Corrigé examen 18 janvier 2012

rcice 1

, $\mathbb{E}(X) = \frac{\theta}{1-\theta}\int_0^1 x \cdot x^{\frac{2\theta-1}{1-\theta}}dx = \frac{\theta}{1-\theta}\int_0^1 x^{\frac{\theta}{1-\theta}}dx = \frac{\theta}{1-\theta}(1-\theta)\, x^{\frac{1}{1-\theta}}\Big]_0^1 = \theta$

$\mathbb{E}(X^2) = \frac{\theta}{1-\theta}\int_0^1 x^2 \cdot x^{\frac{2\theta-1}{1-\theta}}dx = \frac{\theta}{1-\theta}\int_0^1 x^{\frac{1}{1-\theta}}dx = \frac{\theta}{1-\theta}\cdot\frac{1-\theta}{2-\theta}\, x^{\frac{2-\theta}{1-\theta}}\Big|_0^1$

$= \frac{\theta}{2-\theta}$

$Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{\theta}{2-\theta} - \theta^2 = \theta\left(\frac{1}{2-\theta} - \theta\right)$

2) $\bar{X}_n = \mathbb{E}(X) \Rightarrow \hat{\theta}_n = \bar{X}_n$

) Par LFGN : $\bar{X}_n \xrightarrow{P.S} \mathbb{E}(X) = \theta \Rightarrow \hat{\theta}_n$ fort. conv.

$\mathbb{E}(\bar{X}_n) = \mathbb{E}(X) = \theta \Rightarrow \hat{\theta}_n$ ~~f~~ sans biais.

$\log f_\theta(x) = \log(\theta) - \log(1-\theta) + \frac{2\theta-1}{1-\theta}\log x$

$\frac{\partial \log f_\theta(x)}{\partial\theta} = \frac{1}{\theta} + \frac{1}{1-\theta} + \frac{2-2\theta+2\theta-1}{(1-\theta)^2}\log x$

$\frac{\partial^2 \log f_\theta(x)}{\partial\theta^2} = -\frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} + \frac{2}{(1-\theta)^3}\log x$

) La loi de $Y = \log X$

La f de rép $F(x)$ de la v.a. X est :

$F(x) = \int_0^x f(t)dt = \int_0^x \frac{\theta}{1-\theta}\, t^{\frac{2\theta-1}{1-\theta}}dt = \frac{\theta}{1-\theta}\cdot\frac{1-\theta}{\theta}\, t\Big|_0^x$

$$= \begin{cases} x^{\frac{\theta}{1-\theta}} & \text{si } 0 < x < 1 \\ 0 & \text{si } x < 0 \\ 1 & \text{si } x > 1 \end{cases}$$

$G(x) = P[\log X \le x] = P[X \le e^x] = F(e^x) = \begin{cases} e^{\frac{\theta x}{1-\theta}}, & x < 0 \\ 1, & x > 0 \end{cases}$

$g(x) = \begin{cases} \frac{\theta}{1-\theta}e^{\frac{\theta x}{1-\theta}} & x < 0 \end{cases}$

$$\mathbb{E}(Y) = \frac{\theta}{1-\theta}\int_{-\infty}^{0} x\, e^{\frac{\theta x}{1-\theta}}dx = x\, e^{\frac{\theta x}{1-\theta}}\Big|_{-\infty}^{0} - \int_{-\infty}^{0} e^{\frac{\theta x}{1-\theta}}dx$$

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

Par LFGN : $\bar{Y}_n \xrightarrow{p.s} \mathbb{E}(Y) = -\frac{1-\theta}{\theta}$. Par le Lemme de Slutsky : $\hat{\theta}_n \xrightarrow{p.s} \frac{1}{1+\frac{1-\theta}{\theta}} = \theta \Rightarrow$

i) Efficacité : $I_1(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log f_\theta(X)}{\partial\theta^2}\right]$

$$= \frac{1}{\theta^2} - \frac{1}{(1-\theta)^2} - \mathbb{E}(Y)\cdot\frac{2}{(1-\theta)^3}$$

