Corrigé examen janv. 2011, première session

Exercice 1

1) $f_\theta(x) = F'_\theta(x) = \frac{c}{\theta} x^{c-1} \exp\left(-\frac{x^c}{\theta}\right) \mathbb{1}_{x>0}$

2) $L_n(\theta) = \prod_{i=1}^n f_\theta(X_i) = \frac{c^n}{\theta^n} \left(\prod_{i=1}^n X_i\right)^{c-1} \cdot \exp\left(-\frac{1}{\theta}\sum_{i=1}^n X_i^c\right)$

$\log L_n(\theta) = n \log c - n \log \theta + (c-1)\sum_{i=1}^n \log X_i - \frac{1}{\theta}\sum_{i=1}^n X_i^c$

$\frac{\partial \log L_n(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n X_i^c = 0 \Rightarrow \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i^c$

3) $G(y) = \mathbb{P}[Y \le y] = \mathbb{P}\left[2\frac{X^c}{\theta} \le y\right] = \mathbb{P}\left[X^c \le \frac{\theta}{2} y\right] = F_\theta\left(\left(\frac{\theta}{2}y\right)^{1/c}\right)$,

$= \left[1 - \exp\left(-\frac{1}{\theta}\frac{\theta}{2}y\right)\right]\mathbb{1}_{y>0} = \left[1 - \exp\left(-\frac{y}{2}\right)\right]\mathbb{1}_{y>0}$

$g(y) = G'(y) = \frac{1}{2}\exp\left(-\frac{y}{2}\right)\mathbb{1}_{y>0}$

4) $\mathbb{E}[Y] = \int_{\mathbb{R}} y g(y)\,dy = \frac{1}{2}\int_0^\infty y \exp\left(-\frac{y}{2}\right)dy = \left[-y\exp\left(-\frac{y}{2}\right)\right]_0^\infty +$

$+ \int_0^\infty \exp\left(-\frac{y}{2}\right)dy = \left[-2\exp\left(-\frac{y}{2}\right)\right]_0^\infty = 2$

$\mathbb{E}[Y^2] = \int_{\mathbb{R}} y^2 g(y)\,dy = \frac{1}{2}\int_0^\infty y^2 \exp\left(-\frac{y}{2}\right)dy = \left[-y^2\exp\left(-\frac{y}{2}\right)\right]_0^\infty$

$+ 2\int_0^\infty y\exp\left(-\frac{y}{2}\right)dy = 4\mathbb{E}[Y] = 8$

$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = 8 - 4 = 4$

5) $Y = \frac{2}{\theta}X^c \Rightarrow \mathbb{E}(X^c) = \frac{\theta}{2}\mathbb{E}(Y) = \theta$

$\hat{\theta}_n \xrightarrow[n\to\infty]{p.s} \mathbb{E}(X^c) = \theta \Rightarrow \hat{\theta}_n$ fortement conv.

$\mathbb{E}(\hat{\theta}_n) = \mathbb{E}(X^c) = \theta \Rightarrow \hat{\theta}_n$ sans biais

$\mathrm{Var}(\hat{\theta}_n) = \frac{1}{n}\mathrm{Var}(X^c) = \frac{1}{n}\cdot\frac{\theta^2}{4}\cdot\mathrm{Var}(Y) = \frac{\theta^2}{n}$

$I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log L_n(\theta)\right] = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\mathbb{E}\left(\sum_{i=1}^n X_i^c\right)$

$= -\frac{n}{\theta^2} + \frac{2n}{\theta^3}\theta = \frac{n}{\theta^2} \Rightarrow \mathrm{Var}(\hat{\theta}_n) = \frac{1}{I_n(\theta)} \Rightarrow \hat{\theta}_n$ efficac

(2)

6) $\hat{\theta}_n$ estimateur sans biais, $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^c = \frac{\theta}{2}\overline{Y_n}$

$1-\alpha = \mathbb{P}[\tilde{a} \le \hat{\theta}_n \le \tilde{b}] = \mathbb{P}[\tilde{a} \le \frac{\theta}{2n}\sum_{i=1}^{n} y_i \le \tilde{b}]$

Il faut trouver la loi de $\sum_{i=1}^{n} y_i$

Soit : $z = Y_1 + Y_2$ et h_1 sa densité, ~~φ une fonction bornée~~

~~$E[\varphi(z)]$~~ La densité h_1 est la convolution entre g et g :

$$h_1(z) = (g*g)(z) = \int_{\mathbb{R}} g(t)\cdot g(z-t)\,dt = \int_{\mathbb{R}} \frac{1}{2^2}\exp\left(-\frac{t}{2}\right)\exp\left(-\frac{z-t}{2}\right) 1\!\!1_{t>0}\, 1\!\!1_{z-t}$$

$$= \frac{1}{2^2}\int_0^z \exp\left(-\frac{z}{2}\right)dt = \frac{1}{2^2}\cdot z\exp\left(-\frac{z}{2}\right) 1\!\!1_{z>0}$$

