Université de Kairouan
Institut Supérieur des Mathématiques Appliquées & Informatique

Statistique paramétrique & non paramétrique

Corrigé Examen Janvier 2015

Exercice 1

$$f(x,\theta)=\begin{cases}\frac{3x^2}{\theta}\exp(-\frac{x^3}{\theta}) & \text{si } x>0,\ \ \theta>0\\ 0 & \text{sinon}\end{cases}$$

1) $Y=\dfrac{X^3}{\theta}$.

Soient F et G les fonctions de répartition des $v.a$ X et Y respectivement.

$$\begin{aligned}G(y)=P(Y\leq y) &= P\left[\frac{X^3}{\theta}\leq y\right]=P\left[X^3\leq \theta y\right]\\ &= P\left[X\leq(\theta y)^{\frac{1}{3}}\right] \text{ car } x \text{ et } \theta \text{ sont tous les deux strictement positifs.}\\ &= F\left((\theta y)^{\frac{1}{3}}\right)\end{aligned}$$

D'où, $g(y)=\frac{1}{3}\theta(\theta y)^{-\frac{2}{3}}f\left((\theta y)^{\frac{1}{3}}\right)=\frac{1}{3}\theta(\theta y)^{-\frac{2}{3}}\frac{3}{\theta}(\theta y)^{\frac{2}{3}}\exp\left(-\frac{1}{\theta}\theta y\right)=\exp(-y)$ si $y>0$.

Donc, la variable aléatoire Y suit une loi exponentielle de paramètre $\lambda=1$. Ainsi $E(Y)=V(Y)=1$.

On a $E(Y)=1$, donc $E\left(\dfrac{X^3}{\theta}\right)=1 \quad\Rightarrow\quad E(X^3)=\theta$.

De même, $V(Y)=1\Rightarrow V\left(\dfrac{X^3}{\theta}\right)=1 \quad\Rightarrow\quad V(X^3)=\theta^2$.

2) La fonction de vraisemblance de $f(x_i,\theta)$ s'écrit comme suit :

$$\ell(x_1,x_2,\ldots,x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta)=\left(\frac{3x^2}{\theta}\right)^n\exp\left(-\frac{\sum_{i=1}^{n}X_i^3}{\theta}\right)$$

D'où, $\mathcal{L}(X,\theta) = \log \ell = n\log(3x^2) - n\log\theta - \frac{\sum_{i=1}^{n} X_i^3}{\theta}$.

$$\frac{\partial \mathcal{L}}{\partial \theta} = 0 \quad \Rightarrow \quad -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} X_i^3}{\theta^2} = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{\sum_{i=1}^{n} X_i^3}{n}$$

Cet estimateur est-il sans biais ?

$E\left(\hat{\theta}\right) = \frac{E\left(\sum X_i^3\right)}{n} = E(X^3) = \theta$. Donc, $\hat{\theta}$ est **sans biais**.

Est-il convergent ?

$V\left(\hat{\theta}\right) = \frac{V\left(\sum X_i^3\right)}{n^2} = \frac{V(X^3)}{n} = \frac{\theta^2}{n} \longrightarrow 0$ quand $n \to \infty$. Donc, $\hat{\theta}$ **est convergent**.

Exercice 2

1) La méthode des moments consiste à égaliser entre les moments théoriques et les moments empiriques. Soit : $E(X) = \overline{X}$, or $E(X) = \lambda$ d'où, $\hat{\lambda} = \overline{X} = \frac{\sum n_i x_i}{n}$.

D'où $\hat{\lambda} = \frac{0 \times 52 + 1 \times 151 + 2 \times 130 + 3 \times 102 + 4 \times 45 + 5 \times 20}{500} = 1.994 \simeq 2$.

2) X est une $v.a$ représentant le nombre de voiture qui arrivent à un poste de péage.

On veut tester $H_0 : X \sim \mathcal{P}(\lambda = 2)$.

La statistique à utiliser est :

$$\chi^2 = \sum_{i=0}^{i=5} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(5)$$

où 0_i : fréquence (effectif) observée et E_i : fréquence (effectif) théorique, $E_i = np_i$, $n = \sum n_i$.

Pour le calcul de la statistique, on aura ce tableau :

x	0	1	2	3	4	5
0_i	52	151	130	102	45	20
E_i	67.66764	135.33528	135.33528	90.22352	45.11176	18.04470
$\frac{(O_i - E_i)^2}{E_i}$	3.627657	1.8131513	0.210331	1.53713	0.0002768	0.211872

D'où χ^2_O=7.4 et on a, d'après la table, $\chi^2_{0.95}(5) = 11.07$, donc $\chi^2_O < \chi^2_{0.95}(5)$, ce qui permet d'accepter H_0 au seuil $\alpha = 5\%$.

Exercice 3

Partie A : $X \sim N(m_X, \sigma_X^2)$ et $Y \sim N(m_Y, \sigma_Y^2)$.

