{cases}$

F est la fonction de répartition des variables aléatoires de la suite $(Z_i)_{i\in\mathbb{N}^*}$.

Notons encore que $(X_i)_{i\in\mathbb{N}^*}$ est une suite de variables aléatoires indépendantes.

Alors les variables de la suite $(Z_i)_{i\in\mathbb{N}^*}$ sont également indépendantes.

V_n prend ses valeurs dans $\mathbb{R}^+$. $\forall x \in]-\infty, 0[$, $F_{V_n}(x) = 0$. Soit $x \in [0, +\infty[$.

$F_{V_n}(x) = P(\min(Z_1, Z_2, \dots, Z_n) \leq x) = 1 - P(\min(Z_1, Z_2, \dots, Z_n) > x)$.

$F_{V_n}(x) = 1 - P(\{Z_1 > x\} \cap \{Z_2 > x\} \cap \dots \cap \{Z_n > x\})$. Par indépendance il vient :

$F_{V_n}(x) = 1 - P(Z_1 > x)\,P(Z_2 > x) \cdots P(Z_n > x) = 1 - (1 - P(Z_1 \leq x))(1 - P(Z_2 \leq x)) \cdots (1 - P(Z_n \leq x))$

$F_{V_n}(x) = 1 - (1 - F(x))^n = 1 - (1 - (1 - e^{-x}))^n = 1 - e^{-nx}$.

$\forall x \in \mathbb{R}$, $F_{V_n}(x) = \begin{cases} 1 - e^{-nx} & \text{si } x \in [0, +\infty[\\ 0 & \text{sinon} \end{cases}$.

Alors <u>V_n suit la loi exponentielle de paramètre n.</u>

Alors $E(V_n)$ (resp. $V(V_n)$) existe et vaut $\frac{1}{n}$ (resp. $\frac{1}{n^2}$).

Rappelons que $Y_n = V_n + \theta$.

Ainsi <u>$E(Y_n)$ existe et vaut $\theta + \frac{1}{n}$ et $V(Y_n)$ existe et vaut $\frac{1}{n^2}$.</u>

$\lim\limits_{n \to +\infty} (E(Y_n) - \theta) = \lim\limits_{n \to +\infty} \frac{1}{n} = 0$. <u>$Y_n$ est donc un estimateur asymptotiquement sans biais de θ</u>

Rappelons que Y_n possède un moment d'ordre 2. Markov permet alors d'écrire :

$\forall \varepsilon \in \mathbb{R}_+^*$, $0 \leq P(|Y_n - \theta| \geq \varepsilon) \leq \dfrac{E((Y_n-\theta)^2)}{\varepsilon^2} = \dfrac{E(V_n^2)}{\varepsilon^2} = \dfrac{V(V_n) + (E(V_n))^2}{\varepsilon^2} = \dfrac{2}{n^2\varepsilon^2}$ et $\lim\limits_{n \to +\infty} \dfrac{2}{n^2\varepsilon^2} = 0$

Donc $\forall \varepsilon \in \mathbb{R}_+^*$, $\lim\limits_{n \to +\infty} P(|Y_n - \theta| \geq \varepsilon) = 0$. <u>$Y_n$ est un estimateur convergent de θ.</u>

$E((Y_n - \theta)^2) = E(V_n^2) = \frac{2}{n^2}$. <u>Le risque quadratique de Y_n est $\frac{2}{n^2}$.</u>

Posons $\forall n \in \mathbb{N}^*$, $U_n = Y_n - \frac{1}{n}$.

Soit $n \in \mathbb{N}^*$, $E(U_n)$ existe et vaut $\theta + \frac{1}{n} - \frac{1}{n}$ donc θ et $V(U_n)$ existe et vaut $V(Y_n)$ donc $\frac{1}{n^2}$. <u>U_n est un estimateur sans biais de θ.</u>

$\forall \varepsilon \in \mathbb{R}_+^*$, $P(|U_n - \theta| \geq \varepsilon) = P(|U_n - E(U_n)| \geq \varepsilon) \leq \dfrac{V(U_n)}{\varepsilon^2} = \dfrac{1}{n^2\varepsilon^2}$ et $\lim\limits_{n \to +\infty} \dfrac{1}{n^2\varepsilon^2} = 0$.

