Université de Nice Sophia-Antipolis
L2 MASS
TD de statistique

Année Universitaire 2014/2015
Statistique

Correction détaillée du QCM 2

1. Rappel des résultats du cours

Définition 1. *Soit θ un paramètre. On appelle estimateur de θ toute variable aléatoire $\hat{\theta}_n$ de la forme*

$$\hat{\theta}_n = f(X_1, \ldots, X_n)$$

où f est une fonction qui ne dépend pas de θ.

Proposition 2 (Méthode des moments)**.** *Soit Q_θ une loi dépendant du paramètre θ. Soit $X_1, \ldots, X_n$ des v.a. i.i.d. de loi Q_θ. Si on connaît deux fonctions f et g telles que*

$$\mathbb{E}\left[f(X_1)\right] = g(\theta),$$

alors on prend $\hat{\theta}_n$ comme estimateur où $\hat{\theta}_n$ est une solution (quand cela est possible) de

$$g(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Si g est inversible, alors cette solution est connue. C'est

$$\hat{\theta}_n = g^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right).$$

Définition 3. *Soit $\alpha \in]0,1[$. On appelle intervalle de confiance de niveau α (ou de coefficient de sécurité $1-\alpha$) pour le paramètre θ, tout intervalle aléatoire $I_{1-\alpha}(\mathbf{X})$ basé sur $\mathbf{X} = (X_1, ..., X_n)$ et ne dépendant pas du θ inconnu, telle que*

$$\mathbb{P}_\theta\left(\theta \in I_{1-\alpha}(\mathbf{X})\right) \geq 1-\alpha,$$

quelque soit la valeur du paramètre θ.

Théorème 4. *Soient m, σ deux réels. Soit $(X_i)_{i\geq 1}$ une suite de variables i.i.d. de loi $\mathcal{N}(m, \sigma^2)$. Alors,*

$$Z = \sqrt{n}\frac{\bar{X}_n - m}{\sigma} \sim \mathcal{N}(0,1),$$

où $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$.

Théorème 5. *Soient m, σ deux réels. Soit $(X_i)_{i\geq 1}$ une suite de variables i.i.d. de loi $\mathcal{N}(m, \sigma^2)$. Alors,*

$$V = \frac{n\widetilde{\sigma}_n^2}{\sigma} \sim \chi^2(n),$$

où $\widetilde{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - m)^2$.

2. Correction

La correction est divisée en trois sections qui correspondent aux deux exercices et au problème. Les notations définies sur le sujet ne sont pas rappelées ici. Les résultats qui faisaient partis des bonnes réponses à cocher sont soulignés.

2.1. **Exercice 1.**

Correction 1. La loi Q_θ est en fait la loi exponentielle de paramètre $\lambda = \theta^2$. Son espérance est connue. Son calcul a été fait en TD. On a $\mathbb{E}[X_1] = 1/\lambda = \underline{1/\theta^2}$.

Correction 2. On est dans le cadre de la méthode des moments (Proposition 2) avec $f = \mathrm{id}$ et $g : x \mapsto 1/x^2$. Or, la fonction g est inversible sur $\mathbb{R}^*$ et $g^{-1}(x) = 1/\sqrt{x}$. Ainsi, l'estimateur par méthode des moments est

$$\underline{\hat{\theta}_n = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i}} = \frac{1}{\sqrt{\bar{X}_n}}}.$$

Les autres réponses proposées sont donc automatiquement fausses.

