Final SY02 Automne 2007

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 (10 points)

Un producteur de vin désire savoir si sa production est conforme aux normes imposées et, en particulier, si le *titre alcoométrique volumique naturel* (c'est-à-dire la quantité d'alcool naturellement présente dans le vin, mesurée par rapport au volume total) est bien compris dans la fourchette $[11\%, 14\%]$ imposée par le gouvernement. Le vigneron souhaite donc estimer le titrage moyen de sa dernière cuvée par une série de mesures réalisées sur un nombre n de cuves. On supposera que la mesure du titre alcoométrique volumique naturel X suit une loi normale d'espérance μ et de variance σ^2 où μ est le titrage moyen.

1. Proposer un estimateur du titrage moyen et une fonction pivotale de ce paramètre en s'appuyant sur cet estimateur.

On peut prendre $\widehat{\theta} = \overline{X}$ qui est un estimateur sans biais et convergent de l'espérance $\mathbb{E}(X) = \mu$.
Sa fonction pivotale est

$$\frac{\overline{X} - \mu}{\frac{S^*}{\sqrt{n}}} \sim \mathcal{T}_{n-1}$$

.

2. Donner (avec démonstration) l'expression d'un intervalle de confiance bilatéral au niveau $1 - \alpha$ sur le titrage moyen. Application numérique : quel est l'intervalle de confiance sachant que on a obtenu 10 mesures $x_1, \ldots, x_{10}$ vérifiant $\sum_i x_i = 130.5$ et $\sum_i x_i^2 = 1713.73$ et que l'on prend $\alpha = 0.05$?

Intervalle de confiance :
Partant de la fonction pivotale, on obtient

$$\mathbb{P}\left(-u_{1-\alpha^*/2} < \frac{\overline{X} - \mu}{\frac{S^*}{\sqrt{n}}} < u_{1-\alpha^*/2}\right) = 1 - \alpha$$

$$\mathbb{P}\left(\overline{X} - \frac{S^*}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}} < \mu < \overline{X} + \frac{S^*}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$IC = \left[\overline{X} - \frac{S^*}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \overline{X} + \frac{S^*}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}\right]$$

Application numérique :
$\overline{x} = 13.05$, $s^{*2} = \frac{1}{9}(1713.73 - 1013.05^2) = 1.19$, $s^* = 1.09$, $t_{9,0.975} = 2.26$, $IC = [13.05 - 0.78, 13.05 + 0.78] = [12.27, 13.83]$.

3. On cherche à tester l'hypothèse $H_0 : \sigma^2 = \sigma_O^2$ vs. l'hypothèse $H_1 : \sigma^2 \neq \sigma_O^2$. Donner la fonction pivotale à employer, puis en déduire la région critique du test. Application numérique : peut-on accepter l'hypothèse $H_0 : \sigma^2 = 1$ au niveau de signification $\alpha^* = 0.05$, si l'on utilise les données de la question 2 ?

Fonction pivotale :

$$\frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2_{n-1}$$

Région critique :
Si on note T la statistique $\frac{(n-1)S^{*2}}{\sigma_0^2}$, on a vu en cours que la région critique prend la forme :

$$W = \{t < A \quad \text{ou} \quad t > B\}$$

$$\mathbb{P}\left(T < A \quad \text{ou} \quad T > B\right) = \alpha^*$$

D'où $A = \chi^2_{n-1,\frac{\alpha^*}{2}}$ et $B = \chi^2_{n-1,1-\frac{\alpha^*}{2}}$ et finalement

$$W = \left\{ \frac{(n-1)s^{*2}}{\sigma_0^2} < \chi^2_{n-1,\frac{\alpha^*}{2}} \quad \text{ou} \quad \frac{(n-1)s^{*2}}{\sigma_0^2} > \chi^2_{n-1,1-\frac{\alpha^*}{2}} \right\}.$$

Application numérique :
$\frac{(n-1)s^{*2}}{\sigma_0^2} = \frac{9\times 1.19}{1} = 10.71$, $\chi^2_{n-1,\frac{\alpha^*}{2}} = \chi^2_{9,0.025} = 2.70$, $\chi^2_{n-1,1-\frac{\alpha^*}{2}} = \chi^2_{9,0.975} = 19$
On conserve donc l'hypothèse H_0.