$\forall \varepsilon \in \mathbb{R}_+^*$, $\lim\limits_{n \to +\infty} P(|U_n - \theta| \geq \varepsilon) = 0$. $(U_n)_{n \geq 1}$ converge en probabilité vers la variable aléatoire certaine égale à θ. <u>U_n est un estimateur convergent de θ.</u>

Son risque quadratique est sa variance donc $1/n^2$.

c) T_n et U_n sont deux estimateurs sans biais et convergents. Le risque quadratique de T_n est $\frac{1}{n}$ et celui de U_n est $\frac{1}{n^2}$. $\frac{1}{n^2} < \frac{1}{n}$ pour $n \geq 2$!

<u>U_n est donc un meilleur estimateur de θ que T_n.</u>

Exercice **Estimation d'une variance.** Ex 5

X est une variable aléatoire possèdant une espérance m et une variance σ^2.

$(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires réelles mutuellement indépendantes sur $(\Omega, \mathcal{A}, P)$ de même loi que X.

Q1. On suppose que X possède un moment d'ordre 4.

Montrer que $U_n = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} (X_k - m)^2$ est un estimateur sans biais et convergent de σ^2. Quel est son inconvénient ?

Q2 a) Montrer que la variance empirique $S_n^2 = \dfrac{1}{n} \displaystyle\sum_{k=1}^{n} \left(X_k - \overline{X}_n\right)^2$ est un estimateur asymptotiquement sans biais de la variance σ^2 (partir de $\left(X_k - \overline{X}_n\right)^2 = \left((X_k - m) - (\overline{X}_n - m)\right)^2$).

$\mathbf{S_n^2}$ est une simple notation, usuelle...

b) Montrer que la variance empirique corrigée (ou modifiée) $S'^2_n = \dfrac{1}{n-1} \displaystyle\sum_{k=1}^{n} \left(X_k - \overline{X}_n\right)^2$ est un estimateur sans biais de la variance σ^2.

On supposant que X possède un moment d'ordre 4, on peut montrer que les estimateurs précédents sont convergents.

Q1 Posons $\forall k \in \mathbb{N}^*$, $Y_k = (X_k - m)^2$. $\forall n \in \mathbb{N}^*$, $U_n = \dfrac{Y_1 + Y_2 + \cdots + Y_n}{n}$.

* $(Y_n)_{n\geqslant 1}$ est une suite de variables aléatoires (deux à deux...) indépendantes.

* $\forall k \in \mathbb{N}^*$, $Y_k = (X_k - m)^2 = (X_k - E(X_k))^2$ et pour tout $k \in \mathbb{N}^*$, X_k possède une variance qui vaut σ^2. Ainsi pour tout $k \in \mathbb{N}^*$, Y_k possède une espérance qui vaut σ^2.

* Soit $k \in \mathbb{N}^*$. $Y_k^2 = (X_k - m)^4 = X_k^4 - 4m X_k^3 + 6m^2 X_k^2 - 4m^3 X_k + m^4$.

X_k possède un moment d'ordre 4 donc X_k possède un moment d'ordre i pour tout $i \in [\![0,4]\!]$. Donc $E(X_k^4)$, $E(X_k^3)$, $E(X_k^2)$, $E(X_k)$ existent.

Y_k^2 possède alors une espérance comme combinaison linéaire de variables aléatoires qui possèdent une espérance. Alors Y_k possède un moment d'ordre 2 donc une variance.

$E(Y_k^2) = E(X_k^4) - 4m E(X_k^3) + 6m^2 E(X_k^2) - 4m^3 E(X_k) + m^4$

$E(Y_k^2) = E(X^4) - 4m E(X^3) + 6m^2 E(X^2) - 4m^3 E(X) + m^4 = E((X-m)^4)$.

Ainsi toutes les variables aléatoires de la suite (Y_k) possèdent la même espérance : $E((X-m)^4) - \sigma^4$.

Les trois points précédents et la loi faible des grands nombres permettent de dire que la suite de terme général $U_n = \frac{Y_1+Y_2+\dots+Y_n}{n} = \frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2$ converge en probabilité vers la variable certaine égale à σ^2.