Soit h_2 la densité de $Y_1+Y_2+Y_3$:

$$h_2(z) = (g*h_1)(z) = \int_{\mathbb{R}} g(t)\cdot h_1(z-t)\,dt$$

$$= \int_{\mathbb{R}} \frac{1}{2^3}\exp\left(-\frac{t}{2}\right)\cdot(z-t)\cdot\exp\left(-\frac{z-t}{2}\right) 1\!\!1_{t>0}\; 1\!\!1_{z-t>}\,dt$$

$$= \frac{1}{2^3}\exp\left(-\frac{z}{2}\right)\int_0^z (z-t)\,dt = \frac{1}{2^3}\cdot z^2\exp\left(-\frac{z}{2}\right) 1\!\!1_{z>0}$$

En général, la densité de $\sum_{i=1}^{n} Y_i$ est :

$$h(z) = \frac{1}{2^n}\cdot z^{n-1}\cdot\exp\left(-\frac{z}{2}\right) 1\!\!1_{z>0} \qquad (*)$$

Donc, la loi de $\sum_{i=1}^{n} Y_i$ est complètement connue, pas dépendante de θ. Donc la transformation cherchée est : $W_n = \sum_{i=1}^{n} Y_i$

$1-\alpha = \mathbb{P}[a \le W_n \le b]$ avec $\begin{cases} a = u_{\frac{\alpha}{2}} \text{ le fractile d'ordre } \frac{\alpha}{2} \text{ de la loi } (*) \\ b = u_{1-\frac{\alpha}{2}} \text{ le fractile d'ordre } (1-\frac{\alpha}{2}) \text{ de la loi } (*) \end{cases}$

$$1-\alpha = \mathbb{P}\left[u_{\frac{\alpha}{2}} \le \sum_{i=1}^{n} Y_i \le u_{1-\frac{\alpha}{2}}\right] = \mathbb{P}\left[u_{\frac{\alpha}{2}} \le \frac{2}{n\theta}\sum_{i=1}^{n} X_i^c \le u_{1-\frac{\alpha}{2}}\right]$$

③

$$1-\alpha = \mathbb{P}\left[\frac{2}{n}\sum_{i=1}^{n} X_i^c \cdot \frac{1}{u_{1-\frac{\alpha}{2}}} \leq \theta \leq \frac{2}{n}\sum_{i=1}^{n} X_i^c \cdot \frac{1}{u_{\frac{\alpha}{2}}} \right]$$

$$= \mathbb{P}\left[\underbrace{\frac{2}{u_{1-\frac{\alpha}{2}}} \hat{\theta}_n}_{A_n} \leq \theta \leq \underbrace{\frac{2}{u_{\frac{\alpha}{2}}} \hat{\theta}_n}_{B_n} \right]$$

7) Soit la v.a. $Z_i = X_i^c$ alors, $\hat{\theta}_n = \overline{Z}_n$

$\mathbb{E}(\hat{\theta}_n) = \mathbb{E}(Z) = \mathbb{E}(X^c) = \theta$

$Var(\hat{\theta}_n) = Var(X^c) = \theta^2$

Le TCL pour $(Z_1, \dots, Z_n)$:

$$\sqrt{n}\, \frac{\overline{Z}_n - \theta}{\theta} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Alors, asymptotiquement

$1-\alpha = \mathbb{P}\left[a \leq \sqrt{n}\, \frac{\overline{Z}_n - \theta}{\theta} \leq b \right]$ avec $a = u_{\frac{\alpha}{2}}$, $b = u_{1-\frac{\alpha}{2}}$ fractiles de $\mathcal{N}(0,1)$

$$= \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\, \frac{\overline{Z}_n}{\theta} - \sqrt{n} \leq u_{1-\frac{\alpha}{2}} \right]$$

$$= \mathbb{P}\left[\sqrt{n} - u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\, \frac{\overline{Z}_n}{\theta} \leq \sqrt{n} + u_{1-\frac{\alpha}{2}} \right]$$

$$= \mathbb{P}\left[1 - \frac{1}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \frac{\overline{Z}_n}{\theta} \leq 1 + \frac{1}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \right]$$

$$= \mathbb{P}\left[\underbrace{\frac{\overline{Z}_n}{1 + \frac{u_{1-\alpha/2}}{\sqrt{n}}}}_{A_n} \leq \theta \leq \underbrace{\frac{\overline{Z}_n}{1 - \frac{1}{\sqrt{n}} u_{1-\frac{\alpha}{2}}}}_{B_n} \right]$$

Puisque $n \to \infty$, $u_{1-\frac{\alpha}{2}} < \sqrt{n}$.

8) $\overline{X_n^c} = \mathbb{E}(X^c) \Rightarrow \overline{X_n^c} = \tilde{\theta} \Rightarrow \tilde{\theta}_n = \hat{\theta}_n = \overline{X_n^c} = \overline{Z_n}$

Exercice 2

1) $\hat{\theta}_n = \max_{1 \le i \le n} X_i$

2) $F_\theta(x) = \int_{-\infty}^{x} f_\theta(t)\,dt = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } 0 \le x \le \theta \\ 1 & \text{si } x \geq \theta \end{cases}$

Soit $G(y)$ la f ~~densité~~ répartition de $\hat{\theta}_n$ et $g(y)$ la densité.