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

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

$$\left.\begin{array}{l} Var(\hat{\theta}_n) = Var(\bar{X}_n) = \frac{\theta}{n}\left(\frac{1}{2-\theta}-\theta\right) \\ I_n(\theta) = n I_1(\theta) \end{array}\right\} \Rightarrow Var(\hat{\theta}_n) \neq \frac{1}{I_n(\theta)}$$

$\Rightarrow \hat{\theta}_n$ pas efficace.

ii) $L_n(\theta) = \frac{\theta^n}{(1-\theta)^n}\prod_{i=1}^{n} X_i^{\frac{2\theta-1}{1-\theta}}\, \mathbb{1}_{0\leq X_i<1}$

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

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

$$\Rightarrow \theta - 1 = \theta\bar{Y}_n \Rightarrow \tilde{\theta}_n = \frac{1}{1-\bar{Y}_n}$$

vérif

$$\frac{\partial^2 \log L_n(\theta)}{\partial\theta^2} = -\frac{n}{\theta^2} + \frac{n}{(1-\theta)^2} + \frac{2n}{(1-\theta)^3}\bar{Y}_n$$

$$\frac{\log L_n(\tilde\theta_n)}{\partial\theta^2} = n\,\frac{\tilde\theta_n^2-(1-\tilde\theta_n)^2}{\tilde\theta_n^2(1-\tilde\theta_n)^2} + \frac{2n}{(1-\tilde\theta_n)^3}\cdot\frac{\tilde\theta_n-1}{\tilde\theta_n}$$

$$= \frac{-2n}{\tilde\theta_n(1-\tilde\theta_n)^2} + n\,\frac{2\tilde\theta_n-1}{\tilde\theta_n^2(1-\tilde\theta_n)^2}$$

$$= n\,\frac{-2\tilde\theta_n+2\tilde\theta_n-1}{\tilde\theta_n^2(1-\tilde\theta_n)^2} = -\frac{n}{\tilde\theta_n^2(1-\tilde\theta_n)^2} < 0$$

$\Rightarrow \tilde\theta_n$ point de max.

7) $\tilde\theta_n = \frac{1}{1-\bar{Y}_n}$. Par Slutsky : $\tilde\theta_n \xrightarrow{p.s} \frac{1}{1-E(Y)} = \frac{1}{1+\frac{1-\theta}{\theta}} = \theta$

$\Rightarrow \tilde\theta_n$ fortement conv.

8) On considère l'estimateur obtenus par la méthode des moments : $\hat\theta_n = \bar{X}_n$.

Par le TCL :

$$\sqrt{n}\,\frac{\bar{X}_n-\theta}{\sqrt{\operatorname{Var} X}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1)$$

$$\operatorname{Var}(X) = \theta\left(\frac{1}{2-\theta}-\theta\right)$$

Soit la v.a : $Z_n = \sqrt{n}\,\frac{\bar{X}_n-\theta}{\sqrt{\bar{X}_n\left(\frac{1}{2-\bar{X}_n}-\bar{X}_n\right)}}$

$$= \underbrace{\sqrt{n}\,\frac{\bar{X}_n-\theta}{\sqrt{\theta\left(\frac{1}{2-\theta}-\theta\right)}}}_{\xrightarrow{\mathcal{L}} N(0,1)} \cdot \underbrace{\sqrt{\frac{\theta\left(\frac{1}{2-\theta}-\theta\right)}{\bar{X}_n\left(\frac{1}{2-\bar{X}_n}-\bar{X}_n\right)}}}_{\xrightarrow{p.s} 1}$$

