Final SY02 Automne 2006

Nom :	**Signature :**
Prénom :	

Répondre sur ce document, en ne reportant que les grandes lignes du raisonnement et les résultats (faire d'abord les calculs au brouillon). La qualité de la présentation sera prise en compte dans la notation. Aucune copie supplémentaire ne sera acceptée. Le seul document autorisé est le recueil de tables. Le stockage d'informations relatives au cours de sy02 sur une calculette programmable est interdit.

Exercice 1 (7 points)

Une v.a. aléatoire X suit une loi $\mathcal{N}(\mu, \sigma^2)$ d'écart-type connu $\sigma = 4$. Au vu d'un échantillon i.i.d. $X_1, \ldots, X_n$ de v.a. parente X, on veut tester l'hypothèse $H_0 : \mu = 1$ contre l'hypothèse alternative $H_1 : \mu = 5$.

1. Soit $L(\mu; x_1, \ldots, x_n)$ la fonction de vraisemblance. Donner l'expression du rapport $\lambda = \frac{L(5;x_1,\ldots,x_n)}{L(1;x_1,\ldots,x_n)}$.

$$\begin{aligned}
\lambda &= \frac{(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-5)^2\right)}{(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-1)^2\right)} \\
&= \exp\left(\frac{n}{2\sigma^2}(8\overline{x}-24)\right) \\
&= exp\left(\frac{n}{4}(\overline{x}-3)\right)
\end{aligned}$$

2. En déduire la région critique du test de Neyman-Pearson, en fonction du niveau de signification α^*.

La RC du test de NP est de la forme $W = \{\lambda > c\}$. Or, λ est fonction de croissante de $\overline{x}$. Donc, W est de la forme $W = \{\overline{x} > k\}$, où la constante k est solution de l'équation

$$\mathbb{P}_{H_0}(\overline{X} > k) = \alpha^*.$$

Sachant que sous H_0 $\overline{X} \sim \mathcal{N}(1, 4^2/n)$, on trouve

$$k = 1 + \frac{4u_{1-\alpha^*}}{\sqrt{n}}.$$

3. Calculer la puissance de ce test dans le cas où $n = 5$ et $\alpha^* = 0.05$.

$$\begin{aligned}
\pi &= \mathbb{P}_{H_1}(W) \\
&= 1 - \phi\left(\frac{1 + \frac{4u_{1-\alpha^*}}{\sqrt{n}} - 5}{4/\sqrt{n}}\right) \\
&= 1 - \phi\left(u_{1-\alpha^*} - \sqrt{n}\right) = \phi\left(\sqrt{n} - u_{1-\alpha^*}\right)
\end{aligned}$$

A.N. : $\pi = \phi(0.59) = 0.72$.

4. Quelle doit être la taille minimum n_0 de l'échantillon pour que la puissance soit supérieure à 0.95, en supposant toujours $\alpha^* = 0.05$?

$$\begin{aligned}
\pi > 0.95 &\Leftrightarrow \phi\left(\sqrt{n} - u_{1-\alpha^*}\right) > 0.95 \\
&\Leftrightarrow n > (u_{1-\alpha^*} - u_{0.05})^2 = 10.82
\end{aligned}$$

Donc, $n_0 = 11$.

5. On a observé $\overline{x} = 6$ toujours avec $n = 5$. Quel est le degré de signification du test ?

$$\begin{aligned}
\widehat{\alpha}(x_1, \ldots, x_n) &= \mathbb{P}_{H_0}(\overline{X} > 6) \\
&= 1 - \phi\left(\frac{6-1}{4/\sqrt{n}}\right) \\
&= 1 - \phi(2.79) = 1 - 0.9974 \approx 0.003
\end{aligned}$$

Exercice 2 (5 points)

On a demandé à 500 personnes de choisir au hasard un nombre entre 1 et 10. On obtient les résultats suivants :

Nombre choisi	1	2	3	4	5	6	7	8	9	10
fréquence observée	63	54	30	31	52	49	46	57	58	60

D'après ces données, peut-on accepter, au niveau de signification de 1 %, l'hypothèse selon laquelle les nombres ont été choisi de manière uniforme.