$G(y) = \mathbb{P}[\max X_i \le y] = F_\theta^n(y)$

$g(y) = G'(y) = n f_\theta(y) F_\theta^{n-1}(y) = \frac{n}{\theta^n} x^{n-1} \mathbb{1}_{0 \le x \le \theta}$

3) $H_0 : \theta = 1 \qquad H_1 : \theta \neq 1.$

La vraisemblance : $L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = \frac{1}{\theta^n} \mathbb{1}_{0 \le \min X_i \le \max X_i}$

La vraisemblance sous H_0 :

$L_0 = L_n(1) = \mathbb{1}_{0 \le \min X_i \le \max X_i \le 1}$

La vraisemblance sous H_1 : $L_1 = L_n(\theta)$, $\theta \neq 1$

La zone de rejet, par le Lemme de Neyman-Pearson :

$R = \{ L_1 > k L_0 \} = \{ \frac{1}{\theta^n} \mathbb{1}_{0 \le \min X_i \le \max X_i \le \theta} > k \mathbb{1}_{0 \le \min \ldots} \}$

$= \{ \frac{1}{\theta^n} \mathbb{1}_{\max X_i \le \theta} > k \mathbb{1}_{\max X_i \le 1} \} = \{ \mathbb{1}_{\max X_i \le \theta} > k_1 \mathbb{1}_{\max_i \le 1} \}$

Soit $z = \mathbb{1}_{\max X_i \le \theta} \sim \mathcal{B}(F_0(\theta))$, G_0 la ~~de~~ f de rép pour $\theta = 1$.

Soit le rapport de vraisemblance :

$\lambda(X_1 \ldots X_n) = \frac{L_n(1)}{L_n(\hat{\theta}_n)} = \hat{\theta}_n$

La zone de rejet : $R = \{ \lambda(X_1, \ldots, X_n) < c \} = \{ \max_{1 \le i \le n} X_i < c \}$

t.q. : $\alpha = \mathbb{P}[\max X_i < c]$ alors la cte c est le fractile d'ordre α de la loi $G_0(y) = F_1^n(y)$

5

4) Le test de Wald ne peut pas être utilisé puisque l'information de Fisher n'existe pas.

Exercice 3 X : la distance jusqu'à l'objet $\sim N(m, \sigma^2)$

$n = 15$ $H_0 : m = 20$ $H_1 : m \neq 20$

Test sur la moyenne d'une loi Normale de variance inconnue.

Stat de test :

$$Z = \sqrt{n}\,\frac{\overline{X}_n - 20}{S_n^*} \underset{H_0}{\sim} t(14)$$

zone de rejet : $R = \{|Z| > u_{1-\frac{\alpha}{2}}\} = \{|Z| > 2,145\}$

valeur stat test :

$$z = \sqrt{15}\,\frac{19,13 - 20}{2,8} = -1,20 \notin R \Rightarrow H_0 \text{ acceptée}$$

Exercice 4

1) Ozone 1 ne suit pas une loi Normale, pendant que son log suit une loi N.

2) $n - 1 = 3 + 106 \Rightarrow n = 110$

3) Modèle de régression multiple :

Y : log(Ozone1), X_1 : ensolei, X_2 : vent, X_3 = temp

(1) $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

$\varepsilon_i \perp \varepsilon_j$ $i \neq j$, $i = 1, \ldots, n$

4) H_0 : (1) non signif $\Leftrightarrow$ la pollution n'est influencée par aucune des 3 var $\Leftrightarrow b_1 = b_2 = b_3 = 0$

$\Leftrightarrow$ Modèle réduit : $Y_i = b_0 + \varepsilon_i$

H_1 : (1) signif $\Leftrightarrow$ (1)

Stat de test : $Z = \dfrac{SM/3}{SR/106} \underset{H_0}{\sim} F(3, 106)$

$z = 72,91 \Rightarrow$ p-value $= 10^{-16} \Rightarrow H_0$ rejetée $\Rightarrow$ (1) signif.

5) $\hat{b}_0 = 0,26$ $\hat{b}_1 = 0,002$ $\hat{b}_2 = -0,069$, $\hat{b}_3 = 0,044$

$\hat{\sigma} = 0,1666$

6) H_0 : l'ensoleilment n'influe pas la pollution si le vent et la temp. sont dans le modèle
$\Leftrightarrow b_1 = 0 \mid X_2, X_3$ dans le modele
$\Leftrightarrow$ Modèle : $Y_i = b_0 + b_2 X_{2_i} + b_3 X_{3_i} + \varepsilon_i$

Stat de test : $z = \dfrac{\hat{B}_1}{\sqrt{\hat{Var}(B_1)}} \underset{H_0}{\sim} t(106)$

$z = 4{,}24 \Rightarrow$ p-value $= 10^{-5} \Rightarrow H_0$ rejetée

Pareil : le vent, la temp sont influentes si on garde dans le modèle les autres var.
On garde toutes les 3 var : X_1, X_2, X_3.

7) $R^2 = 0{,}67 \Rightarrow$ modèle d'assez bonne qualité