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

$$1-\alpha = P[a < Z_n < b] \xrightarrow[n\to\infty]{} P[a < N(0,1) < b]$$

$u_{\frac{\alpha}{2}}$, $b = u_{1-\frac{\alpha}{2}}$ fractiles de la loi $\mathcal{N}(0,1)$

donc, l'estimateur asymptotique de θ, de niveau $1-\alpha$:

$$\mathbb{P}\left[-u_{1-\frac{\alpha}{2}} < \sqrt{n}\,\frac{\overline{X_n}-\theta}{\sqrt{\overline{X_n}\left(\frac{1}{2-\overline{X_n}}-\overline{X_n}\right)}} \leq u_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\Rightarrow \boxed{\overline{X_n} - \sqrt{\frac{\overline{X_n}}{n}\left(\frac{1}{2-\overline{X_n}}-\overline{X_n}\right)}\cdot u_{1-\frac{\alpha}{2}} \leq \theta \leq \overline{X_n} + \sqrt{\frac{\overline{X_n}}{n}\left(\frac{1}{2-\overline{X_n}}-\overline{X_n}\right)}\cdot u_{1-\frac{\alpha}{2}}}$$

Exercice 2 $\quad X \sim \mathcal{N}(m,1)$

$H_0 : m = 1 \qquad H_1 : m \neq 1$.

) On utilise le Lemme de Neyman-Pearson :

$$L_1 = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-m)^2\right)$$

$$L_0 = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-1)^2\right)$$

$$\frac{L_1}{L_0} = \exp\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-1)^2 - \frac{1}{2}\sum_{i=1}^{n}(x_i-m)^2\right) = \exp\left(\frac{n}{2} - \frac{nm^2}{2} - n\overline{x_n} + nm\overline{x_n}\right)$$

Par le Lemme de Neyman-Pearson, la zone de rejet est :

$\cancel{\alpha = \mathbb{P}[(x_1,\dots,x_n)}$

$$R = \left\{(x_1,\dots,x_n) \;\middle|\; \frac{L_1}{L_0} > k\right\} = \left\{(m-1)\overline{X_n} > k_1\right\}$$

Test de risque α :

$$\alpha = \mathbb{P}\left[(X_1,\dots,X_n)\in R \mid H_0 \text{ vraie}\right] = \mathbb{P}\left[(m-1)\overline{X_n} > k_1 \mid m = 1\right]$$

$$\Rightarrow \begin{cases} \frac{\alpha}{2} = \mathbb{P}\left[\overline{X_n} > k_2 \mid m = 1\right] \\ \frac{\alpha}{2} = \mathbb{P}\left[\overline{X_n} < k_3 \mid m = 1\right] \end{cases}$$

(5)

$= \mathbb{P}\left[\sqrt{n}\,\frac{\overline{X_n}-1}{1} > k_4 \,\middle|\, m=1\right]$

. $\frac{\alpha}{2} = \mathbb{P}\left[\sqrt{n}\,\frac{\overline{X_n}-1}{1} < k_5 \,\middle|\, m=1\right]$

Sous H_0, $Z = \sqrt{n}\,\frac{\overline{X_n}-1}{1} \sim \mathcal{N}(0,1)$. Donc, $k_4 = u_{1-\frac{\alpha}{2}}$ et $k_5 = -u_{1-\frac{\alpha}{2}}$

Stat de test : $Z = \sqrt{n}(\overline{X_n}-1)$, zone de rejet :

$$R = \left\{ |Z| > u_{1-\frac{\alpha}{2}} \right\}$$

2) Test de Wald : $R(\theta) = \theta - 1$ avec $\theta = m$.

L'EMV de θ : $\hat{\theta}_n = \overline{X_n}$. L'information de Fisher :

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

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

$\Rightarrow I_n(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log L_n(\theta)\right] = n$

Alors, la statistique de Wald est :

$$W_n = \frac{1}{n}(\overline{X_n}-1)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

La zone de rejet : $R : \left\{ W_n > u_{1-\alpha;\, \chi^2(1)} \right\}$

) Entre les 2 statistiques, on préfère la première : elle est UPP et en plus la loi est exacte

Exercice 3

X : la teneur en cacao $\sim \mathcal{N}(m, \sigma^2)$ avec σ^2 inconnu.

