Université Paris Dauphine Année 2002-2003
U.F.R. Mathématiques de la Décision
MD4 - Modèles Linéaires et généralisations

Correction T.D. 1

Tests d'hypothèses

Exercice 4

1 $X \sim \mathcal{E}\left(\frac{1}{\theta}\right)$, donc $f(x;\theta) = \frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\} 1_{\mathbb{R}_+}(x)$.
Nous devons tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$.
Nous appliquons le lemme de Neyman-Pearson :

$$\frac{f\left(x;\theta_1\right)}{f\left(x;\theta_0\right)} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left\{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^n x_i\right\}.$$

Puisque $\theta > \theta_0$, $\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) > 0$, et donc la région critique optimale du test est $\left\{\sum_{i=1}^n x_i > K\right\}$, il s'agit de la région de rejet de l'hypothèse H_0.
Pour un seuil α donné, nous déterminons k à l'aide de l'équation suivante :

$$\mathbb{P}_{H_0}\left(\sum_{i=1}^n X_i > K\right) = \alpha.$$

Aussi $X_i \sim \mathcal{E}\left(\frac{1}{\theta}\right)$ et donc $\sum_{i=1}^n X_i \sim \gamma\left(n, \frac{1}{\theta}\right)$, car les X_i sont indépendants.
Par conséquent, $\frac{2}{\theta}\sum_{i=1}^n X_i \sim \chi^2(2n)$ et donc $\alpha = 1 - F_{\chi^2(2n)}\left(\frac{2K}{\theta_0}\right)$...

2 Pour $\theta > \theta_0$, la fonction puissance du test est donnée par

$$p(\theta) = 1 - F_{\chi^2(2n)}\left(\frac{2K}{\theta}\right).$$

Exercice 5

1 X est de fonction de répartition F continue, strictement monotone, $\theta = F^{-1}\left(\frac{1}{2}\right)$ est médiane de X. Nous disposons d'un échantillon de n réalisations de la variable aléatoire X et nous souhaitons tester $H_0 : \theta = 0$ contre $H_1 : \theta > 0$. Soit K_n le nombre de réalisations positives dans l'échantillon de taille n. Nous allons utiliser la stratégie de décision suivante :

– si $k_n > K$, nous rejetons H_0 ;
– si $k_n = K$, nous rejetons H_0 avec la probabilité π ;
– si $k_n < K$, nous ne pouvons pas rejeter H_0.
Plus nous avons de réalisations positives dans l'échantillon, plus il est intuitif de penser que la médiane s'éloigne positivement de 0.
Pour α donné, nous avons :

$$\alpha = \mathbb{P}_{H_0}(K_n > K) + \pi \mathbb{P}_{H_0}(K_n = K).$$

Nous déterminons avant tout la valeur de l telle que :

$$\mathbb{P}_{H_0}(K_n > K) \leq \alpha < \mathbb{P}_{H_0}(K_n \geq K).$$

Puis, nous ajustons à l'aide de π :

$$\pi = \frac{\alpha - \mathbb{P}_{H_0}(K_n > K)}{\mathbb{P}_{H_0}(K_n = K)}.$$

Aussi, sous H_0, $K_n \sim \mathcal{B}\left(n, \frac{1}{2}\right)$. En effet, sous H_0 la médiane étant nulle, il y a autant de chance pour que la réalisation de la variable aléatoire soit négative ou positive.

2 Nous avons $n = 5$ et $\alpha = 0.05$.
Nous obtenons : $K = 4$ et $\pi = \frac{0.05 - 0.0313}{0.1562} \simeq 0.12$. La stratégie de décision est donc :
– si $k_5 > 4$ (et donc $k_5 = 5$), nous rejetons H_0 ;
– si $k_5 = 4$, nous rejetons H_0 avec la probabilité 0.12 ;
– si $k_5 < 4$, nous ne pouvons pas rejeter H_0.

Exercice 2

1 Nous avons $\overline{X} \sim \mathcal{N}\left(\theta_0, \frac{\sigma^2}{n}\right)$, $\overline{Y} \sim \mathcal{N}\left(\theta_1, \frac{\sigma^2}{m}\right)$,
$\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sim \chi^2(n-1)$, $\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 \sim \chi^2(m-1)$.

2 Nous souhaitons tester l'hypothèse H_0 selon laquelle $\theta = \theta_0$ contre l'hypothèse alternative H_1 selon laquelle $\theta = \theta_1$.
Posons

$$T = \frac{\left(\overline{X} - \overline{Y}\right) / \sqrt{1/n + 1/m}}{\sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 + \sum_{i=1}^{m} \left(Y_i - \overline{Y}\right)^2} / \sqrt{n + m - 2}}$$

Une région critique du test précédent est telle que nous rejetons l'hypothèse H_0 si $\{|t| > k\}$.

