## Corrigé du TD 6

**Exercice 1.** (pas de corrigé) Soit $(X_n, n \geq 1)$, une suite de variables aléatoires indépendantes, de loi normale $N(\theta, \theta)$, avec $\theta > 0$. L'objectif de cet exercice est de présenter deux tests pour déterminer si pour une valeur déterminée $\theta_0 > 0$, on a $\theta = \theta_0$ (hypothèse $H_0$) ou $\theta > \theta_0$ (hypothèse $H_1$). On note

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \; = \; \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\bar{X}_n^2$$

**1.** Déterminer $\hat{\theta}_n$, l'estimateur du maximum de vraisemblance de $\theta$. Montrer directement qu'il est convergent, asymptotiquement normal et donner sa variance asymptotique. Est-il asymptotiquement efficace ?

**2.** Construire un test asymptotique convergent à l'aide de l'estimateur du maximum de vraisemblance.

On considère la classe des estimateurs $T_n^\lambda$ de la forme $T_n^\lambda = \lambda\bar{X}_n + (1-\lambda)V_n$.

**3.** Montrer que la suite d'estimateurs $(T_n^\lambda, n \geq 2)$ est convergente, sans biais. Donner la variance de $T_n^\lambda$ .

**Exercice 2.** L'information dans une direction de l'espace prise par un radar de surveillance aérienne se présente sous la forme d'un n-échantillon $X = (X_1, ..., X_n)$ de variables aléatoires indépendantes de même loi gaussienne de moyenne $\theta$, paramètre inconnu, et de variance $\sigma^2$ connu. On notera $f(x, \theta)$,$x \in \mathbb{R}^n$ la densité du vecteur aléatoire.

En l'absence de tout Objet Volant (Hypothèse $H_0$) $\theta = \theta_0 \in \mathbb{R}^+$, sinon (Hypothèse $H_1$),$\theta = \theta_1$, avec $\theta_1 > \theta_0$.

**1. En utilisant Neyman-Pearson, construire un test de l'hypothèse $\theta = \theta_0$ contre l'hypothèse $\theta = \theta_1$ de niveau $\alpha$.**

On calcule d'abord le rapport de vraisemblance :

$$\frac{L(x,\theta_1)}{L(x,\theta_0)} = \exp\left(-\sum_{i=1}^{n}\frac{(x_i-\theta_1)^2-(x_i-\theta_0)^2}{2\sigma^2}\right) = \exp\left(M\frac{\sqrt{n}(\bar{x}_n-\theta_0)}{\sigma} - \frac{1}{2}M^2\right)$$

avec $M = \frac{\sqrt{n}(\theta_1-\theta_0)}{\sigma}$. La région critique optimale est donc de la forme :

$$W_\alpha = \left\{(x_1, ..., x_n), \frac{\sqrt{n}(\bar{x}_n-\theta_0)}{\sigma} > k_\alpha\right\}$$

On obtient ainsi la statistique de test $\frac{\sqrt{n}(\bar{X}_n-\theta_0)}{\sigma}$ qui suit, sous $H_0$, la loi gaussienne $\mathcal{N}(0,1)$. On doit donc choisir $k_\alpha$ tel que $\alpha = P_{\theta_0}(W_\alpha)$. On obtient donc $k_\alpha = \phi^{-1}(1-\alpha)$ avec $\phi$ la fonction de répartition de la gaussienne centrée réduite.

Remarque : La région critique est indépendante de la valeur de $\theta_1$ (avec $\theta_1 > \theta_0$. Ce test est donc aussi un test de niveau $\alpha$ pour tester $\{\theta = \theta_0\}$ contre $\{\theta > \theta_0\}$.

**2. Quelle est la plus petite valeur de $n$ permettant de construire un test de niveau $\alpha$, $\alpha \in [0,1]$ et d'erreur de deuxième espèce inférieure ou égale à $\beta$, $\beta \in [0,1]$, avec $\alpha < \beta$ ?**

Sous $H_1$, i.e. $\theta = \theta_1$, $\frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} - M$ suit la loi $\mathcal{N}(0,1)$. L'erreur de deuxième espèce est donnée par

$$\begin{aligned} P_{\theta_1}(\bar{W}_\alpha) &= P_{\theta_1}\left(\frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} < k_\alpha\right) \\ &= P_{\theta_1}\left(\frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} - M < k_\alpha - M\right) \leq \beta \end{aligned}$$

On veut que l'erreur de deuxième espèce $\leq \beta$. Donc,

$$\begin{aligned} P_{\theta_1}\left(\frac{\sqrt{n}(\bar{X}_n - \theta_0)}{\sigma} - M < k_\alpha - M\right) &\leq \beta \\ \text{d'où} \quad k_\alpha - M &\leq \phi^{-1}(\beta) \end{aligned}$$