2.2. **Exercice 2.**

Correction 3. La vraisemblance de l'échantillon $X_1, \ldots, X_n$ est donnée par

$$L_{X_1,\ldots,X_n}(\eta) = \prod_{i=1}^n \eta^2 X_i \exp(-\eta X_i) = \eta^{2n} \prod_{i=1}^n X_i \exp\left(-\eta \sum_{i=1}^n X_i\right).$$

Donc, la log-vraisemblance est donnée par

$$\ell_{X_1,\ldots,X_n}(\eta) = 2n\log(\eta) + \sum_{i=1}^n \log(X_i) + \left(-\eta \sum_{i=1}^n X_i\right).$$

Sa dérivée par rapport à η est égale à

$$\frac{d\ell_{X_1,\ldots,X_n}}{d\eta}(\eta) = \frac{2n}{\eta} + 0 - \sum_{i=1}^n X_i.$$

Donc, sa dérivée est positive si et seulement si $\frac{2n}{\eta} \geq \sum_{i=1}^n X_i$, i.e. $\eta \leq 2/\bar{X}_n$. Donc, <u>$\hat{\eta}_n = 2/\bar{X}_n$ est l'estimateur du maximum de vraisemblance de η</u>. Les autres réponses proposées sont donc automatiquement fausses.

2.3. **Problème.**

Correction 4. D'après l'énoncé, la loi de X est la loi normale $\mathcal{N}(\mu, s^2)$. C'est une loi à densité. Celle-ci est connue, c'est

$$\underline{f : x \mapsto \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(x-\mu)^2}{2s^2}\right).}$$

Correction 5. Une application du Théorème 4 permet de dire que $Z \sim \mathcal{N}(0,1)$.

Correction 6. Soit $\alpha < \beta \in]0,1[$, on a

$$\begin{aligned}
\mathbb{P}(t_\beta \leq Z \leq t_\alpha) &= 1 - \mathbb{P}(Z > t_\alpha) - \mathbb{P}(Z < t_\beta) \\
&= \mathbb{P}(Z \leq t_\alpha) - \mathbb{P}(Z < t_\beta) \\
&= \mathbb{P}(Z \leq t_\alpha) - \mathbb{P}(Z \leq t_\beta) \quad \text{(car la loi de } Z \text{ est continue.)} \\
&= 1 - \alpha - 1 - \beta = \underline{\beta - \alpha.}
\end{aligned}$$

De plus, en prenant $\beta = 1-\alpha$, on en déduit que $\mathbb{P}\left(t_{1-\alpha} \leq Z \leq t_\alpha\right) = 1 - 2\alpha$. Or, 2α n'est pas égal à $1-2\alpha$ pour tout $\alpha \in]0, 1/2[$, donc une des inégalités est fausse. Pour finir, en utilisant encore une fois le fait que la loi de Z est continue, on a $\mathbb{P}\left(Z \geq t_\alpha\right) = 1 - \mathbb{P}\left(Z \leq t_\alpha\right) = 1 - \alpha \neq \alpha/2$. Et donc, la dernière inégalité est fausse.

Correction 7. Soit $\alpha < \beta \in]0, 1[$, on a

$$\begin{aligned}
\mathbb{P}\left(t_\beta \leq Z \leq t_\alpha\right) &= \mathbb{P}\left(t_\beta \leq \sqrt{n}\frac{\bar{X}_n - \mu}{s} \leq t_\alpha\right) \\
&= \mathbb{P}\left(\frac{t_\beta s}{\sqrt{n}} \leq \bar{X}_n - \mu \leq \frac{t_\alpha s}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu \leq \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}\right).
\end{aligned}$$

Or, $\mathbb{P}\left(t_\beta \leq Z \leq t_\alpha\right) = \beta - \alpha$ par la question précédente. En remarquant que $\beta - \alpha = 1 - (1 - \beta + \alpha)$, on en déduit que $\mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu \leq \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}\right) = 1 - (1-\beta+\alpha)$. Par la Définition 3, ceci implique que $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}]$ est un intervalle de confiance de niveau $1 - \beta + \alpha$ pour μ. En particulier, si on prend $\beta = 1 - \alpha$, on trouve que $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, \bar{X}_n - \frac{t_{1-\alpha} s}{\sqrt{n}}]$ est un IC de niveau 2α pour μ. En revanche, comme $1 - (1 - \beta + \alpha) < 1 - (1 - \beta - \alpha)$, on a

$$\mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu \leq \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}\right) = 1 - (1 - \beta + \alpha) < 1 - (1 - \beta - \alpha),$$

donc $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}]$ n'est pas un IC de niveau $1 - \beta - \alpha$ pour μ. Et de même,

$$\mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu \leq \bar{X}_n - \frac{t_\beta s}{\sqrt{n}}\right) = 1 - 2\alpha < 1 - \alpha,$$

donc $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, \bar{X}_n - \frac{t_{1-\alpha} s}{\sqrt{n}}]$ n'est pas un IC de niveau α pour μ.

