Université de Paris X | Année Universitaire 2008/2009
Licence Sciences Economiques | Statistiques

Corrigé du partiel du 08 décembre 2008

Exercice 1

1) $\bar{X}_n$ est bien un estimateur sans biais de μ, puisque

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \frac{n}{n}\ \mathbb{E}\left[X_i\right] = \mathbb{E}\left[X_i\right] = \mu$$

2)

$$\begin{aligned}
\mathbb{E}[T_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2)\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i\bar{X}_n + \sum_{i=1}^{n}\bar{X}_n^2\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n\sum_{i=1}^{n} X_i + n\bar{X}_n^2\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2\right] \\
&= \frac{1}{n}\ \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right] \\
&= \frac{1}{n}\left(\mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] - n\mathbb{E}\left[\bar{X}_n^2\right]\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2] - n\mathbb{E}\left[\bar{X}_n^2\right]\right)
\end{aligned}$$

Or, $\mathbb{E}[X_i^2] = var(X_i) + \mathbb{E}[X_i]^2 = \sigma^2 + \mu^2$ et $\mathbb{E}[\bar{X}_n^2] = var(\bar{X}_n) + \mathbb{E}[\bar{X}_n]^2 = \frac{\sigma^2}{n} + \mu^2$. Alors

$$\begin{aligned}
\mathbb{E}[T_n] &= \frac{1}{n}\left(n(\sigma^2+\mu^2) - n(\frac{\sigma^2}{n}+\mu^2)\right) \\
&= \frac{1}{n}\left(n\sigma^2 - \sigma^2\right) \\
&= \frac{n-1}{n}\ \sigma^2
\end{aligned}$$

3) A partir de la réponse à la question précédente, un estimateur sans biais de σ^2 est donc

$$\frac{n}{n-1}\ T_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Exercice 2

Soit un n-échantillon $(X_1, \ldots, X_n)$ de loi de densité

$$f_\theta(x) = \frac{x^2}{2\theta^3}\ \exp\left(-\frac{x}{\theta}\right)\mathbf{1}_{[0,+\infty[}(x)$$

avec $\theta > 0$

1) Par hypothèse, $\mathbb{E}(X) = 3\theta$. On peut alors dire que $\theta = \mathbb{E}\left(X\right)/3$ et comme un estimateur sans biais de l'espérance est la moyenne empirique (c'est-à-dire $\bar{X}_n$), alors on propose $\bar{X}_n/3$ comme estimateur sans biais de θ.

2) <u>Fonction de vraisemblance</u> :

$$L_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

avec $f_\theta(x_i) = \frac{x_i^2}{2\theta^3}\ \exp\left(-\frac{x_i}{\theta}\right)\mathbf{1}_{[0,+\infty[}(x_i)$

$$\begin{aligned}
L_\theta(x_1, \ldots, x_n) &= \prod_{i=1}^{n}\left(\frac{x_i^2}{2\theta^3}\ \exp\left(-\frac{x_i}{\theta}\right)\mathbf{1}_{[0,+\infty[}(x_i)\right) \\
&= \frac{\prod_{i=1}^{n} x_i^2}{2^n\theta^{3n}} \times \exp\left(-\frac{\sum_{i=1}^{n} x_i}{\theta}\right) \times \mathbf{1}_{[0,+\infty[}(x_1) \times \ldots \times \mathbf{1}_{[0,+\infty[}(x_n) \\
&= \frac{\prod_{i=1}^{n} x_i^2}{2^n\theta^{3n}} \times \exp\left(-\frac{\sum_{i=1}^{n} x_i}{\theta}\right) \times \mathbf{1}_{\{\min_{1\le i\le n} x_i \ge 0\}}
\end{aligned}$$