Pour tout $k \in \mathbb{N}^*$, $E(Y_k) = E((X_k-m)^2)$ existe et vaut σ^2.

Donc pour tout $n \in \mathbb{N}^*$, $E(U_n)$ existe et vaut $\frac{1}{n}\sum_{k=1}^{n}E(Y_k) = \frac{1}{n}\sum_{k=1}^{n}\sigma^2 = \sigma^2$

Pour tout $n \in \mathbb{N}^*$, $E(U_n)$ existe et vaut σ^2.

De plus $U_n = \frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2$ est un estimateur sans biais et convergent de σ^2.

L'inconvénient de cet estimateur est qu'il suppose connue l'espérance m de X.

Q2 a) Soit $k \in \mathbb{N}^*$. Soit $n \in \mathbb{N}^*$. $E(X_k) = m$ et $E(\overline{X}_n) = E\left(\frac{X_1+\dots+X_n}{n}\right) = \frac{E(X_1)+\dots+E(X_n)}{n} = \frac{nm}{n} = m$.

$(X_k-\overline{X}_n)^2 = ((X_k-m)-(\overline{X}_n-m))^2 = (X_k-m)^2 - 2(X_k-m)(\overline{X}_n-m) + (\overline{X}_n-m)^2$.

$(X_k-\overline{X}_n)^2 = (X_k-E(X_k))^2 - 2(X_k-E(X_k))(\overline{X}_n-E(\overline{X}_n)) + (\overline{X}_n-E(\overline{X}_n))^2$.

X_k possède une variance qui vaut σ^2 donc $E((X_k-E(X_k))^2)$ existe et vaut σ^2.

$X_1, X_2, \dots, X_n$ sont indépendantes et possèdent une variance qui vaut σ^2.

Alors $V(X_1+\dots+X_n)$ existe et vaut $n\sigma^2$. Donc $V\left(\frac{X_1+\dots+X_n}{n}\right)$ existe et vaut $\frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$.

Donc $E((\overline{X}_n-E(\overline{X}_n))^2)$ existe et vaut $\frac{\sigma^2}{n}$.

X_k et $\overline{X}_n$ possèdent un moment d'ordre 2 donc $\mathrm{cov}(X_k,\overline{X}_n)$ existe.

Alors $E((X_k-E(X_k))(\overline{X}_n-E(\overline{X}_n)))$ existe et vaut $\mathrm{cov}(X_k,\overline{X}_n)$.

$\mathrm{cov}(X_k,\overline{X}_n) = \mathrm{cov}\left(X_k, \frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{cov}(X_k,X_i) = \frac{1}{n}\mathrm{cov}(X_k,X_k) = \frac{1}{n}V(X_k) = \frac{\sigma^2}{n}$

indépendance...

Ainsi $E((X_k-E(X_k))(\overline{X}_n-E(\overline{X}_n)))$ existe et vaut $\frac{\sigma^2}{n}$.

Ainsi $(X_\ell - \bar{X}_n)^2$ possède une espérance comme combinaison linéaire de trois variables aléatoires qui possèdent une espérance.

$$E((X_\ell-\bar{X}_n)^2) = \sigma^2 - 2\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \sigma^2\left(1-\frac{1}{n}\right) = \frac{n-1}{n}\sigma^2.$$

Alors S_n^2 possède une espérance comme combinaison linéaire de n variables aléatoires qui possèdent une espérance.

$$E(S_n^2) = \frac{1}{n}\sum_{\ell=1}^{n} E((X_\ell-\bar{X}_n)^2) = \frac{1}{n}\sum_{\ell=1}^{n}\left(\frac{n-1}{n}\right)\sigma^2 = \frac{n-1}{n}\sigma^2.$$

$E(S_n^2) = \frac{n-1}{n}\sigma^2$ et ceci pour tout $n\in\mathbb{N}^*$. Notons que $\lim_{n\to+\infty} E(S_n^2) = \sigma^2$.

$S_n^2 = \frac{1}{n}\sum_{\ell=1}^{n}(X_\ell-\bar{X}_n)^2$ est un estimateur asymptotiquement sans biais de σ^2.