Soit $\alpha \in]0, 1[$, on a

$$\begin{aligned}
1 - \alpha = \mathbb{P}\left(Z \leq t_\alpha\right) &= \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{s} \leq t_\alpha\right) \\
&= \mathbb{P}\left(\bar{X}_n - \mu \leq \frac{t_\alpha s}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu\right).
\end{aligned}$$

Ainsi, on a bien que $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, +\infty[$ est un IC de niveau α pour μ. Or, $1 - \alpha \geq 1 - 2\alpha$, donc $\mathbb{P}\left(\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}} \leq \mu\right) \geq 1 - 2\alpha$ ce qui est équivalent à dire que $[\bar{X}_n - \frac{t_\alpha s}{\sqrt{n}}, +\infty[$ est un IC de niveau 2α pour μ.

Correction 8. Une application du Théorème 5 permet de dire que $V \sim \chi^2(n)$.

Correction 9. Le travail effectué pour répondre à la question 6 ne dépend pas de la loi considérée. On utilise uniquement le fait que Z admet une loi continue. Comme V admet également une loi continue, les résultats sont identiques.

Correction 10. Soit $\alpha \in]0,1[$, on a

$$
\begin{aligned}
\mathbb{P}\left(t_{1-\alpha} \leq V \leq t_\alpha\right) &= \mathbb{P}\left(t_{1-\alpha} \leq \frac{n\widetilde{s}_n^2}{s^2} \leq t_\alpha\right) \\
&= \mathbb{P}\left(\frac{t_{1-\alpha}}{n\widetilde{s}_n^2} \leq \frac{1}{s^2} \leq \frac{t_\alpha}{n\widetilde{s}_n^2}\right) \\
&= \mathbb{P}\left(\frac{n\widetilde{s}_n^2}{t_\alpha} \leq s^2 \leq \frac{n\widetilde{s}_n^2}{t_{1-\alpha}}\right).
\end{aligned}
$$

D'après la question précédente, on a $\mathbb{P}\left(t_{1-\alpha} \leq V \leq t_\alpha\right) = 1-2\alpha$. En se rappelant la Définition 3, on peut dire que $[\frac{n\widetilde{s}_n^2}{t_\alpha}, \frac{n\widetilde{s}_n^2}{t_{1-\alpha}}]$ est un intervalle de confiance de niveau 2α pour s^2. En revanche, comme $1-2\alpha < 1-\alpha$ ou $1-\alpha/2$, on en déduit que I n'est pas un IC de niveau α ou $\alpha/2$ pour s^2.

Soit $\alpha \in]0,1[$, on a

$$
\begin{aligned}
1-\alpha = \mathbb{P}\left(V \leq t_\alpha\right) &= \mathbb{P}\left(\frac{n\widetilde{s}_n^2}{s^2} \leq t_\alpha\right) \\
&= \mathbb{P}\left(\frac{1}{s^2} \leq \frac{t_\alpha}{n\widetilde{s}_n^2}\right) \\
&= \mathbb{P}\left(\frac{n\widetilde{s}_n^2}{t_\alpha} \leq s^2\right).
\end{aligned}
$$

En particulier, on peut dire que $J = [\frac{n\widetilde{s}_n^2}{t_\alpha}, +\infty[$ est un IC de niveau α pour s^2. Comme $1-\alpha \geq 1-2\alpha$, on a $\mathbb{P}\left(\frac{n\widetilde{s}_n^2}{t_\alpha} \leq s^2\right) \geq 1-2\alpha$ et donc J est un IC de niveau 2α pour s^2. En revanche, comme $1-\alpha < 1-\alpha/2$, on en déduit que J n'est pas un IC de niveau $\alpha/2$ pour s^2.