3) <u>Estimateur du maximum de vraisemblance</u> : sur le domaine $\mathcal{D} = \{\min_{1\le i\le n} x_i \ge 0\}$,

$$\ln L_\theta(x_1, \ldots, x_n) = -n\ln 2 - 3n\ln\theta + 2\sum_{i=1}^{n}\ln x_i - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

$$\frac{\partial \ln L_\theta(x_1, \ldots, x_n)}{\partial \theta} = -\frac{3n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i$$

$$\begin{aligned}
\frac{\partial \ln L_\theta(x_1, \ldots, x_n)}{\partial \theta} = 0 \quad &\Longleftrightarrow \quad -\frac{3n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i = 0 \\
&\Longleftrightarrow \quad \frac{1}{\theta^2} \left(-3n\theta + \sum_{i=1}^{n} x_i \right) = 0 \\
&\Longleftrightarrow \quad -3n\theta + \sum_{i=1}^{n} x_i = 0
\end{aligned}$$

Le point $t_n = \frac{1}{3n} \sum_{i=1}^{n} x_i$ est donc stationnaire, vérifions maintenant que c'est bien un maximum.

$$\frac{\partial^2 \ln L_\theta(x_1, \ldots, x_n)}{\partial \theta^2} = \frac{3n}{\theta^2} - \frac{2}{\theta^3} \sum_{i=1}^{n} x_i \quad \text{qui prend comme valeur en } t_n :$$

$$\frac{3n}{t_n^2} - \frac{2}{t_n^3} \sum_{i=1}^{n} x_i = \frac{1}{t_n^2} \left(3n - \frac{2}{t_n} \sum_{i=1}^{n} x_i \right) = \frac{1}{t_n^2} \left(3n - \frac{2}{\frac{1}{3n} \sum_{i=1}^{n} x_i} \sum_{i=1}^{n} x_i \right) = \frac{1}{t_n^2} \left(3n - 2 \times 3n \right) = -\frac{3n}{t_n^2} < 0$$

Le point t_n est bien un maximum. L'estimateur du maximum de vraisemblance est donc

$$T_n = \frac{1}{3n} \sum_{i=1}^{n} X_i$$

<u>L'estimateur T_n est-il sans biais ?</u>

$$\mathbb{E}(T_n) = \frac{1}{3n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{n}{3n} \; \mathbb{E}(X_i) = \frac{\mathbb{E}(X_i)}{3} = \frac{3\theta}{3} = \theta$$

Par conséquent, T_n est bien un estimateur sans biais de θ.

<u>L'estimateur T_n est-il convergent ?</u>

$$var(T_n) = \frac{1}{9n^2} \sum_{i=1}^{n} var(X_i) = \frac{n}{9n^2} \; var(X_i) = \frac{var(X_i)}{9n} = \frac{3\theta^2}{9n} = \frac{\theta^2}{3n}$$

Donc $\lim_{n \to +\infty} var(T_n) = 0$. Par conséquent, puisque l'estimateur T_n est un estimateur sans biais de θ et que $\lim_{n \to +\infty} var(T_n) = 0$, alors l'estimateur T_n est un estimateur convergent de θ.

L'estimateur T_n est-il efficace ?

$$\begin{aligned}
I_n(\theta) &= -\mathbb{E}\left[\frac{\partial^2 \ln L_\theta(X_1,\ldots,X_n)}{\partial\theta^2}\right] \\
&= -\mathbb{E}\left[\frac{3n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^n X_i\right] \\
&= -\frac{3n}{\theta^2} + \frac{2}{\theta^3}\ \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\
&= -\frac{3n}{\theta^2} + \frac{2}{\theta^3}\ \sum_{i=1}^n \mathbb{E}\left[X_i\right] \\
&= -\frac{3n}{\theta^2} + \frac{2n}{\theta^3}\ \mathbb{E}[X_i] \\
&= -\frac{3n}{\theta^2} + \frac{6n\theta}{\theta^3} \\
&= -\frac{3n}{\theta^2} + \frac{6n}{\theta^2} \\
&= \frac{3n}{\theta^2} \\
&= \frac{1}{var(T_n)}
\end{aligned}$$

Par conséquent, l'estimateur T_n est efficace.