Autrement dit :

$$\begin{aligned} \frac{\sqrt{n}(\theta_1 - \theta_0)}{\sigma} &> k_\alpha - \phi^{-1}(\beta) \\ n &\geq \sigma^2 \frac{\left[\phi^{-1}(1-\alpha) - \phi^{-1}(\beta)\right]^2}{(\theta_1 - \theta_0)^2} \end{aligned}$$

**3. Supposons maintenant qu'en presence d'objet volant l'information fournie par le radar est un n-échantillon $X = (X_1, ..., X_n)$ de variables aléatoires indépendantes de même loi gaussienne de moyenne $\theta \neq \theta_0$, $\theta \in \mathbb{R}$ et de variance $\sigma^2$. Peut-on construire un test de l'hypothèse $\theta = \theta_0$ contre l'hypothèse $\theta \neq \theta_0$ de niveau $\alpha$ donné uniformément plus puissant ?**

C'est un test bilatéral, la même méthode s'applique ! On a donc la même statistique de test mais la région de rejet est de la forme :

$$W_\alpha = \left\{(x_1, ..., x_n), k_{(\alpha,1)} < \frac{\sqrt{n}(\bar{x}_n - \theta_0)}{\sigma} < k_{(\alpha,2)}\right\}$$

**Exercice 3.** Une agence de voyage souhaite cibler sa clientèle. Elle sait que les coordonnées du lieu de vie d'un client $(X, Y)$ rapportées au lieu de naissance $(0,0)$ sont une information significative pour connaître le goût de ce client. Elle distingue :

- La population 1 (Hypothèse $H_0$) dont la loi de répartition a pour densité :

$$p_1(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

- La population 2 (Hypothèse $H_1$) dont la loi de répartition a pour densité :

$$p_2(x,y) = \frac{1}{16}\mathbf{1}_{[-2,2]}(x)\mathbf{1}_{[-2,2]}(y)$$

L'agence souhaite tester l'hypothèse qu'un nouveau client vivant en $(x, y)$ appartient à la population 1 plutôt qu'à la population 2.

**1. Proposer un test de niveau inférieur à $\alpha = 5\%$ et de puissance maximale, construit à partir du rapport de vraisemblance.**

Cet exemple rentre dans le cadre du Lemme de Neyman-Pearson. On définit la région critique du test :

$$A_\alpha = \left\{ (x,y) \in \mathbb{R}^2 ; \frac{p_2(x,y)}{p_1(x,y)} > ka \right\}.$$

On obtient

$$A_\alpha = \left\{ (x,y) \in \mathbb{R}^2 \cap ([-2;2] \times [-2;2]) ; x^2 + y^2 > ka \right\}.$$

Sous $H_0$, $X^2 + Y^2$ suit une loi du $\chi^2_2$ (à 2 degrés de liberté). On a $5\% = P(\chi^2_2) = 5.99)$. Soit $\alpha$ tel que $K_\alpha = 5.99$. On en déduit donc que $P_{H_0}(A_\alpha) = \alpha = 5\%$.

**2. Donner une statistique de test et caractériser graphiquement la région critique dans $\mathbb{R}^2$.**
Une statistique de test est $X^2 + Y^2$. La région critique est l'intersection de l'extérieur du cercle de rayon $\sqrt{5.99}$ avec le carré $[-2;2] \times [-2;2]$.

**Exercice 4.** Soit $X_1, ..., X_n$ un n-échantillon de loi exponentielle de paramètre $\frac{1}{\theta} > 0$.
**1. Construire le test de niveau $\alpha$ $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta > \theta_0\}$.**
La méthode à utiliser pour construire un test unilatéral est la même que pour un test simple, à savoir la méthode de Neyman-Pearson. Soit un certain $\theta' > \theta$. On étudie le rapport de vraisemblance :

$$\frac{p_n(x_1, ..., x_n, \theta')}{p_n(x_1, ..., x_n, \theta_0)} = \left(\frac{\theta_0}{\theta'}\right)^n \exp\left\{ - \left( \frac{1}{\theta'} - \frac{1}{\theta_0} \right) \sum_{i=1}^n x_i \right\}$$

