# Corrigé examen 14 janvier 2015

1) $\mathbb{E}(X) = \left(\frac{1}{\theta}-1\right)\int_0^\infty x \cdot e^{-\left(\frac{1}{\theta}-1\right)x} dx = \int_0^\infty e^{-\left(\frac{1}{\theta}-1\right)x} dx$

$= \frac{\theta}{\theta-1} e^{-\left(\frac{1}{\theta}-1\right)x}\Big|_0^\infty = \frac{\theta}{1-\theta}$

$\mathbb{E}(X^2) = \left(\frac{1}{\theta}-1\right)\int_0^\infty x^2 e^{-\left(\frac{1}{\theta}-1\right)x} dx = -x^2 e^{-\left(\frac{1}{\theta}-1\right)x}\Big]_0^\infty +$

$+ 2\int_0^\infty x \cdot e^{-\left(\frac{1}{\theta}-1\right)x} dx = 2\left(\frac{\theta}{1-\theta}\right)^2$

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

2) On considère l'équation en $\theta$: $\mathbb{E}(X) = \bar{x}_n$

$\frac{\theta}{1-\theta} = \bar{x}_n \Rightarrow \theta = \frac{\bar{x}_n}{1+\bar{x}_n}$. Donc, on estimateur par la méthode des moments: $\hat{\theta}_n = \frac{\bar{X}_n}{1+\bar{X}_n}$.

Par LFGN on a $\bar{X}_n \xrightarrow{P.S} \mathbb{E}(X)$. Donc, $\bar{X}_n \xrightarrow{P.S} \frac{\theta}{1-\theta}$

$\Rightarrow \hat{\theta}_n \xrightarrow{P.S} \theta \Rightarrow \hat{\theta}_n$ fortement consistant pour $\theta$.

3) $f_\theta(x) = \left(\frac{1}{\theta}-1\right) \cdot \exp\left(-\left(\frac{1}{\theta}-1\right)x\right) \cdot \mathbb{1}_{x>0}$

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

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

avec $C(\theta) = -\left(\frac{1}{\theta}-1\right)$, $T(x) = x$, $D(\theta) = \log\left(\frac{1}{\theta}-1\right)$

$\tilde{S}(x) = \mathbb{1}_{x>0}$

Alors $\sum_{i=1}^n T(X_i) = \sum_{i=1}^n X_i$ est une statistique exhaustive pour $\theta$.

4) $F(x) = \int_0^x f_\theta(t)dt = \begin{cases} 0 & \text{si } x<0 \\ \int_0^x \left(\frac{1}{\theta}-1\right) e^{-\left(\frac{1}{\theta}-1\right)t} dt = 1-\exp\left(-\left(\frac{1}{\theta}-1\right)x\right), & \text{si } x>0 \end{cases}$

$p = \mathbb{P}[X \geq 1] = 1 - F(1) = \exp\left(-\left(\frac{1}{\sigma} - 1\right)\right)$

5) $Y = \mathbb{1}_{X \geq 1} = \begin{cases} 1 & \text{si } X \geq 1 \\ 0 & \text{si } X < 1 \end{cases} \Rightarrow Y \sim \mathcal{B}(p')$

$p' = \mathbb{P}[X \geq 1] = p = \exp\left(-\left(\frac{1}{\sigma} - 1\right)\right)$

6) $\mathbb{E}(\bar{Y}_n) = \mathbb{E}(Y) = p = \exp\left(-\left(\frac{1}{\sigma} - 1\right)\right)$

$Var(\bar{Y}_n) = \frac{Var(Y)}{n} = \frac{p(1-p)}{n}$

7) Puisque $Y_i \sim \mathcal{B}(p)$, alors l'EMV de $p$ est : $\hat{p}_n = \bar{Y}_n$

8) $Y \sim \mathcal{B}(p)$, alors sa fonction de fréquence est :

$g(x) = \mathbb{P}[Y = x] = p^x (1-p)^{1-x}$ avec $x \in \{0, 1\}$

$= \exp\left(x \cdot \log p + (1-x)\log(1-p)\right)$

$= \exp\left(\log \frac{p}{1-p} \cdot x + \log(1-p)\right)$

$= \exp\left(C(p) \cdot T(x) + D(p) + S(x)\right)$

avec $C(p) = \log \frac{p}{1-p}$, $T(x) = x$, $D(p) = \log(1-p)$

$S(x) \equiv 0$.