On fait un test du χ^2.

k	1	2	3	4	5	6	7	8	9	10
p_k	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
np_k	50	50	50	50	50	50	50	50	50	50
n_k	63	54	30	31	52	49	46	57	58	60

On obtient

$$D^2 = \sum \frac{n_k^2}{np_k} - n = 23.60$$

et $\chi^2_{9;0.99} = 21.7$.
On rejette donc H_0 au niveau $\alpha^* = 0.01$.

Exercice 3 (8 points)

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de variable parente X, de densité $f(x) = \frac{\theta^4}{6}\, x^3 e^{-\theta x} 1_{[0,+\infty[}(x)$ θ étant un paramètre positif.

1. Donner l'expression de $\widehat{\theta}$, estimateur du maximum de vraisemblance de θ.

$$L(\theta, x_1, \ldots, x_n) = \theta^{4n} 6^{-n} (\prod x_i)^3 e^{-\theta \sum_i x_i}$$

$$\ln L(\theta, x_1, \ldots, x_n) = 4n \ln(\theta) - \theta n \bar{x} + cste$$

$$\frac{\partial \ln L}{\partial \theta}(\theta, x_1, \ldots, x_n) = \frac{4n}{\theta} - n\bar{x}$$

$$\frac{\partial \ln L}{\partial \theta}(\theta, x_1, \ldots, x_n) = 0 \Leftrightarrow \theta = \frac{4}{\bar{x}}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2}(\theta, x_1, \ldots, x_n) = -\frac{4n}{\theta^2}$$

La dérivée seconde est donc négative pour $\frac{4}{\bar{x}}$ et on obtient

$$\hat{\theta} = \frac{4}{\bar{X}}$$

2. Calculer l'information de Fisher $I_n(\theta)$ relative au paramètre θ. En déduire une fonction asymptotiquement pivotale pour θ que l'on exprimera en fonction de $\widehat{\theta}$.

Information de Fisher

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}(\theta, X_1, \ldots, X_n)\right) = -E\left(-\frac{4n}{\theta^2}\right) = \frac{4n}{\theta^2}$$

Fonction asymptotiquement pivotale de θ (en fonction de $\widehat{\theta}$)
En utilisant les propriétés asymptotiques d'un estimateur du maximum de vraisemblance, on obtient

$$\frac{\widehat{\theta} - \theta}{1/\sqrt{I_n(\theta)}} = \frac{\widehat{\theta} - \theta}{\theta/(2\sqrt{n})} \xrightarrow{L} \mathcal{N}(0, 1)\,.$$

3. On considère le problème de test $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ avec $\theta_1 > \theta_0$. Montrer que la région critique W du test le plus puissant pour ce problème au niveau α^* s'exprime en fonction de $\widehat{\theta}$, puis donner une approximation de W en supposant n grand.

Expression de W en fonction de $\widehat{\theta}$
Il s'agit d'un problème de comparaison de deux hypothèses simples. Le test le plus puissant est donné par le théorème de Neyman-Pearson.

$$\begin{aligned}\frac{L(\theta_1; x_1, \ldots, x_n)}{L(\theta_0; x_1, \ldots, x_n)} &= \left(\frac{\theta_1}{\theta_0}\right)^{4n} \exp\left((\theta_0 - \theta_1)\sum_{i=1}^{n} x_i\right) \\ &= \left(\frac{\theta_1}{\theta_0}\right)^{4n} \exp\left(\frac{4n}{\widehat{\theta}}(\theta_0 - \theta_1)\right)\end{aligned}$$

Comme $\theta_0 - \theta_1 < 0$, ce rapport est fonction croissante de $\widehat{\theta}$. La région critique du test de Neyman-Pearson, de la forme

$$W = \left\{\frac{L(\theta_1; x_1, \ldots, x_n)}{L(\theta_0; x_1, \ldots, x_n)} > k\right\},$$

peut donc s'écrire $W = \{\widehat{\theta} > k'\}$ pour une certaine constante k'.