Pour finir, soit $\alpha \in]0,1[$, on a

$$
\begin{aligned}
\mathbb{P}\left(V \geq t_{1-\alpha}\right) &= \mathbb{P}\left(\frac{n\widetilde{s}_n^2}{s^2} \geq t_{1-\alpha}\right) \\
&= \mathbb{P}\left(\frac{t_{1-\alpha}}{n\widetilde{s}_n^2} \leq \frac{1}{s^2}\right) \\
&= \mathbb{P}\left(s^2 \leq \frac{n\widetilde{s}_n^2}{t_{1-\alpha}}\right).
\end{aligned}
$$

Or, $\mathbb{P}\left(V \geq t_{1-\alpha}\right) = 1 - \mathbb{P}\left(V \leq t_{1-\alpha}\right) = 1-\alpha$, donc $[0, \frac{n\widetilde{s}_n^2}{t_{1-\alpha}}]$ est un IC de niveau α pour s^2 (cette affirmation n'était pas proposée). En revanche, comme $1-\alpha < 1-\alpha/2$, on peut en déduire que $[0, \frac{n\widetilde{s}_n^2}{t_{1-\alpha}}]$ n'est pas de niveau $\alpha/2$ pour s^2.

Correction 11. Sur la table de la loi normale centrée réduite, on lit

$$
\mathbb{P}\left(\mathcal{N}(0,1) \leq 1,96\right) = 0,975 \quad \text{et} \quad \mathbb{P}\left(\mathcal{N}(0,1) \leq 1\right) \approx 0,8413.
$$

On en déduit facilement que

$$
\mathbb{P}\left(\mathcal{N}(0,1) \leq 1,96\right) \leq 0,975 \quad \text{et} \quad \mathbb{P}\left(-1,96 \leq \mathcal{N}(0,1) \leq 1,96\right) = 1-2(1-0,975) = 0,95 \geq 0,95.
$$

et

$$
\mathbb{P}\left(\mathcal{N}(0,1) \geq 1,96\right) \approx 1-0,975 < 0,03 \quad \text{et} \quad \mathbb{P}\left(-1 \leq \mathcal{N}(0,1) \leq 1\right) \approx 1-2(1-0,8413) > 0,5.
$$

Sur la table de la loi du chi deux, on lit pour $\nu = 10$

$$\mathbb{P}\left(\chi^2(10) \geq 2,156\right) = 0,995, \quad \mathbb{P}\left(\chi^2(10) \geq 25,188\right) = 0,005 \quad \text{et} \quad \mathbb{P}\left(\chi^2(10) \geq 15,987\right) = 0,10,$$

et pour $\nu = 20$

$$\mathbb{P}\left(\chi^2(20) \geq 10,851\right) = 0,95 \quad \text{et} \quad \mathbb{P}\left(\chi^2(20) \geq 12,443\right) = 0,90.$$

On en déduit facilement que

$$\begin{cases} \mathbb{P}\left(\chi^2(20) \leq 11\right) \geq \mathbb{P}\left(\chi^2(20) \leq 10,851\right) = 1 - 0,95 = 0,05 \\ \mathbb{P}\left(2,156 \leq \chi^2(10) \leq 15,987\right) = (1 - 0,10) - (1 - 0,995) = 0,895 \geq 0,85. \end{cases}$$

et

$$\begin{cases} \mathbb{P}\left(\chi^2(20) \geq 10\right) \geq \mathbb{P}\left(\chi^2(20) \geq 10,851\right) = 0,95 \\ \mathbb{P}\left(2 \leq \chi^2(10) \leq 26\right) \geq \mathbb{P}\left(2,156 \leq \chi^2(10) \leq 25,188\right) = 0,995 - 0,005 > 0,98. \end{cases}$$