Puisque la loi de $Y$ est de type exponentiel, on a que :

$\frac{1}{n} \sum_{i=1}^{n} T(Y_i) = \bar{Y}_n = \hat{p}_n$ est un estimateur sans biais, efficace pour $\mathbb{E}[T(Y)] = \mathbb{E}(Y) = p$.

On a aussi que $\sum_{i=1}^{n} T(Y_i) = n\bar{Y}_n$ exhaustif pour $p$, donc $\bar{Y}_n = \hat{p}_n$ exhaustif pour $p$ (exhaustivité valable jusqu'à une cte)

Enfin. $\bar{Y}_n \xrightarrow{P.S} \mathbb{E}(Y)$, donc $\hat{p}_n \xrightarrow{P.S} p \Rightarrow \hat{p}_n$ fortement convergent pour $p$.

9) Un estimateur de $p$ sans biais : $\hat{p}_n = \bar{Y}_n$

On considère la transf : $Z_n = \sqrt{n} \frac{\bar{Y}_n - p}{\sqrt{p(1-p)}} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0,1)$

Alors, pour trouver l'estimateur asymptotique par intervalle, il faut trouver les bornes aléatoires $A_n$ et $B_n$ telles que : $1-\alpha = \lim_{n\to\infty} P[A_n < p < B_n]$ (1)

ce qui est équivalent, par à trouver les constantes a et b telles que :

$$1-\alpha = \lim_{n\to\infty} \mathbb{P}[a \leq Z_n \leq b] \qquad (2)$$

$$\Leftrightarrow \begin{cases} \frac{\alpha}{2} = \lim_{n\to\infty} \mathbb{P}[Z_n < a] \Rightarrow a = u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}} \\ 1-\frac{\alpha}{2} = \lim_{n\to\infty} \mathbb{P}[Z_n < b] \Rightarrow b = u_{1-\frac{\alpha}{2}} \end{cases}$$

avec $u_{\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}}$ fractile de la loi (limite) $N(0,1)$.

On remplace en (2) :

$$(3) \quad 1-\alpha = \lim_{n\to\infty} \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\overline{Y}_n - p}{\sqrt{p(1-p)}} \leq u_{1-\frac{\alpha}{2}}\right]$$

D'autre part, $\overline{Y}_n \xrightarrow{P.S.} p$, $Z_n \xrightarrow{\mathcal{L}} N(0,1)$, alors par le théorème de Slutsky, on a : $\sqrt{n}\,\frac{\overline{Y}_n - p}{\sqrt{\overline{Y}_n(1-\overline{Y}_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1)$

Alors, par (3), on a :

$$(4) \quad 1-\alpha = \lim_{n\to\infty} \mathbb{P}\left[-u_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\overline{Y}_n - p}{\sqrt{\overline{Y}_n(1-\overline{Y}_n)}} \leq u_{1-\frac{\alpha}{2}}\right]$$

$$= \lim_{n\to\infty} \mathbb{P}\left[\underbrace{\overline{Y}_n - \sqrt{\frac{\overline{Y}_n(1-\overline{Y}_n)}{n}}\cdot u_{1-\frac{\alpha}{2}}}_{A_n} \leq p \leq \underbrace{\overline{Y}_n + \sqrt{\frac{\overline{Y}_n(1-\overline{Y}_n)}{n}}\cdot u_{1-\frac{\alpha}{2}}}_{B_n}\right]$$

10) Par 8); on a $\bar{Y}_n \xrightarrow{P.S} p$ , donc $\hat{\theta}_n = \frac{1}{1-\log \bar{Y}_n} \xrightarrow{P.S} \frac{1}{1-\log(p)}$

D'autre part, $p = \exp\left(-\left(\frac{1}{\theta}-1\right)\right) \Rightarrow \log p = -\left(\frac{1}{\theta}-1\right)$

$\Rightarrow 1-\log p = \frac{1}{\theta} \Rightarrow \theta = \frac{1}{1-\log p} \Rightarrow \hat{\theta}_n \xrightarrow{P.S} \theta \Rightarrow \hat{\theta}_n$ fortement consistant.

11) $\mu = p^2$ , $p \in ]0,1[ \Rightarrow \mu \in ]0,1[$

Considérons la fonction $h: ]0,1[ \to \underbrace{]0,1[}_{\text{intervalle borné}}$ , définie par $h(p) = p^2$.

