Corrigé examen du 18 janvier 2017 (M1)

Exercice 1

1) $$\mathbb{E}[X] = \frac{1+\theta}{1-\theta}\int_0^1 x^{\frac{1+\theta}{1-\theta}}\,dx = \frac{1+\theta}{2}\cdot\left[x^{\frac{2}{1-\theta}}\right]_0^1 = \frac{1+\theta}{2}$$

$$\mathbb{E}[X^2] = \frac{1+\theta}{1-\theta}\int_0^1 x^{\frac{2}{1-\theta}}\,dx = \frac{1+\theta}{3-\theta}\left.x^{\frac{3-\theta}{1-\theta}}\right]_0^1 = \boxed{\frac{1+\theta}{3-\theta}}$$

2) $$\frac{1+\theta}{2} = \bar{x}_n \Rightarrow \theta = 2\bar{x}_n - 1 \Rightarrow \boxed{\hat{\theta}_n^{(1)} = 2\bar{X}_n - 1}.$$

3) Par LFGN: $\bar{X}_n \xrightarrow{P.S} \mathbb{E}[X]$. Alors $\hat{\theta}_n^{(1)} \xrightarrow{P.S} 2\mathbb{E}[X]-1$

$\Rightarrow \hat{\theta}_n^{(1)} \xrightarrow{P.S} \theta \Rightarrow \hat{\theta}_n^{(1)}$ fortement consistant.

$\mathbb{E}[\hat{\theta}_n^{(1)}] = 2\mathbb{E}[\bar{X}_n] - 1 = 2\mathbb{E}[X] - 1 = \theta \Rightarrow \hat{\theta}_n^{(1)}$ sans biais

$$\mathrm{Var}[\hat{\theta}_n^{(1)}] = 4\,\mathrm{Var}[\bar{X}_n] = \frac{4}{n}\mathrm{Var}[X]$$

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1+\theta}{3-\theta} - \frac{(1+\theta)^2}{4}$$

$$= \frac{(1+\theta)}{4(3-\theta)}(4 - 3 - 2\theta + \theta^2) = \frac{(1+\theta)(1-\theta)^2}{4(3-\theta)}$$

$$\boxed{\mathrm{Var}[\hat{\theta}_n^{(1)}] = \frac{(1+\theta)(1-\theta)^2}{n(3-\theta)}}$$

4) $\hat{\theta}_n^{(1)} = 2\bar{X}_n - 1$. Par TCL on a que:

$$\sqrt{n}\,\frac{\bar{X}_n - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Définition de l'estimateur par intervalle asymp de niveau de confiance $1-\alpha$, pour θ:

trouver les statistiques A_n et B_n telles que

2

$$1-\alpha = \lim_{n\to\infty} \mathbb{P}[A_n \leq \theta \leq B_n]$$

Pour trouver A_n et B_n, on doit trouver une v.a Z_n (à partir de $\hat{\theta}_n$!!!) de loi connue asymptotiquement, indép de θ ; On considère alors : $Z_n = \sqrt{n}\,\dfrac{\overline{X}_n - \mathbb{E}[X]}{\sqrt{Var[X]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

$$\mathbb{E}[X] = \frac{1+\theta}{2}$$

Parce que $Var[X] = \dfrac{(1+\theta)(1-\theta)^2}{4(3-\theta)}$ et l'intervalle est asymptotique, on va prendre à la place de Z_n, la v.a. $\tilde{Z}_n = \sqrt{n}\,\dfrac{\overline{X}_n - \mathbb{E}[X]}{S_n^*}$

$$S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \xrightarrow{p.s} Var(X)$$

$Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ $\Rightarrow$

$\Rightarrow$ par le lemma de Slutsky : $\tilde{Z}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