4) Les réponses aux questions précédentes nous permettent de déterminer l'espérance et la variance de T_n : $\mathbb{E}(T_n) = \theta$ et $var(T_n) = \frac{\theta^2}{3n}$. D'après le TCL, on sait que

$$\frac{T_n - \theta}{\frac{\theta}{\sqrt{3n}}} \quad \text{a pour loi asymptotique la loi } \mathcal{N}(0,1)$$

$$\text{Ainsi, } \frac{\sqrt{3n}(T_n - \theta)}{\theta} \quad \text{a pour loi asymptotique } \mathcal{N}(0,1).$$

$$\text{Donc } \sqrt{3n}(T_n - \theta) \quad \text{a pour loi asymptotique } \mathcal{N}(0,\theta^2).$$

Exercice 3

On désire estimer la proportion de clients qui portent des lunettes.

1) Une estimation ponctuelle de la proportion p de personnes portant des lunettes est :

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i = \frac{260}{780} = \frac{1}{3}$$

2) La construction d'un intervalle de confiance pour la proportion p au niveau de confiance 0,95 nécessite plusieurs étapes qui sont présentées ci-dessous.

Etape 1 : Il faut d'abord trouver un estimateur ayant les "bonnes" propriétés. On propose $\bar{X}_n$ comme estimateur de p, avec $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, avec X_i qui suit une loi de Bernouilli de paramètre p.

$$X_i = \begin{cases} 1 & \text{si le client porte des lunettes} \\ 0 & \text{sinon} \end{cases}$$

Il est nécessaire de vérifier que cet estimateur est un "bon" estimateur, autrement dit qu'il est un estimateur sans biais et convergent de p. $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_i) = p \Rightarrow \bar{X}_n$ est bien un estimateur sans biais de p. De plus, $\bar{X}_n$ est un estimateur convergent de p, car $\lim_{n\to+\infty} var(\bar{X}_n) = 0$. En effet,

$$var(\bar{X}_n) = \frac{var(X_i)}{n} = \frac{p(1-p)}{n}$$

Etape 2 : Il faut ensuite introduire Z, une nouvelle variable aléatoire, dite pivotale, qui dépend de $(X_1, ..., X_n)$ et du paramètre p et de loi indépendante de p.
D'après le TCL, la variable aléatoire

$$Z = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

converge vers une variable aléatoire de loi normale centrée réduite $\mathcal{N}(0{,}1)$, dès que $n > 30, np > 5$ et $n(1-p) > 5$. Puisqu'ici, n est suffisamment grand ($n > 30$), et que les variables aléatoires X_i sont indépendantes et de même loi de Bernouilli de paramètre p, les hypothèses du TCL sont satisfaites. On considère donc que Z suit asymptotiquement une loi $\mathcal{N}(0{,}1)$.

Etape 3 : On construit un intervalle de confiance pour la proportion p au niveau de confiance 0,95 à partir de la variable aléatoire Z.
On cherche a tel que $\mathbb{P}(-a \leq \bar{X}_n - p \leq a) = 0,95$ $\qquad (*)$

$$(*) \iff \mathbb{P}\left(\frac{-a}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{a}{\sqrt{\frac{p(1-p)}{n}}}\right) = 0,95$$

$$(*) \iff \mathbb{P}\left(\frac{-a\sqrt{n}}{\sqrt{p(1-p)}} \leq Z \leq \frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right) = 0,95$$

$$(*) \iff 2\ \Phi\left(\frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 = 0,95$$

$$
\begin{aligned}
(*) \quad &\Longleftrightarrow \quad \Phi\left(\frac{a\sqrt{n}}{\sqrt{p(1-p)}}\right) = \frac{1,95}{2} = 0,975 \\
(*) \quad &\Longleftrightarrow \quad \frac{a\sqrt{n}}{\sqrt{p(1-p)}} = \Phi^{-1}(0,975) \\
(*) \quad &\Longleftrightarrow \quad a = \sqrt{\frac{p(1-p)}{n}} \times \Phi^{-1}(0,975)
\end{aligned}
$$

où Φ est la fonction de répartition d'une loi normale centrée réduite.