Puisque $\hat{p}_n = \bar{Y}_n$ est l'EMV de $p$, alors $h(\hat{p}_n) = (\bar{Y}_n)^2$ est l'EMV de $\mu$ $\Rightarrow \hat{\mu}_n = (\bar{Y}_n)^2$ .

12) $\bar{Y}_n \xrightarrow{P.S} p \Rightarrow (\bar{Y}_n)^2 \xrightarrow{P.S} p^2 \Rightarrow \hat{\mu}_n \xrightarrow{P.S} \mu \Rightarrow \hat{\mu}_n$ fortement conv.

$$\mathbb{E}\left[(\bar{Y}_n)^2\right] = Var(\bar{Y}_n) + \mathbb{E}^2(\bar{Y}_n) = \frac{p(1-p)}{n} + p^2$$

$$= \frac{p}{n} + p^2\left(1-\frac{1}{n}\right) \xrightarrow[n\to\infty]{} p^2 \Rightarrow \hat{\mu}_n \text{ biaisé, mais}$$

asymptotiquement sans biais.

## Exercice 2

1) $\lambda = \frac{1}{\theta}-1$ , $s = 1 \Rightarrow X \sim \gamma\left(1, \frac{1}{\theta}-1\right) \Rightarrow \sum_{i=1}^{n} X_i \sim \gamma\left(n, \frac{1}{\theta}-1\right)$

2) $H_0 : \theta = \theta_0$ , $H_1 : \theta > \theta_0$ .

La vraisemblance: $L_n(\theta) = \left(\frac{1}{\theta}-1\right)^n \cdot \exp\left(-n\left(\frac{1}{\theta}-1\right)\bar{X}_n\right) \cdot \mathbb{1}_{\min x_i}$

$L_1(\theta) = L_n(\theta)$ , $L_0 = L_n(\theta_0)$

Par le Lemme de Neyman-Pearson, on a :

$\alpha = \mathbb{P}\left[L_1 > k L_0 \mid X_i \sim \gamma\left(1, \frac{1}{\theta_0}-1\right)\right]$

La règle de décision étant : $\delta = \begin{cases} H_1 & \text{si } L_1 > k L_0 \\ H_0 & \text{si } L_1 < k L_0 \end{cases}$

Alors, le niveau $\alpha$ du test est :

$$\alpha = \alpha(\theta_0) = \mathbb{P}\left[\exp\left(-n\bar{x}_n\left(\frac{1}{\theta}-\frac{1}{\theta_0}\right)\right) > k_1 \;\middle|\; x_i \sim \gamma\left(1, \frac{1}{\theta_0}-1\right)\right]$$

$$= \mathbb{P}\left[\bar{X}_n\left(\frac{1}{\theta_0}-\frac{1}{\theta}\right) > k_2 \;\middle|\; x_i \sim \gamma\left(1, \frac{1}{\theta_0}-1\right)\right]$$

$$= \mathbb{P}\left[\bar{X}_n > k_3 \;\middle|\; X_i \sim \gamma\left(1, \frac{1}{\theta_0}-1\right)\right]$$

$$= \mathbb{P}\left[\sum_{i=1}^{n} x_i > k_4 \;\middle|\; x_i \sim \gamma\left(1, \frac{1}{\theta_0}-1\right)\right]$$

Mais $\sum_{i=1}^{n} x_i \sim \gamma\left(n, \frac{1}{\theta_0}-1\right)$, donc $k_4$ est le fractile d'ordre $1-\alpha$ de la loi $\gamma\left(n, \frac{1}{\theta_0}-1\right)$.

Exercice 3 $\quad n = 1600$

1) $X_i \sim \begin{cases} 1 & \text{si poisson Carpe} \\ 0 & \text{sinon} \end{cases} \Rightarrow X_i \sim \mathcal{B}(p)$

$p$ = la proba qu'un poisson pêché soit une Carpe.