1) On veut tester $H_0 :\ m_X = 190$ vs $H_1 :\ m_X = 195$.

a) La région critique au sens de Neyman-Pearson est définie par :

$$W = \left\{ (X_1, \ldots, X_n) \in \mathbb{R}^n / \frac{L_1}{L_0} > k \right\}$$

Or $L(X_i, m_X) = \prod_{i=1}^{n} f(x_i, m_X) = \left(2\pi\sigma_x^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_X^2}\sum_{i=1}^{n}(X_i - m_X)^2\right]$

$$\begin{aligned}
\frac{L_1}{L_0} > k &\Longrightarrow \exp\left[-\frac{1}{2\sigma_X^2}\sum_{i=1}^{n}\left(X_i - m_X^0\right)^2 - \sum_{i=1}^{n}\left(X_i - m_X^1\right)^2\right] > k \\
&\Longrightarrow \sum_{i=1}^{n}\left(X_i - m_X^0\right)^2 - \sum_{i=1}^{n}\left(X_i - m_X^1\right)^2 > k' \qquad (k' = 2\sigma_X^2 \log k) \\
&\Longrightarrow \cancel{\sum_{i=1}^{n} X_i^2} - 2n\overline{X}m_X^0 + (m_X^0)^2 - \cancel{\sum_{i=1}^{n} X_i^2} + 2n\overline{X}m_X^1 - (m_X^1)^2 > k \\
&\Longrightarrow \left(2n\overline{X} - (m_X^1 + m_X^0)\right)\left(m_X^1 - m_X^0\right) > k'. \qquad \text{Soit } \boxed{\overline{X} > C}.
\end{aligned}$$

Ainsi $\boxed{W = \left\{(X_1, \ldots, X_n) \in \mathbb{R}^n / \overline{X} > C\right\}}$.

Déterminons la valeur de C pour $\alpha = 5\%$. On a $X \sim N\left(m_X, \sigma_X^2\right)$ avec σ_X^2 inconnu, donc on l'estime par $S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$. D'où $T = \frac{\overline{X} - m_X}{S_x/\sqrt{n}} \sim St(n-1)$.

$$\begin{aligned}
\alpha &= p\left[D_1/H_0\right] = p\left[\overline{X} > C/m_X = 190\right] = p\left[T > \sqrt{n}\frac{C - 190}{S_x}\right] \\
&= 1 - p\left[T \leq \sqrt{n}\frac{C - 190}{S_x}\right].
\end{aligned}$$

D'où $p\left[T \leq \sqrt{n}\frac{C-190}{S_x}\right] = 95\%$. D'après la table de la loi de Student à 8 ddl, on a : $\sqrt{n}\frac{C-190}{S_x} = 1.86$.

Donc $C = 190 + 1.86\frac{S_x}{\sqrt{9}}$; avec $S_x = 159.397$. Ainsi $\boxed{C = 288.826}$.

Soit $\boxed{W = \left\{(X_1, \ldots, X_n) \in \mathbb{R}^n / \overline{X} > 288.826\right\}.}$

b) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 190.47 \notin W$, donc, au seuil $\alpha = 5\%$, on décide H_0.

c)

$$\begin{aligned}
\beta &= p\left[D_0/H_1\right] = p\left[\overline{X} \leq C/m_x = 195\right] \\
&= p\left[T \leq \sqrt{n}\frac{C - 195}{S_x}\right] = p\left[T \leq 1.7658\right] = 94.22\%.
\end{aligned}$$

(Cette dernière valeur est calculée à l'aide de l'interpolation linéaire, i.e, $1.397 \longrightarrow 90\%$,

$1.7658 \longrightarrow ?, 1.86 \longrightarrow 95\%.)$.

Ainsi, la puissance du test est : $\eta = 1 - \beta = 5.78\%.$.

2) a) Nous sommes dans le cas où les variances sont inconnues et inégales, d'où la statistique à utiliser est :

$$T = \sqrt{n}\frac{(\overline{X} - \overline{Y}) - (m_x - m_y)}{\sqrt{S_x^2 + S_y^2}} \sim St(2n-2)$$

D'où l'intervalle de confiance est donné par :

$$IC_{0.95}\,(mx - m_y) = \left[(\overline{X} - \overline{Y}) \pm t_{1-\alpha/2}(2n-2)\frac{S_x - S_y}{\sqrt{n}}\right]$$

Soit $IC_{0.95}\,(mx - m_y) = [-156.576; 162.310]$.

b)Il s'agit de tester $H_0 : m_x = my$ vs $H_1 : m_x \neq m_y$

Les moyennes théoriques sont estimés par les moyennes empiriques, i.e, $\hat{m}_x = \overline{X}$ et $\hat{m}_y = \overline{Y}$. On a $\overline{X} - \overline{Y} = 3.088 \in IC(m_x - m_y)$, donc, on accepte H_0. Au seuil, $\alpha = 5\%$, le cours moyen des actions n'est pas affecté par l'opération terroriste du 11/09/2001.

Partie B :

1) Dans ce cas, on observe deux échantillons issus de la même population (Bourse de New York) et on ne sait pas la loi de ces deux échantillons, donc pour étudier s'il y a une différence entre eux, on peut utiliser le test de Wilcoxon.

2) Appliquer le test proposé et donner la décision au seuil $\alpha = 5\%$.