Appliquant le Théorème de Neyman-Pearson, on obtient que la région de rejet optimale de niveau $\alpha$, $W_\alpha$, est de la forme :

$$\begin{aligned} W_\alpha &= \left\{ (x_1, ..., x_n) ; \frac{p_n(x_1, ..., x_n, \theta')}{p_n(x_1, ..., x_n, \theta')} > k_\alpha \right\} \\ &= \left\{ (x_1, ..., x_n) ; \sum_{i=1}^n x_i > C_\alpha \right\} \end{aligned}$$

où $\alpha = P_{\theta_0}(W_\alpha)$. On déduit également la statistique de test $\sum_{i=1}^n X_i$ qui suit, sous $H_0$, la loi gamma $\Gamma(n, \frac{1}{\theta_0})$. Donc, $\frac{2}{\theta_0} \sum_{i=1}^n X_i$ suit la loi Gamma $\Gamma(n, \frac{1}{2})$, i.e. la loi du $\chi^2$ à $2n$ degrés de liberté (toujours sous l'hypothèse nulle). Donc, on peut déterminer $C_\alpha$ :

$$\alpha = P_{\theta_0}\left(W_\alpha) = P_{\theta_0}(\sum_{i=1}^n X_i > C_\alpha\right) = P_{\theta_0}\left( \frac{2}{\theta_0} \sum_{i=1}^n X_i > \frac{2}{\theta_0} C_\alpha \right)$$

Donc, $\frac{2}{\theta_0} C_\alpha = F^{-1}_{\chi^2_{2n}}(1 - \alpha)$. On obtient donc la région critique :

$$W_\alpha = \left\{ (x_1, ..., x_n) ; \sum_{i=1}^n x_i > \frac{\theta_0}{2} F^{-1}_{\chi^2_{2n}}(1 - \alpha) \right\}$$

**2. Construire le test de niveau $\alpha$, $H_0 = \{\theta = \theta_0\}$ contre $H_1 = \{\theta \neq \theta_0\}$.**

La région critique optimale du test bilatéral de niveau $\alpha$ est donnée par :

$$W_\alpha = \left\{(x_1, ..., x_n); \sum_{i=1}^{n} x_i < k_1, \sum_{i=1}^{n} x_i > k_2\right\}$$

avec $k_1$ et $k_2$ tels que $\alpha = P_{\theta_0}(W_\alpha)$.

$$\begin{aligned} 1-\alpha &= P_{\theta_0}\left(k_1 < \sum_{i=1}^{n} X_i < k_2\right) \\ &= P_{\theta_0}\left(\frac{2}{\theta_0}k_1 < \frac{2}{\theta_0}\sum_{i=1}^{n} X_i < \frac{2}{\theta_0}k_2\right) \end{aligned}$$

Sous $h_0$, $\frac{2}{\theta_0}\sum_{i=1}^{n} X_i$ suit la loi $\chi^2_{2n}$, on choisit donc $\frac{2}{\theta_0}k_1 = F^{-1}_{\chi^2_{2n}}(\alpha_1)$ et $\frac{2}{\theta_0}k_2 = F^{-1}_{\chi^2_{2n}}(1-\alpha_2)$, avec $\alpha_1 + \alpha_2 = \alpha$. Un choix classique consiste à prendre $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

(a). Le rapport de vrais. $\frac{L(\lambda_1, x_1, \ldots, x_n)}{L(\lambda_0, x_1, \ldots, x_n)} = e^{n(\lambda_0-\lambda_1)}\left(\frac{\lambda_1}{\lambda_0}\right)^{n\widehat{\lambda}}$ ;

où $\widehat{\lambda} = \overline{X}$.

C'est une fonction croissante de $\widehat{\lambda}$. Donc, d'après le théorème de N-P, la RC optimale au niveau $\alpha^*$ est de la forme $W : \widehat{\lambda} > A$ pour une constante $A$ vérifiant

$$P_{\lambda_0}(\widehat{\lambda} > A) = \alpha^*.$$

(b) D'après le TLC, on a approximativement pour $n$ grand et sous l'hypothèse $H_0$ :

$$\widehat{\lambda} \sim \mathcal{N}(\lambda_0, \lambda_0/n)$$

On en déduit :

$$A \approx \lambda_0 + u_{1-\alpha^*}\sqrt{\frac{\lambda_0}{n}}$$

A.N. : $A \approx 1.87$, $\widehat{\lambda} = 1.8$, donc on conserve l'hypothèse $H_0$.

(c) On a :

$$\pi = P_{\lambda_1}(\widehat{\lambda} > A) \approx 1 - \phi\left(\frac{A - \lambda_1}{\sqrt{\lambda_1/n}}\right)$$

A.N. : $\pi \approx 0.69$.

(d) Soit le problème de test :

$$\begin{array}{ll} H_0 : & \lambda = \lambda_0 \\ h_1 : & \lambda = \lambda_1 \end{array}$$

avec $\lambda_1 > \lambda_0$. La RC optimale pour ce problème a été trouvée précédemment, elle ne dépend pas de $\lambda_1$. Par conséquent, le test précédent est UPP pour le problème de test $H_0$ contre $H_1$.

D'après les résultats numériques trouvés précédemment, il n'y a donc pas lieu de remettre en cause l'affirmation du Ministère de l'Intérieur (au niveau de signification de 5 %).