Correction du TD 17

Estimation

Correction de l'exercice 4 - EDHEC 14

1. On a déjà, de manière évidente : $u_k \geqslant 0$ pour tout $k \in \mathbb{N}$. Il reste à vérifier que la série $\sum u_k$ converge et que sa somme vaut 1. Or, la suite (u_k) est une suite géométrique de raison $\dfrac{\theta}{1+\theta}$ et $0 \leqslant \dfrac{\theta}{1+\theta} < 1$ (car $0 \leqslant \theta < \theta + 1$). On en déduit que la série $\sum u_k$ converge. De plus :

$$\sum_{k=0}^{+\infty} u_k = \frac{1}{1+\theta} \sum_{k=0}^{+\infty} \left(\frac{\theta}{1+\theta}\right)^k = \frac{1}{1+\theta} \times \frac{1}{1-\frac{\theta}{1+\theta}} = \frac{1}{(1+\theta)-\theta} = 1$$

Par conséquent, (u_k) définit une loi de probabilité.

2. La variable aléatoire Y est à valeurs dans $\mathbb{N}^*$ et, pour tout $k \in \mathbb{N}^*$:

$$P(Y=k) = P(X+1=k) = P(X=k-1) = \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{k-1} = \frac{1}{1+\theta}\left(1-\frac{1}{1+\theta}\right)^{k-1}$$

On en déduit que Y suit une loi géométrique de paramètre $p = \dfrac{1}{1+\theta}$.

On sait alors que $E(Y) = \dfrac{1}{p} = 1+\theta$. Comme $X = Y-1$, on en déduit que X admet une espérance et que $E(X) = E(Y) - 1 = \theta$.

De même, on sait que $V(Y) = \dfrac{1-p}{p^2}$, c'est-à-dire :

$$V(Y) = \frac{1-\frac{1}{1+\theta}}{\left(\frac{1}{1+\theta}\right)^2} = \frac{\frac{\theta}{1+\theta}}{\left(\frac{1}{1+\theta}\right)^2} = \theta(1+\theta)$$

Et comme $X = Y-1$, on en déduit que X admet une variance et que $V(X) = V(Y) = \theta(1+\theta)$.

3. (a) Comme $x_1, \dots, x_n$ sont des éléments de $X(\Omega)$, on a $P(X_k = x_k) > 0$ pour tout $k \in [\![1;n]\!]$. D'où :

$$\begin{aligned}
\ln(L(\theta)) &= \ln\left[\prod_{k=1}^{n} P(X_k = x_k)\right] \\
&= \sum_{k=1}^{n} \ln\Big(P(X_k = x_k)\Big) \\
&= \sum_{k=1}^{n} \ln\left[\frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{x_k}\right] \\
&= \sum_{k=1}^{n} \ln\left[\frac{\theta^{x_k}}{(1+\theta)^{x_k+1}}\right] \\
&= \sum_{k=1}^{n} \Big[x_k \ln(\theta) - (x_k+1)\ln(1+\theta)\Big] \\
&= \ln(\theta)\sum_{k=1}^{n} x_k - \ln(1+\theta)\sum_{k=1}^{n}(x_k+1)
\end{aligned}$$

Ainsi, $\ln(L(\theta)) = \ln(\theta) S_n - \ln(1+\theta)(S_n + n)$.

(b) La fonction φ est dérivable sur $]0;+\infty[$ comme somme de fonctions dérivables sur $]0;+\infty[$ et, pour tout $\theta > 0$, on a :

$$\begin{aligned}\varphi'(\theta) &= \frac{S_n}{\theta} - \frac{S_n+n}{1+\theta}\\ &= \frac{S_n(1+\theta)-\theta(S_n+n)}{\theta(1+\theta)}\\ &= \frac{S_n - n\theta}{\theta(1+\theta)}\end{aligned}$$

Or, $\theta(1+\theta) > 0$. Donc $\varphi'(\theta)$ est du signe de $S_n - n\theta$ sur $]0;+\infty[$. D'où le tableau de variations de φ :

x	0		$\frac{S_n}{n}$		$+\infty$
$\varphi'(x)$		$+$	0	$-$	
$\varphi(x)$		$\nearrow$	$\varphi\left(\frac{S_n}{n}\right)$	$\searrow$	

La fonction φ admet un maximum sur $]0;+\infty[$, atteint en un seul réel qui est $\widehat{\theta_n} = \dfrac{S_n}{n}$.
Comme $L(\theta) = \exp(\varphi(\theta))$ et que la fonction exp est strictement croissante sur $\mathbb{R}$, la fonction L a les mêmes variations que la fonction φ sur $]0;+\infty[$.
La fonction L admet également un maximum sur $]0;+\infty[$, qui est également atteint (uniquement) en $\widehat{\theta_n} = \dfrac{S_n}{n}$.

4. (a) Les variables aléatoires $(X_1,\dots,X_n)$ sont iid et T_n ne dépend pas de θ. Donc T_n est un estimateur de θ.

(b) La variable aléatoire T_n admet une espérance car $X_1,\dots,X_n$ admettent des espérances (elles ont la même loi que X et X admet une espérance). De plus, par linéarité de l'espérance :

$$\begin{aligned}E(T_n) &= \frac{1}{n}\sum_{k=1}^{n} E(X_k)\\ &= \frac{1}{n}\sum_{k=1}^{n} E(X) \quad \text{car } X_1,\dots,X_n \text{ ont la même loi que } X\\ &= \frac{1}{n}\times nE(X)\\ &= E(X)\\ &= \theta \quad \text{(question 2.a)}\end{aligned}$$

On en déduit que T_n est un estimateur sans biais de θ. Comme T_n est un estimateur sans biais, son risque quadratique est égal à sa variance : $r_{T_n}(\theta) = V(T_n)$ (sous réserve d'existence de cette variance).

Les variables aléatoires $X_1,\dots,X_n$ admettent des variances (elles ont la même loi que X et X admet une variance), et sont indépendantes. On en déduit que T_n admet une variance et :

$$\begin{aligned}
V(T_n) &= V\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \\
&= \frac{1}{n^2} V\left[\sum_{k=1}^{n} X_k\right] \\
&= \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) \quad \text{par indépendance de } X_1,\dots,X_n \\
&= \frac{1}{n^2}\sum_{k=1}^{n} V(X) \quad \text{car } X_1,\dots,X_n \text{ ont la même loi que } X \\
&= \frac{1}{n^2} \times nV(X) \\
&= \frac{V(X)}{n} \\
&= \frac{\theta(1+\theta)}{n} \quad \text{(question 2.a)}
\end{aligned}$$

(c) Avec BT (T_n admet une variance), on a pour tout $\varepsilon > 0$:

$$P(|T_n - E(T_n)| \geq \varepsilon) \leq \frac{V(T_n)}{\varepsilon^2},$$

donc

$$0 \leq P(|T_n - \theta| \geq \varepsilon) \leq \frac{\theta(1+\theta)}{n\varepsilon^2}$$

Par encadrement, $\lim_{n\to+\infty} P(|T_n - \theta| \geq \varepsilon) = 0$. Donc T_n converge en probabilité vers θ. Cela signifie que, plus n est grand, plus T_n prend des valeurs proche du paramètre θ à estimer presque sûrement.