Remarque.. Voyons une petite piste pour une variante de ce qui précède.

$$nS_n^2 = \sum_{\ell=1}^{n}(X_\ell-m)^2 - 2\sum_{\ell=1}^{n}(X_\ell-m)(\bar{X}_n-m) + \sum_{\ell=1}^{n}(\bar{X}_n-m)^2$$

$$nS_n^2 = \sum_{\ell=1}^{n}(X_\ell-m)^2 - 2(\bar{X}_n-m)\left[\sum_{\ell=1}^{n}X_\ell - nm\right] + n(\bar{X}_n-m)^2 = \sum_{\ell=1}^{n}(X_\ell-m)^2 - 2(\bar{X}_n-m)(n\bar{X}_n-nm)+n(\bar{X}_n-m)^2$$

$$nS_n^2 = \sum_{\ell=1}^{n}(X_\ell-m)^2 - 2(\bar{X}_n-m)^2 + n(\bar{X}_n-m)^2 = \sum_{\ell=1}^{n}(X_\ell-m)^2 - n(\bar{X}_n-m)^2.$$

Alors $S_n^2 = \frac{1}{n}\sum_{\ell=1}^{n}(X_\ell-m)^2 - (\bar{X}_n-m)^2$... donc $E(S_n^2)$ existe et vaut $\frac{1}{n}\times n\,\sigma^2 - V(\bar{X}_n)$.

Je vous laisse la suite.

b) $S_n'^2 = \frac{n}{n-1}S_n^2$, $E(S_n^2)$ existe et vaut $\frac{n-1}{n}\sigma^2$.

Alors $E(S_n'^2)$ existe et vaut $\frac{n}{n-1}\,\frac{n-1}{n}\sigma^2$ donc σ^2.

$S_n'^2$ est un estimateur sans biais de σ^2.

Contente la fille M. B. ?

Exercice 1 **PC et CC** ★ ESCP 97 Q1. On désire estimer le paramètre λ d'une loi de Poisson à l'aide d'un n-échantillon i.i.d. $(X_1, X_2, \dots, X_n)$ de variables aléatoires indépendantes suivant cette loi. $S_n = X_1 + X_2 + \cdots + X_n$.

Vérifier que $M_n = \dfrac{S_n}{n}$ est un estimateur sans biais et convergent de λ.

Q2. On cherche maintenant à estimer, à l'aide du même échantillon, la probabilité $e^{-q\lambda}$ de n'observer que des 0 au cours de q expériences consécutives.

On pose $T_n = e^{-q\frac{S_n}{n}}$. T_n est-il un estimateur sans biais de $e^{-q\lambda}$? Asymptotiquement sans biais ? Convergent ?

Q1 $\forall i \in [\![1,n]\!]$, $E(X_i)$ existe et vaut λ. Donc $E(S_n)$ existe et vaut $\sum_{i=1}^{n} E(X_i)$ donc $n\lambda$.

Alors $E(M_n)$ existe et vaut λ.

<u>M_n est un estimateur sans biais de λ.</u>

- Pour tout $i \in [\![1,n]\!]$, $E(X_i)$ existe et vaut λ, $V(X_i)$ existe et vaut λ
- $X_1, X_2, \dots, X_n$ sont indépendantes.

La loi faible des grands nombres montre alors que "M_n" converge en probabilité vers λ. <u>Ainsi M_n est un estimateur sans biais et convergent de λ.</u>

Q2 Visiblement nous allons être amené à étudier $E(T_n)$ et $E(T_n^2)$.

Etudions alors $E(T_n^r)$ pour tout r dans $\mathbb{N}$. $\forall r \in \mathbb{N}, T_n^r = e^{-\frac{rq}{n} S_n}$

Soit $r \in \mathbb{N}$. $E(T_n^r)$ existe si et seulement si la série de terme général général $e^{-\frac{rq}{n}k} P(S_n = k)$ est absolument convergente d'après le théorème de transfert.

Or $\forall k \in \mathbb{N}$, $e^{-\frac{rq}{n}k} P(S_n = k) \geqslant 0$. Ainsi $E(T_n^r)$ existe si et seulement si la série de terme général $e^{-\frac{rqk}{n}} P(S_n = k)$ converge.