H_0 : le chocolat est de qualité supérieure : $m \geq 430$

H_1 : le chocolat est de qualité inférieure : $m < 430$

(6)

st sur la moyenne d'une loi Normale, de variance inconnue :

Stat de test : $z = \sqrt{n}\,\dfrac{\overline{X}_n - 430}{S_n^*}$

La zone de rejet : $R = \{ z < u_{8,\alpha} \} = \{ z < -1,86 \}$, $\alpha = 0,05$
$= \{ z < -0,889 \}$ $\alpha = 0,20$

fractile de la loi $t(8)$

$\overline{x}_n = \cancel{447,8889}$ 431.222 , $s_n^* = \cancel{37,25736}$ 26,85

~~$z = 3\,\dfrac{447,8889 - 430}{37,25736} = 1,44 \notin R \Rightarrow H_0$ acceptée~~

$z = 0,13 \notin R \Rightarrow H_0$ acceptée

Exercice 4

Y : log(concentration NO_2) $\sim \mathcal{N}$.
X_1 : log(nb. voitures)
X_2 : temp au sol.
X_3 : vitesse vent
X_4 : différence temp.

1) $n - 1 = 4 + 495 \Rightarrow n = 500$

2) Modèle de régression multiple.

(1) $y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + b_4 X_{4i} + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$
$i = 1, \dots, n$

3) H_0 : (1) non signif $\Leftrightarrow$ NO_2 n'est infl par aucune des 4 var $\Leftrightarrow$ $b_1 = b_2 = b_3 = b_4 = 0$.

Modèle : (2) $y_i = b_0 + \varepsilon_i$

H_1 : (1) signif $\Leftrightarrow$ NO_2 infl par au moins une des 4 var $\Leftrightarrow \exists\, b_j \neq 0$ $j = 1, 2, 3, 4 \Leftrightarrow$ (1)

Stat de test : $z = \dfrac{SM/4}{SR/495} \underset{H}{\sim} F(4, 495)$

$z = 116$, p-value $= 10^{-16} \Rightarrow H_0$ rejetée $\Rightarrow$ (1) signif

) $\hat{b}_0 = 1{,}10$ $\hat{b}_1 = 0{,}43$ $\hat{b}_2 = -0{,}019$ $\hat{b}_3 = -0{,}133$

$\hat{b}_4 = 0{,}13$, $\hat{\sigma} = 0{,}54$

Interprétation: la pollution (son log) est égale à 1,10 si il y a 1 voiture, température 0, vitesse 0, diff = 0

Avec chaque log(nb voiture) la pollution augment avec 0,43

chaque degré en plus, la pollution diminue de $-0{,}01$

chaque m/s de plus de vent, la poll diminue de $-0{,}13$

chaque degré en plus à la différence de temp, la pollution augmente de 0,13

5) Test de la température

H_0: la température n'influe pas la pollution | nb voitures, vitesse vent, diff temp dans le modèle:

$b_2 = 0 \mid \ldots$ Modèle: (3) $Y_i = b_0 + b_1 X_{1i} + b_3 X_{3i} + b_4 X_{4i} + \varepsilon_i$

H_1: la temp influe le NO_2 | nb voitures, X_3, X_4 dans le modèle

$b_2 \neq 0 \mid \ldots \longrightarrow$ Modèle: (1)

Stat de test:

$$Z = \frac{\hat{B}_2}{\sqrt{\widehat{Var}(B_2)}} \underset{H_0}{\sim} t(495)$$

8

-value $= 10^{-06}$ $\Rightarrow H_0$ rejetée $\Rightarrow$ la température a une infl sur la ~~sto~~ conc de NO_2 ($z = -\cancel{0},744$)

Pour les autres : X_1, X_3, X_4 signif.

6) $R^2 = 0,54 \Rightarrow$ modèle de qualité assez médiocre

7) $\hat{Y} = 1,10 + 0,43 \cdot \log(10) - 0,019 - 0,13 + 0,13$

$= 2,07$