Alors, il faut trouver a et b tels que :

$$1-\alpha = \lim_{n\to\infty} \mathbb{P}[a \leq \tilde{Z}_n \leq b] \Rightarrow \begin{cases} a = u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}} \\ b = u_{1-\alpha/2} \end{cases}$$

fractiles de $\mathcal{N}(0,1)$

$$= \lim_{n\to\infty} \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\overline{X}_n - \frac{1+\theta}{2}}{S_n^*} \leq u_{1-\frac{\alpha}{2}}\right]$$

$$= \lim_{n\to\infty} \mathbb{P}\left[-S_n^* u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\left(\overline{X}_n - \frac{1+\theta}{2}\right) \leq S_n^* u_{1-\frac{\alpha}{2}}\right]$$

$$= \lim_{n\to\infty} \mathbb{P}\left[2\left(\overline{X}_n - \frac{S_n^* u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) - 1 \leq \theta \leq 2\left(\overline{X}_n + \frac{S_n^* u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) - 1\right]$$

Donc $A_n = 2\left(\overline{X_n} - \frac{S_n^* u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) - 1$

$B_n = 2\left(\overline{X_n} + \frac{S_n^* u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) + 1$

5) $P[0<X<1] = F(1) - F(0) = \int_0^1 f(x)dx = 1 \Rightarrow X$ prend des valeurs entre 0 et 1

6) $L_n(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \left(\frac{1+\theta}{1-\theta}\right)^n \prod_{i=1}^{n} x_i^{\frac{2\theta}{1-\theta}} \mathbb{1}_{0 \leq x_i \leq 1}$

On a que l'ensemble: $A = \{x ; f_\theta(x) > 0\} =]0,1[$ ne dépend pas de θ, De plus $f_\theta(x)$ est dérivable 2 fois par rapport à θ sur A. Alors, sur A on peut considérer le $\log f_\theta(x)$ qui est aussi deux fois dérivable par rapport à θ. Trouver l'EMV revient à maximiser la log-vraisemblance:

$$\log L_n(\theta) = n\log(1+\theta) - n\log(1-\theta) + \frac{2\theta}{1-\theta}\sum_{i=1}^{n}\log X_i$$

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

$$\Rightarrow \hat{\theta}_n^{(2)} = \frac{1+\overline{Y_n}}{1-\overline{Y_n}} \text{, avec } Y_i = \log X_i$$

Vérification: $\frac{\partial^2}{\partial\theta^2}\log L_n(\theta) = \frac{2n\theta}{(1-\theta^2)^2} + \frac{4n\overline{Y_n}}{(1-\theta)^3}$

$$\Rightarrow \frac{\partial^2}{\partial\theta^2}\log L_n(\hat{\theta}) = \frac{2n\hat{\theta}^{(2)}}{\left(1-\hat{\theta}^{(2)2}\right)^2} + \frac{4n\overline{Y_n}}{\left(1-\hat{\theta}^{(2)}\right)^3}$$

4

$$= \frac{2n}{(1-\hat{\theta}_n)^2} \left[\frac{\hat{\theta}}{(1+\hat{\theta})^2} + \frac{2\overline{Y_n}}{(1-\hat{\theta})} \right]$$

$$= \frac{2n}{(1-\hat{\theta})^2} \frac{\hat{\theta}-\hat{\theta}^2+2\overline{Y_n}(1+\hat{\theta})^2}{(1-\hat{\theta})(1+\hat{\theta})^2} \to > 0$$

$$\hat{\theta}-\hat{\theta}^2+2\overline{Y_n}(1+\hat{\theta})^2 = \hat{\theta}(1-\hat{\theta})+2\overline{Y_n}(1+\hat{\theta})^2$$

$$= \frac{1+\overline{Y_n}}{1-\overline{Y_n}}\left(1-\frac{1+\overline{Y_n}}{1-\overline{Y_n}}\right)+2\overline{Y_n}\left(1+\frac{1+\overline{Y_n}}{1-\overline{Y_n}}\right)^2$$

$$= \frac{(1+\overline{Y_n})(-2\overline{Y_n})}{(1-\overline{Y_n})^2} + \frac{2\overline{Y_n}\cdot 4}{(1-\overline{Y_n})^2}$$

$$= \frac{-2\overline{Y_n}-2(\overline{Y_n})^2+8\overline{Y_n}}{(1-\overline{Y_n})^2} = \frac{6\overline{Y_n}-2(\overline{Y_n})^2}{(1-\overline{Y_n})^2}$$

$$= \frac{2\overline{Y_n}(3-\overline{Y_n})}{(1-\overline{Y_n})^2}$$

$$\left.\overline{Y}_n \in (-\infty, 0) \Rightarrow 3-\overline{Y_n} > 0 \right\} \Rightarrow \frac{\overline{Y_n}(3-\overline{Y_n})}{(1-\overline{Y_n})^2} \leq 0$$