L'objectif est de déterminer a à partir de l'expression précédente. Mais on ne connaît pas p. On se propose alors d'approximer p par $\bar{x}_n$ que l'on a calculé dans la première question de cet exercice. On sait également que 780 ménages ont été interrogés, autrement dit, $n = 780$.

$$
\begin{aligned}
a &= \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \times \Phi^{-1}(0,975) \\
a &= \sqrt{\frac{1/3 \times 2/3}{780}} \times \Phi^{-1}(0,975) \\
a &= \sqrt{\frac{2}{9} \times \frac{1}{780}} \times \Phi^{-1}(0,975) \\
a &= \frac{\Phi^{-1}(0,975)}{\sqrt{9 \times 390}} \\
a &= \frac{\Phi^{-1}(0,975)}{\sqrt{3510}} \\
a &= \frac{1,96}{\sqrt{3510}} \\
a &\simeq 0,03308
\end{aligned}
$$

Pour déterminer $\Phi^{-1}(0,975)$, on a lu dans la table de la loi normale centrée réduite la valeur t telle que $\mathbb{P}(Z \leq t) = 0,975$, avec $Z \sim \mathcal{N}(0,1)$, ce qui donne $t = 1,96$. Par conséquent, $\Phi^{-1}(0,975) = 1,96$.

L'intervalle de confiance pour p est $IC = [\bar{X}_n - a; \bar{X}_n + a]$. L'intervalle de confiance observé est défini par $IC_{obs} = [\bar{x}_n - a; \bar{x}_n + a]$. Par conséquent, dans cet exercice, l'intervalle de confiance pour p, au niveau de confiance 0,95, est égal à $[0,30025; 0,36642]$.

3) On cherche les niveaux de confiance $(1-\alpha)$ pour lesquels la longueur de l'intervalle

de confiance est inférieure ou égale à 0,05, ce qui s'écrit, connaissant la définition de a,

$$\begin{aligned}
\bar{x}_n + a - (\bar{x}_n - a) &\leq 0,05 \\
2a &\leq 0,05 \\
2\ \Phi^{-1}\left(\frac{1+(1-\alpha)}{2}\right)\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} &\leq 0,05 \\
2\ \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} &\leq 0,05 \\
\Phi^{-1}\left(1-\frac{\alpha}{2}\right) &\leq \frac{0,05}{2}\sqrt{\frac{n}{\bar{x}_n(1-\bar{x}_n)}} \\
\Phi^{-1}\left(1-\frac{\alpha}{2}\right) &\leq \frac{0,05}{2}\sqrt{\frac{780}{1/3 \times 2/3}} \\
\Phi^{-1}\left(1-\frac{\alpha}{2}\right) &\leq 1,48113
\end{aligned}$$

Donc, d'après la lecture de la table de la loi normale centrée réduite,

$$\begin{aligned}
1-\frac{\alpha}{2} &\leq 0,9306 \\
2-\alpha &\leq 1,8612 \\
1-\alpha &\leq 0,8612
\end{aligned}$$

En conclusion, pour que la longueur de l'intervalle de confiance soit inférieure ou égale à 0,05, le niveau de confiance doit être inférieur ou égal à 0,8612.

4) De même, on cherche n, la taille de l'échantillon, telle que

$$\begin{aligned}
2a &\leq 0,05 \\
2 \times 1,96\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} &\leq 0,05 \\
\sqrt{n} &\geq \frac{2 \times 1,96}{0,05}\sqrt{(\bar{x}_n(1-\bar{x}_n))} \\
n &\geq \frac{2^2 \times 1,96^2}{0,05^2}(\bar{x}_n(1-\bar{x}_n)) \\
n &\geq \frac{2^2 \times 1,96^2}{0,05^2} \times \frac{1}{3} \times \frac{2}{3} \\
n &\geq 1365,9022
\end{aligned}$$

Donc, pour que la longueur de l'intervalle de confiance soit inférieure ou égale à 0,05, la taille de l'échantillon doit être au minimum de 1366.