On suppose à présent que la variance σ^2 est connue et égale à $\sigma_0^2 = 1$.

4. Donner (sans démonstration) l'expression d'un nouvel intervalle de confiance sur le titrage moyen. Application numérique avec les mêmes données que dans la question 2. Combien de mesures doit-on réaliser pour que sa largeur n'excède pas 0.5 ?

Intervalle de confiance

$$IC = \left[\overline{X} - \frac{\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}}, \overline{X} + \frac{\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right]$$

Application numérique
$u_{0.975} = 1.96$, $IC = [13.05 - 0.62, 13.05 + 0.62] = [12.43, 13.67]$.

Nombre de mesures : La largeur de l'intervalle de confiance est égale à $\ell = \frac{2\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$. On obtient donc

$$\ell \leq \varepsilon$$

$$\frac{2\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \varepsilon$$

$$n \geq \left(\frac{2\sigma_0 u_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2$$

$$n \geq 62.$$

5. Tester l'hypothèse $H_0 : \mu = \mu_0$ vs. l'hypothèse $H_1 : \mu \neq \mu_0$. Application numérique : que décide-t-on pour $\mu_0 = 13$ avec les données de la question 2 ?

Région critique :
En utilisant la statistique de test $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}$ qui suit sous l'hypothèse H_0 une loi normale centrée réduite, on obtient la région critique

$$W = \{|z| > u_{1-\frac{\alpha^*}{2}}\}.$$

Application numérique :
$z = \frac{\overline{x}-\mu_0}{\sigma_0}\sqrt{n} = \frac{13.05-13}{1/\sqrt{10}} = 0.158$, $u_{1-\frac{\alpha^*}{2}} = 1.96$. On conserve donc l'hypothèse H_0.

Exercice 2 (5 points)

On cherche à estimer la probabilité p qu'un dindon mécanique tombe en panne. Pour cela, on modélise la durée de vie en secondes par une variable aléatoire X qui est supposée suivre une loi géométrique de paramètre $p : \mathbb{P}(X = x) = (1-p)^{x-1}p \quad$ pour $x \in \mathbb{N}^*$. La mise en service de $n = 100$ dindons mécaniques a donné les résultats suivants :

date de panne	1	2	3	4	5
nombre d'occurences	57	26	10	5	2

1. Existe-t-il un estimateur efficace de p ou d'une fonction de p ? Si oui, donner son espérance et sa variance. En déduire l'espérance et la variance de la variable aléatoire X.

Estimateur efficace :

$$L(p; x_1, \ldots, x_n) = \prod_i \left((1-p)^{x_i-1}p\right) = (1-p)^{n(\overline{x}-1)}p^n$$

$$\ln L(p; x_1, \ldots, x_n) = n(\overline{x}-1)\ln(1-p) + n\ln p$$

$$\frac{\partial \ln L}{\partial p}(p; x_1, \ldots, x_n) = n(\overline{x}-1)\frac{-1}{1-p} + \frac{n}{p}$$

$$\frac{\partial \ln L}{\partial p}(p; x_1, \ldots, x_n) = \frac{n(1-p\overline{x})}{p(1-p)}$$

On peut donc écrire

$$\frac{\partial \ln L}{\partial p}(p; X_1, \ldots, X_n) = -\frac{n}{1-p}(\overline{X} - \frac{1}{p})$$

Sachant que le support de X, égal à l'ensemble $\mathbb{N}^*$, ne dépend pas du paramètre p, on peut en déduire (CNS d'efficacité) que $\overline{X}$ est l'estimateur efficace de $\frac{1}{p}$.

Espérance et variance de l'estimateur efficace :
Un estimateur efficace est sans biais : $\mathbb{E}(\overline{X}) = \frac{1}{p}$.
La CNS d'efficacité entraîne que $\mathrm{Var}(\overline{X}) = \frac{-\frac{1}{p^2}}{-\frac{n}{1-p}} = \frac{1-p}{np^2}$.

Espérance et variance de X :
$\mathbb{E}(X) = \mathbb{E}(\overline{X}) = \frac{1}{p}$ et $\mathrm{Var}(X) = n\mathrm{Var}(\overline{X}) = \frac{1-p}{p^2}$.