$\Rightarrow \hat{\theta}_n^{(2)}$ point de max.

7) Soit $G(x)$ la fonction de répartition de Y et g la densité. Alors, $\forall x \in \mathbb{R}$:

$$G(x) = \mathbb{P}[Y \leq x] = \mathbb{P}[\log X \leq x] = \mathbb{P}[X \leq \exp(x)]$$

$= F(e^x)$, avec F la fonction de rép pour X.

$$g(x) = G'(x) = e^x f(e^x) = e^x \cdot \frac{1+\theta}{1-\theta} (e^x)^{\frac{2\theta}{1-\theta}} \mathbb{1}_{0 < e^x < 1}$$

$$g(x) = \frac{1+\theta}{1-\theta} e^{\left(\frac{2\theta}{1-\theta}+1\right)x} \mathbb{1}_{x<0}$$

$$= \frac{1+\theta}{1-\theta} e^{\frac{1+\theta}{1-\theta}x} \mathbb{1}_{x<0}.$$

8) $\mathbb{E}[Y] = \frac{1+\theta}{1-\theta}\int_{-\infty}^{0} x\, e^{\frac{1+\theta}{1-\theta}x}dx = x\, e^{\frac{1+\theta}{1-\theta}x}\Big]_{-\infty}^{0} - \int_{-\infty}^{0} e^{\frac{1+\theta}{1-\theta}x}dx$

$$= -\frac{1-\theta}{1+\theta} e^{\frac{1+\theta}{1-\theta}x}\Big]_{-\infty}^{0} = \frac{\theta-1}{\theta+1} < 0.$$

Par LFGN $\overline{Y}_n \xrightarrow{P.S} \mathbb{E}[Y]$. Alors:

$$\hat{\theta}_n^{(2)} \xrightarrow[n\to\infty]{P.S} \frac{1+\mathbb{E}[Y]}{1-\mathbb{E}[Y]} = \frac{1+\frac{\theta-1}{\theta+1}}{1-\frac{\theta-1}{\theta+1}} = \frac{2\theta}{2} = \theta$$

$\Rightarrow \hat{\theta}_n^{(2)}$ fortement convergent.

9) $h(\theta) = 1 - \frac{2}{1+\theta}$

$1 < 1+\theta < 2 \Rightarrow 1 < \frac{2}{1+\theta} < 2$ $\Big\} \Rightarrow h(\theta) \in (-1, 0)$

10) L'EMV de θ est $\hat{\theta}_n^{(2)} = \frac{1+\overline{Y}_n}{1-\overline{Y}_n}$. La fonction h prend des valeurs dans un intervalle borné, alors, l'EMV pour $\psi = h(\theta)$ est $\hat{\psi}_n = h(\hat{\theta}_n^{(2)}) = \overline{Y}_n$.

11) La définition d'une v. a. de type exponentiel: sa densité s'écrit sous la forme:

$$\exp(C(\theta)\cdot T(x) + D(\theta) + S(x)) = \exp(C(\theta)\cdot T(x) + D(\theta))\tilde{S}(x)$$

avec $C, T, D, S, \tilde{S}$ fonctions mesurables.

$$f_\theta(x) = \exp\left(\log\frac{1+\theta}{1-\theta} + \frac{2\theta}{1-\theta}\cdot \log x\right)\mathbb{1}_{0<x<1}$$

(6)

$\tilde{S}(x) = \mathbb{1}_{0<x<1}$, $D(\theta) = \log \frac{1+\theta}{1-\theta}$

$c(\theta) = \frac{2\theta}{1-\theta}$, $T(x) = \log x$.