Approximation de W pour n grand
On obtient une expression approchée pour k' en imposant que le niveau du test soit approximativement égal à une valeur α^*.

$$\Pr_{H_0}(\widehat{\theta} > k') = \alpha^* \Leftrightarrow \Pr_{H_0}\left(\frac{\widehat{\theta} - \theta_0}{\theta_0/(2\sqrt{n})} > \frac{k' - \theta_0}{\theta_0/(2\sqrt{n})}\right) = \alpha^*$$

D'après la question 2, $\frac{\widehat{\theta} - \theta_0}{\theta_0/(2\sqrt{n})}$ suit approximativement pour H_0, pour n grand, une loi normale centrée-réduite. On en déduit

$$\begin{aligned}&\frac{k' - \theta_0}{\theta_0/(2\sqrt{n})} \approx u_{1-\alpha^*} \\ \Leftrightarrow\ & k' \approx \theta_0\left(1 + \frac{u_{1-\alpha^*}}{(2\sqrt{n})}\right).\end{aligned}$$

4. On considère maintenant le problème de test suivant $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. Existe-t-il un test UPP pour ce problème ?

Soit le problème de test $H_0 : \theta = \theta_0$ contre $h_1 : \theta = \theta_1$ avec $\theta_1 \neq \theta_0$.
On a vu que, si $\theta_1 > \theta_0$, la RC du test de NP est de la forme $W = \{\widehat{\theta} > k'\}$.
En revanche, si $\theta_1 < \theta_0$, le rapport des vraisemblances est fonction décroissante de $\widehat{\theta}$, et la RC du test de NP est donc de la forme $W = \{\widehat{\theta} < k''\}$.
La région critique du test de NP dépend donc de l'hypothèse simple h_1 considérée : par conséquent, il n'existe pas de test UPP.

5. Calculer la statistique du rapport de vraisemblance λ, exprimée en fonction de $\widehat{\theta}$, pour le problème de test de la question 4.

$$\begin{aligned}
\lambda &= \frac{sup_{\theta \in H_0} L(\theta; x_1, \ldots, x_n)}{sup_{\theta \in \Theta} L(\theta; x_1, \ldots, x_n)} \\
&= \frac{\theta_0^{4n} \exp(-4n\theta_0/\widehat{\theta})}{\widehat{\theta}^{4n} \exp(-4n\widehat{\theta}/\widehat{\theta})} \\
&= \left(\frac{\theta_0}{\widehat{\theta}}\right)^{4n} \exp\left[4n\left(1 - \frac{\theta_0}{\widehat{\theta}}\right)\right]
\end{aligned}$$

6. En utilisant la statistique $\ln \lambda$ et en supposant que n est grand, proposer une région critique pour le test de la question 4. Quelle décision prendra-t-on si $\theta_0 = 2$, $n = 50$, $\sum_i x_i = 115$ et $\alpha^* = 0.05$.

Région critique
La fonction $-2\ln\lambda$ suit asymptotiquement, sous H_0, une loi du χ^2_1. La région critique du test du RV est donc

$$W = \{-2\ln\lambda \geq c\} \qquad \text{avec} \qquad c \simeq \chi^2_{1;1-\alpha^*}.$$

Application numérique
Région critique : $W = \{-2\ln\lambda \geq \chi^2_{1;0.95} = 3.84\}$
Calcul de $-2\ln\lambda$:

$$\begin{aligned}
-2\ln\lambda &= -2\left(4n(\ln\theta_0 - \ln\widehat{\theta}) + 4n\left(1 - \frac{\theta_0}{\widehat{\theta}}\right)\right) \\
&= -8n\left(\ln\theta_0 - \ln\widehat{\theta} + 1 - \frac{\theta_0}{\widehat{\theta}}\right) \\
&= 4.09
\end{aligned}$$

On rejette donc l'hypothèse H_0.