$X_1, X_2, \dots, X_n$ sont indépendantes et suivent la loi de Poisson de paramètre λ.

Alors $S_n = X_1 + X_2 + \cdots + X_n$ suit la loi de Poisson de paramètre $n\lambda$.

$\forall k \in \mathbb{N}$, $e^{-\frac{rq}{n}k} P(S_n = k) = e^{-\frac{rq}{n}k} \dfrac{(\lambda n)^k}{k!} e^{-n\lambda} = e^{-n\lambda} \dfrac{\left(e^{-\frac{rq}{n}} \lambda n\right)^k}{k!}$.

Or pour tout x dans $\mathbb{R}$, la série de terme général $\dfrac{x^k}{k!}$ converge et $\sum_{k=0}^{+\infty} \dfrac{x^k}{k!} = e^x$.

Ainsi la série de terme général $e^{-\frac{rq}{n}k} P(S_n=k)$ converge et :

$$\sum_{k=0}^{+\infty} e^{-\frac{rq}{n}k} P(S_n=k) = e^{-n\lambda} e^{e^{-\frac{rq}{n}}(n\lambda)} = e^{n\lambda\left(e^{-\frac{rq}{n}}-1\right)}.$$

Donc $E(T_n^r)$ existe et vaut $e^{n\lambda\left(e^{-\frac{rq}{n}}-1\right)}$ et ceci pour tout $r\in\mathbb{N}$.

Soit $r\in\mathbb{N}^*$. $n\lambda\left(e^{-\frac{rq}{n}}-1\right) \underset{n\to+\infty}{\sim} n\lambda\left(-\frac{rq}{n}\right) = -rq\lambda$. Donc $\lim\limits_{n\to+\infty} E(T_n^r) = e^{-rq\lambda}$.

... ce qui est encore vrai pour $r=0$.

$E(T_n)$ existe et vaut $e^{n\lambda\left(e^{-\frac{q}{n}}-1\right)}$.

$E(T_n) = e^{-q\lambda} \iff e^{-\frac{q}{n}}-1 = -\frac{q\lambda}{n\lambda} \iff e^{-\frac{q}{n}}-1 = -\frac{q}{n}$.

Or $\forall x\in\mathbb{R}$, $e^x-1\geq x$ avec égalité si et seulement si $x=1$.

Donc $e^{-\frac{q}{n}}-1 = -\frac{q}{n} \iff -\frac{q}{n}=1 \iff n=-q$!

Donc T_n n'est pas un estimateur sans biais de $e^{-q\lambda}$. Cependant $\lim\limits_{n\to+\infty} E(T_n) = e^{-q\lambda}$

donc T_n est un estimateur asymptotiquement sans biais de $e^{-q\lambda}$.

Rappelons que T_n possède un moment d'ordre 2 et posons $\theta = e^{-q\lambda}$.

Alors $(T_n-\theta)^2$ possède une espérance. Soit $\varepsilon\in\mathbb{R}_+^*$. Markov donne :

$$0\leq P(|T_n-\theta|\geq\varepsilon)\leq \frac{E((T_n-\theta)^2)}{\varepsilon^2} = \frac{V(T_n-\theta)+(E(T_n-\theta))^2}{\varepsilon^2} = \frac{1}{\varepsilon^2}\left[V(T_n)+(E(T_n)-\theta)^2\right]$$

Or $\lim\limits_{n\to+\infty} V(T_n) = \lim\limits_{n\to+\infty}\left[E(T_n^2)-(E(T_n))^2\right] = e^{-2q\lambda}-(e^{-q\lambda})^2 = 0$ et

$\lim\limits_{n\to+\infty}(E(T_n)-\theta)^2 = \lim\limits_{n\to+\infty}\left(E(T_n)-e^{-q\lambda}\right)^2 = 0$.

Alors par encadrement il vient : $\lim\limits_{n\to+\infty} P(|T_n-\theta|\geq\varepsilon) = 0$ et ceci pour tout ε dans $\mathbb{R}_+^*$.

Alors « T_n » converge en probabilité vers $\theta = e^{-q\lambda}$.

T_n est un estimateur convergent de $e^{-q\lambda}$.