Pour $g(x) = \exp\left(\frac{1+\theta}{1-\theta} x + \log \frac{1+\theta}{1-\theta}\right) \mathbb{1}_{x<0}$

$c(\theta) = \frac{1+\theta}{1-\theta}$, $T(x) = x$, $D(\theta) = \log \frac{1+\theta}{1-\theta}$, $\tilde{S}(x) = \mathbb{1}_{x<0}$

Parce que $f_\theta(x)$ est de type exponentiel, alors :

$\frac{1}{n} \sum_{i=1}^{n} T(x_i) = \overline{Y_n}$ est efficace et sans biais pour

$\mathbb{E}[T(X)] = \mathbb{E}[Y] = \frac{\theta-1}{\theta+1} = \psi$

Par LFGN :

$\left.\overline{Y_n} \longrightarrow \mathbb{E}[Y] \Rightarrow \overline{Y_n} \longrightarrow \psi\right\} \Rightarrow \overline{Y_n}$ fortement consistant.

$\hat{\psi}_n = \overline{Y_n}$

12)

$g(x) = \exp\left(-\frac{1}{\psi} x + \log\left(-\frac{1}{\psi}\right)\right) \mathbb{1}_{x<0}$

$\Rightarrow \sum_{i=1}^{n} T(Y_i) = \sum_{i=1}^{n} Y_i$ exhaustif pour ψ.

Donc $\overline{Y_n} = \hat{\psi}_n$ exhaustif pour ψ.

Exercice 2

1) $h(p) = \frac{p-1+1}{1-p} = \frac{1}{1-p} - 1 \Rightarrow h$ croissante.

2) $p_1 \leq p_2 \Rightarrow h(p_1) \leq h(p_2) \Rightarrow \frac{h(p_2)}{h(p_1)} \geq 1$

3) H_0 est composite donc on va étudier si la vrais L_n est monotone par rapport à une stat $T_n = T(X_1, \dots, X_n)$

$$L_n(p) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

$$\frac{L_n(p_2)}{L_n(p_1)} = \left(\frac{p_2}{p_1}\right)^{n\bar{x}_n}\left(\frac{1-p_2}{1-p_1}\right)^{n-n\bar{x}_n}$$

$$= \left(\frac{p_2}{1-p_2} \cdot \frac{1-p_1}{p_1}\right)^{n\bar{x}_n} \cdot \left(\frac{1-p_2}{1-p_1}\right)^{n}$$

$$= \left(\frac{h(p_2)}{h(p_1)}\right)^{n\bar{x}_n}\left(\frac{1-p_2}{1-p_1}\right)^{n}$$

$\frac{h(p_2)}{h(p_1)} \geq 1$ pour $p_2 \geq p_1$

$\Rightarrow L_n(p)$ croissante par rapport à $\bar{X}_n$

4) Pour tester H_0 contre H_1, la règle de décision suivante:

$$\delta(x_1, \dots, x_n) = \begin{cases} H_1 & \text{si } T_n < t_0 \\ H_0 & \text{si } T_n > t_0 \end{cases}$$

donne le test UPP :

$$\alpha(p) = \mathbb{P}[H_1 \mid X_i \sim \mathcal{B}(p),\ p \geq p_0]$$
$$= \mathbb{P}[T_n < t_0 \mid X_i \sim \mathcal{B}(p),\ p \geq p_0]$$

$$= \mathbb{P}\left[\overline{X}_n < t_0 \mid X_i \sim \mathcal{B}(p),\ p \geq p_0\right]$$

$$= \mathbb{P}\left[\sqrt{n}\ \frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} < c(p) \mid X_i \sim \mathcal{B}(p),\ p \geq p_0\right]$$

$$c(p) = \sqrt{n}\ \frac{t_0 - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}$$

$$Y_n = \sqrt{n}\ \frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Parce que $c(p) \leq c(p_0)$, avec proba 1, pour $p \geq p_0$, alors :

$$\alpha(p) = \mathbb{P}\left[Y_n < c(p)\right] \leq \mathbb{P}\left[Y_n < c(p_0)\right] = \alpha(p_0)$$

Alors $\alpha = \sup_{p \geq p_0} \alpha(p) = \alpha(p_0) = \mathbb{P}\left[\sqrt{n}\ \frac{\overline{X}_n - p_0}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} < c(p_0)\right]$

avc $c(p_0) = u_\alpha$ fractile de la loi $\mathcal{N}(0.1)$.