$\alpha = 0.05$. $\quad p_0 = \frac{1}{5}$

$H_0 : p = p_0 \qquad H_1 : p \neq p_0$ (test sur la proportion)

Statistique de test : $\quad Z = \sqrt{n}\,\dfrac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}}$

Zone de rejet : $R = \{|Z| > u_{1-\frac{\alpha}{2}}\}$, avec $u_{1-\frac{\alpha}{2}}$ fractile d'ordre $1-\frac{\alpha}{2}$ de la loi $\mathcal{N}(0,1) \Rightarrow u_{1-\frac{\alpha}{2}} = 1{,}96$

La réalisation de la stat de test :

$$z = \sqrt{1600}\,\frac{\frac{211}{1600}-\frac{1}{5}}{\sqrt{\frac{1}{5}\cdot\frac{4}{5}}} = -40\,\frac{\frac{109}{1600}}{\frac{2}{5}} = -6{,}8 \in R$$

$\Rightarrow H_0$ rejetée $\Rightarrow$ la population des Carpes est $\neq \frac{1}{5}$ de la population

2) $n = 211$, $X$ : le poids d'une Carpe $\sim \mathcal{N}(m, \sigma^2)$

Le poids est d'au moins 1k : $m \geq 1000$ $(m_0 = 1000)$

$H_0 : m \geq 1000$ $H_1 : m < 1000$ (test sur la moyenne d'une loi Normale de variance inconnue)

Stat test :

$$Z = \sqrt{n} \, \frac{\overline{X_n} - m_0}{S_n^*}$$

zone de rejet : $R = \{ Z < -u_{1-\alpha} \} = \{ Z < -1,65 \}$

fractile d'ordre $1-\alpha$ de la loi $t(210)$

Réalisation stat test :

$$z = \sqrt{211} \, \frac{950 - 1000}{\sqrt{1600}} = \sqrt{211} \, \frac{-50}{40} = -\frac{5}{4}\sqrt{211} = -18,1$$

$\Rightarrow z \in R \Rightarrow H_0$ rejetée, donc le poids est $< 1000$g.

Exercice 4

$Y$ : log(Resist), $X1 : LP$, $X2 : PC$, $X3 : LDR$, $X4 : BDR$, $X5 : LBR$ $X6 : NR\,FROUDE$.

1) $n - 1 = 302 + 5 = 307 \Rightarrow n = 308$

2) On a un modèle de régression multiple, à 5 régresseurs :

(1) $Y_i = b_0 + b_1 X1_i + \cdots + b_5 X5_i + \varepsilon_i$, $\varepsilon_i \sim N(0,1)$

$\varepsilon_i \perp\!\!\!\perp \varepsilon_j$ $i \neq j$

$i = 1, \ldots, n$

3) $H_0$ : (1) non significatif $\Leftrightarrow$ $Y$ n'est influencé par aucune des 5 variables : $b_j = 0$ $\forall j = 1, \ldots, 5$

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

$H_1$ : (1) signif $\Leftrightarrow$ au moins une des 5 var. influ $Y$ $\Leftrightarrow$ $\exists j \in \{1, 2, \ldots, 5\}$ t.q : $b_j \neq 0$

Modèle : (1)

Statistique de test : $Z = \frac{SM/5}{SR/302} \underset{H_0}{\sim} F(5,302)$

réalisation $z = 2048$, p-value $= 10^{-16}$ => $H_0$ rejetée
=> modèle (1) significatif

4) $\hat{b}_0 = -3,48$ $\hat{b}_1 = 0,02$ $\hat{b}_2 = -1,83$ $\hat{b}_3 = 0.02$
$\hat{b}_4 = 0.11$ $\hat{b}_5 = 18.03$ $\hat{\sigma} = 0.31$

5) $H_0$ : X1 n'influe pas y | X2, .., X5 sont dans le modèle
$\Leftrightarrow b_1 = 0 | \ldots$
$\Leftrightarrow$ Modèle réduit : (3) $Y_i = b_0 + b_2 X2_i + \cdots + b_5 X5_i$

$H_1$ : X1 influe | $X_2, \ldots, X5$ dans le modèle
$\Leftrightarrow b_1 \neq 0$ | —— // ——
$\Leftrightarrow$ Modèle complet : (1)

Statistique de test : $Z = \frac{B_1}{\sqrt{Var(B_1)}} \underset{H_0}{\sim} t(302)$

Réalisation stat de test : $z = \frac{0.02}{0.011} = 1,788$

p-value $= 0,07$ => $H_0$ rejetée.

Variables significatives : PC (p-value=0.02), BDR (p-value = 0.004), NRFROUDE (p-value < $10^{-16}$)

Variables non-signif : LP, LDR.

6) $R^2 = 0,97$ => modèle de très bonne qualité