Soit α le risque de première espèce du test précédent.
Nous avons $\mathbb{P}\left(|T| > k \,|H_0\right) = 2 - 2F_{T(n+m-2)}(k) = \alpha$.
Par conséquent, $k = F^{-1}_{T(n+m-2)}\left(1 - \frac{\alpha}{2}\right)$. Il s'agit du test de Student.

Exercice 1

1 Rappelons tout d'abord la définition de la loi du χ^2. Si $X_1, \ldots, X_n$ sont n variables aléatoires i.i.d. de loi $\mathcal{N}(0,1)$, alors la somme $X_1^2 + ... + X_n^2$ suit une loi du χ^2 à n degrés de liberté, notée $\chi^2(n)$, cette loi a pour densité

$$f_{\chi^2(n)}(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} 1_{\{x>0\}}(x).$$

Revenons maintenant à notre problème, $Z = \frac{2}{\theta}\sqrt{X}$. On effectue ce changement de variable pour la densité $f_\theta(x)dx$ donnant ainsi la densité f_Z de Z :

$$\begin{aligned}
z = \frac{2}{\theta}\sqrt{x} \quad \Rightarrow \quad x = \frac{\theta^2 z^2}{4}, \quad dx = \frac{1}{2}\theta^2 z dz \\
f_Z(z) &= \frac{1}{2\theta \frac{\theta}{2} z} e^{-z/2} \frac{1}{2} \theta^2 z 1_{\mathbb{R}_+}(z) \\
&= \frac{1}{2} e^{-z/2} 1_{\mathbb{R}_+}(z)
\end{aligned}$$

ce qui implique au regard de la définition du $\chi^2(n)$ que $Z \sim \chi^2(2)$. De plus, par indépendance

$$\frac{2}{\theta}\sum_{i=1}^n \sqrt{X_i} = \sum_{i=1}^n Z_i \sim \chi^2(2n).$$

2 On applique ici le lemme de Neyman-Pearson : pour $x_i \in \mathbb{R}_+$

$$\begin{aligned}
\frac{f(x;\theta_1)}{f(x;\theta_0)} &= \frac{\left(\frac{1}{2\theta_1}\right)^n \prod_{i=1}^n \frac{1}{\sqrt{x_i}} e^{-\frac{1}{\theta_1}\sqrt{x_i}}}{\left(\frac{1}{2\theta_0}\right)^n \prod_{i=1}^n \frac{1}{\sqrt{x_i}} e^{-\frac{1}{\theta_0}\sqrt{x_i}}} \\
&= \left(\frac{\theta_0}{\theta_1}\right)^n e^{\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{i=1}^n \sqrt{x_i}}
\end{aligned}$$

d'où

$$\begin{aligned}
\frac{f(x;\theta_1)}{f(x;\theta_0)} \geq K \quad &\Longleftrightarrow \quad \log \frac{f(x;\theta_1)}{f(x;\theta_0)} \geq \log K \\
&\Longleftrightarrow \quad n \log \frac{\theta_0}{\theta_1} + \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \sum_{i=1}^n \sqrt{x_i} \geq \log K \\
&\Longleftrightarrow \quad \sum_{i=1}^n \sqrt{x_i} \geq \frac{\log K - n \log \frac{\theta_0}{\theta_1}}{\frac{1}{\theta_0} - \frac{1}{\theta_1}} = K'.
\end{aligned}$$

Le test donne donc : on rejette H_0 si $\sum_{i=1}^n \sqrt{x_i} \geq K'$, et nous ne pouvons pas rejeter H_0 sinon. La région critique est donc :

$$W = \left\{ x \in \mathbb{R}_n^+ \ ; \ \sum_{i=1}^n \sqrt{x_i} \geq K' \right\}.$$

Pour le test de niveau α :

$$\begin{aligned}
\mathbb{P}_{\theta_0}(W) = \alpha \quad & \Longleftrightarrow \quad \mathbb{P}_{\theta_0}\left(\sum_{i=1}^n \sqrt{X_i} \geq K' \right) = \alpha \\
& \Longleftrightarrow \quad \mathbb{P}_{\theta_0}\left(\frac{2}{\theta_0} \sum_{i=1}^n \sqrt{X_i} \geq \frac{2}{\theta_0} K' \right) = \alpha \\
& \Longleftrightarrow \quad \mathbb{P}\left(\chi^2(2n) \geq \frac{2}{\theta_0} K' \right) = \alpha \\
& \Longleftrightarrow \quad \mathbb{P}\left(\chi^2(2n) \leq \frac{2}{\theta_0} K' \right) = 1 - \alpha \\
& \Longleftrightarrow \quad \frac{2}{\theta_0} K' = c_{1-\alpha}
\end{aligned}$$

où $c_{1-\alpha}$ est le quantile d'ordre α de la loi $\chi^2(2n)$. Soit finalement, on rejette l'hypothèse H_0 si $\sum_{i=1}^n \sqrt{x_i} \geq \frac{\theta_0}{2} c_{1-\alpha}$, sinon on ne peut pas rejeter l'hypothèse H_0.