<u>Exercice 3</u>

X : longueur d'un oeuf $\sim \mathcal{N}(m, \sigma^2)$.

$1-\alpha = 0{,}95$.

La longueur d'un oeuf est de 20 mm : $m = 20$

$H_0 : m = 20$ $\quad$ $H_1 : m \neq 20$ $\quad$ $m_0 = 20$

On a un test sur la moyenne d'une loi Normale de variance inconnue.

Stat de test : $Z = \sqrt{n}\ \frac{\overline{X}_n - 20}{S_n^*} \underset{H_0}{\sim} t(n-1) = t(8)$

Zone de rejet :

$$R = \{|z| > u_{1-\frac{\alpha}{2}}\} = \{|z| > 2,306\}$$

$$z = \sqrt{9}\ \frac{20,1-20}{2,368} = 0,12 \notin R \Rightarrow H_0 \text{ acceptée}$$

$\Rightarrow$ La ~~dim~~ longueur est de 20 mm.

Exercice 4

1) Les degrés de libertés totaux sont $n-1 = 5+97$

$\Rightarrow \boxed{n = 103}$

2) On utilise un modèle de regression linéaire :

Y = FLOW, X_2 : Fly, X_3 : Water, X_4 : Coarse_A
X_5 : Fine_A X_1 : Slag.

(1) $Y_i = b_0 + b_1 X_{1i} + \cdots + b_5 X_{5i} + \varepsilon_i$, $i = 1, \ldots, n$
$\varepsilon_i \sim N(0, \sigma^2)$
$\varepsilon_i \perp\!\!\!\perp \varepsilon_j \quad i \neq j$.

3) H_0 : (1) non significatif : aucune var n'influe Y
$b_1 = \cdots = b_5 = 0 \Rightarrow$ Modèle réduit : (2) $Y_i = b_0 + \varepsilon_i$

H_1 : (1) signif : Y influencé par au moins une des 5 var : $\exists\, b_j \neq 0 \Rightarrow$ Modèle (1).

Stat de test : $Z = \dfrac{SM/5}{SR/97} \underset{H_0}{\sim} F(5,97)$

Valeur stat de test $z = 19,47$ p-value $= 2,2 \cdot 10^{-13} < 0,0$

$\Rightarrow H_0$ rejetée $\Rightarrow$ (1) signif.

4) H_0 : X_1 n'influe pas Y / $X_2, \ldots, X_5$ sont dans le modèle : $b_1 = 0 / \ldots$

Modèle réduit : (3) $y_i = b_2 + b_2 X_{2i} + \dots + b_5 X_{5i} + \varepsilon_i$

H_1 : X_1 influe Y | $X_2, \dots, X_5$ dans le modèle

$b_1 \neq 0$ | . . . .

Modèle : (1)

Stat de test : $Z = \dfrac{\hat{b}_1}{\sqrt{Var(b_1)}} \underset{H_0}{\sim} t(97)$

$z = -2{,}73$ p-value $= 0{,}007 \Rightarrow H_0$ rejetée

$\Rightarrow X_1$ influe Y si les autres var sont dans le modèle.

Var influentes : Slag, Water

Var non influentes : Fly, Coarse-A, Fine-A.

4) $\hat{b}_0 = -86$ $\hat{b}_1 = -0{,}07 \dots$ $\hat{b}_5 = 0{,}02$ $\hat{\sigma} = 12{,}73$.

L'écoulement diminue avec "scories" et augmente quand les autres var. augmentent.

5) $R^2 adj = 0{,}47$ modèle de qualité assez moyenne.