2. Au vu des résultats, peut-on admettre, au niveau de signification $\alpha^* = 5\%$, que la variable aléatoire X suit une loi géométrique de paramètre $p = 0.6$? (On utilisera un test du χ^2.)

k	1	2	3	4	>= 5
n_k	57	26	10	5	2
p_k	0.6	0.24	0.096	0.0384	0.0256
np_k	60	24	9.6	3.84	2.54

Les effectifs des deux dernières classes sont inférieurs à 5 ; on les regroupe donc pour obtenir :

k	1	2	3	>= 4
n_k	57	26	10	7
p_k	0.6	0.24	0.096	0.064
np_k	60	24	9.6	6.4

Finalement, la réalisation de la statistique de test $d^2 = \frac{3^2}{60} + \frac{2^2}{24} + \frac{0.4^2}{9.6} + \frac{0.6^2}{6.4} = 0.39$.
Sachant que le région critique s'écrit

$$w = \{d^2 > \chi^2_{K-1,1-\alpha^*} = \chi^2_{3,0.95} = 7.81\},$$

on accepte l'hypothèse H_0.

Rq. Si le regroupement n'est pas fait, on obtient $d^2 = 0.82$, $w = \{d^2 > 9.49\}$ et la même décision.

Exercice 3 (5 points)

On a relevé les durées de vie $x_1, \ldots, x_n$ de n appareils de même type. On suppose que la durée de vie X d'un appareil suit une loi exponentielle de paramètre θ inconnu. On rappelle que la densité de cette loi s'écrit $f(x) = \theta e^{-\theta x}$. Le fabricant prétend que la durée de vie moyenne est de 1000 heures, c'est-à-dire que le paramètre θ est égal à $1/1000$. Ce chiffre est considéré comme sur-estimé par une association des consommateurs. Nous allons utiliser la théorie des tests pour tenter de trancher cette question.

1. On considère le problème de test suivant : $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ avec $\theta_1 > \theta_0$. Donner la forme de la région critique du test optimal au niveau α^* pour ce problème.

Sachant que l'on a 2 hypothèses simples, on peut utiliser le théorème de NP : on a

$$L(\theta; x_1, \ldots, x_n) == \prod_i (\theta e^{-\theta x_i}) = \theta^n e^{-n\theta \overline{x}}$$

Le rapport de vraisemblance est donc :

$$\frac{L(\theta_1; x_1, \ldots, x_n)}{L(\theta_0; x_1, \ldots, x_n)} = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-n(\theta_1 - \theta_0)\overline{x}}$$

où $\widehat{\lambda} = \overline{X}$.
C'est une fonction décroissante de $\overline{X}$. Donc, d'après le théorème de N-P, la RC optimale au niveau α^* est de la forme $W = \{\overline{x} < A\}$ pour une constante A vérifiant

$$P_{\theta_0}(\overline{X} < A) = \alpha^*.$$

2. En supposant n grand, déterminer l'expression littérale de la région critique.

Expression littérale
D'après le TLC, on a approximativement pour n grand et sous l'hypothèse H_0 :

$$\overline{X} \sim \mathcal{N}\left(\frac{1}{\theta_0}, \frac{1}{n\theta_0^2}\right)$$

On en déduit :

$$\mathbb{P}_{H_0}\left(\frac{\overline{X} - \frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}} < \frac{A - \frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}}\right) = \alpha^*$$

$$\frac{A - \frac{1}{\theta_0}}{\sqrt{\frac{1}{n\theta_0^2}}} = u_\alpha^* \Rightarrow A = \frac{1}{\theta_0}\left(1 - \frac{u_{1-\alpha^*}}{\sqrt{n}}\right)$$

et finalement

$$W = \left\{\overline{x} < \frac{1}{\theta_0}\left(1 - \frac{u_{1-\alpha^*}}{\sqrt{n}}\right)\right\}$$

4. On considère cette fois les hypothèses $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$. Montrer qu'il existe un test UPP pour ce problème.

Soit le test intermédiaire $H_0 : \theta = \theta_0$ vs. $h_1 : \theta = \theta_1$ avec $\theta_1 > \theta_0$. La RC optimale pour ce problème a été trouvée précédemment et ne dépend pas de θ_1. Par conséquent, le test est UPP et la RC est la RC critique obtenue avec ce test intermédiaire :

$$W = \left\{\overline{x} < \frac{1}{\theta_0}\left(1 - \frac{u_{1-\alpha^*}}{\sqrt{n}}\right)\right\}$$