La puissance du test est égale à :

$$p(\theta_1) = \mathbb{P}_{\theta_1}(W) = \mathbb{P}_{\theta_1}\left(\sum_{i=1}^n \sqrt{x_i} \geq \frac{\theta_0}{2} c_{1-\alpha} \right).$$

Or sous H_1, on a en fait $\frac{2}{\theta_1} \sum_{i=1}^n \sqrt{X_i} \sim \chi^2(2n)$ d'où

$$p(\theta_1) = \mathbb{P}_{\theta_1}\left(\frac{2}{\theta_1} \sum_{i=1}^n \sqrt{X_i} \geq \frac{\theta_0}{\theta_1} c_{1-\alpha} \right) = \mathbb{P}\left(\chi^2(2n) \geq \frac{\theta_0}{\theta_1} c_{1-\alpha} \right).$$

Exercice 3

1 Soit $x = (x_1, ..., x_n)$, la densité de probabilité de x est égale à :

$$f_\theta(x) = \frac{\theta^{n/2}}{(2\pi)^{n/2}} \exp\left(-\frac{\theta}{2} \sum_{i=1}^n x_i^2 \right).$$

Nous utilisons le lemme de Neyman-Pearson pour déterminer le test UPP de $\theta = 1$ contre $\theta > 1$ soit

$$\frac{f_\theta(x)}{f_1(x)} > K \quad \Longleftrightarrow \quad (\theta)^{n/2} \exp\left(\left(\frac{1-\theta}{2} \right) \sum_{i=1}^n x_i^2 \right) > K$$

$$\iff \quad \left(\frac{1-\theta}{2}\right)\sum_{i=1}^{n} x_i^2 > \log K - \frac{n}{2}\log\theta$$

$$\iff \quad \sum_{i=1}^{n} x_i^2 < \frac{\log K - \frac{n}{2}\log\theta}{\left(\frac{1-\theta}{2}\right)} = K'$$

car $\frac{1-\theta}{2} < 0$ puisque $\theta > 1$. Le test UPP de niveau α est donc :

- si $\sum_{i=1}^{n} x_i^2 < K'$, nous rejetons H_0 ;
- si $\sum_{i=1}^{n} x_i^2 > K'$, nous ne pouvons pas rejeter H_0.

Cette stratégie de décision est très intuitive. En effet, lorsque l'estimateur empirique de la variance est faible il est légitime de penser que H_1 est vraie. Il reste ensuite à calculer K' en fonction du niveau du test considéré.

$$\mathbb{P}_{H_0}\left(\sum_{i=1}^{n} X_i^2 < K'\right) = \alpha \iff \mathbb{P}\left(\chi^2(n) < K'\right) = \alpha \iff K' = F^{-1}_{\chi^2(n)}(\alpha)$$

où $F^{-1}_{\chi^2(n)}$ est l'inverse de la fonction de répartition de la loi du χ^2.

2 Calcul de la puissance.

$$\begin{aligned}
p(\theta) &= \mathbb{P}_{H_1}\left(\sum_{i=1}^{n} X_i^2 < F^{-1}_{\chi^2(n)}(\alpha)\right) \\
&= \mathbb{P}_{H_1}\left(\sum_{i=1}^{n} (\sqrt{\theta}X_i)^2 < \theta F^{-1}_{\chi^2(n)}(\alpha)\right) \\
&= \mathbb{P}\left(\chi^2(n) < \theta F^{-1}_{\chi^2(n)}(\alpha)\right) \\
&= F_{\chi^2(n)}\left[\theta F^{-1}_{\chi^2(n)}(\alpha)\right].
\end{aligned}$$

3 Application numérique à $n = 10$ et $\alpha = 0.05$.
On obtient ainsi

$$K' = F^{-1}_{\chi^2(10)}(0.05) = 3.94$$

$$p(\theta) = F_{\chi^2(10)}(3.94\ \theta)$$

soit pour $\theta = 4$, $p(4) = 0.9$.