Final SY02 Automne 2008

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. Les calculettes sont autorisées à condition qu'elles ne contiennent aucune information relative au cours de sy02.

Exercice 1 (6 points)

On cherche à comparer la durée de vie de deux types de pneu A et B. On dispose pour cela d'un échantillon de 41 durées de vie en milliers de km pour le type A et de 21 durées de vie pour le type B. Les résultats sont résumés dans le tableau suivant :

	n	$\sum x_i$	$\sum x_i^2$
A	41	1840	82996
B	21	828	32752

On admettra que les 2 populations suivent les distributions normales $\mathcal{N}(\mu_A, \sigma_A^2)$ et $\mathcal{N}(\mu_B, \sigma_B^2)$ et dans tout cet exercice, on prendra comme niveau de signification des différents tests la valeur $\alpha^* = 0.05$.

1. Donner les estimations sans biais de μ_A, μ_B, σ_A^2 et σ_B^2.

$$\overline{x}_A = 44.88$$

$$\overline{x}_B = 39.43$$

$$s_A^{*2} = \frac{82996 - 41(44.88)^2}{40} = 10.33$$

$$s_B^{*2} = \frac{32752 - 21(39.43)^2}{20} = 5.14$$

2. Montrer que l'on peut admettre l'hypothèse d'égalité des variances des 2 populations.

Test de Fisher :

$$W = \left\{ \frac{s_A^{*2}}{s_B^{*2}} < F_{n_A-1,n_B-1;\frac{\alpha^*}{2}} \text{ ou } \frac{s_A^{*2}}{s_B^{*2}} > F_{n_A-1,n_B-1;1-\frac{\alpha^*}{2}} \right\}$$

$$F_{n_A-1,n_B-1;\frac{\alpha^*}{2}} = F_{40,20,0.025} = \frac{1}{F_{20,40,0.975}} = \frac{1}{2.07} = 0.483$$

$$F_{n_A-1,n_B-1;1-\frac{\alpha^*}{2}} = F_{40,20,0.975} = 2.29$$

$$\frac{s_A^{*2}}{s_B^{*2}} = \frac{10.33}{5.14} = 2.01$$

Nous ne sommes donc pas dans la région critique : on accepte l'hypothèse d'égalité des variances.

3. En déduire une estimation sans biais de la variance commune σ^2.

$$s^{*2} = \frac{(n_A-1)s_A^{*2} + (n_B-1)s_B^{*2}}{n_A+n_B-2} = \frac{40 \times 10.33 + 20 \times 5.14}{60} = 8.06$$

4. Tester l'égalité des moyennes μ_A et μ_B.

Test de Student

$$W = \left\{ \frac{|\overline{x}_A - \overline{x}_B|}{s^* \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{n_A+n_B-2, 1-\frac{\alpha^*}{2}} \right\}$$

$$t_{n_A+n_B-2,1-\frac{\alpha^*}{2}} = t_{60;0.975} = 2$$

$$\frac{|\overline{x}_A - \overline{x}_B|}{s^* \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{|44.83 - 39.43|}{\sqrt{8.06 \times (1/41 + 1/21)}} = 7.08$$

Nous sommes donc dans la région critique : on ne peut donc pas accepter l'hypothèse d'egalité des moyennes.

Exercice 2 (5 points)

On considère la réalisation suivante d'un échantillon iid de v.a. parente X :

59 44 75 37 3

Peut-on admettre au niveau $\alpha^* = 0.05$ que X suit une loi normale d'espérance 50 et de variance 100 ? (On utilisera un test de Kolmogorov-Smirnov.)

La région critique du test de K-S est la suivante

$$W = \{d_n > d_{n,1-\alpha^*}\}$$

Calcul de d_n (on a noté $z_i = (x_i - \mu)/\sigma$) :

x_i	z_i	$\widehat{F}(x_i)$	$F_0(x_i) = \phi(z_i)$	$\lvert\widehat{F}(x_i) - F_0(x_i)\rvert$	$\lvert\widehat{F}(x_i^-) - F_0(x_i)\rvert$
3	-4.70	0.20	0.00	0.20	0.00
37	-1.30	0.4	0.10	0.30	0.10
44	-0.60	0.6	0.27	**0.33**	0.13
59	0.90	0.8	0.82	0.02	0.22
75	2.50	1	0.99	0.01	0.19

On a donc $d_n^* = 0.33$.
Sachant que $d_{5,0.95} = 0.563$, on ne rejette pas l'hypothèse H_0.

Exercice 3 (9 points)

On considère dans ce problème une variable aléatoire X de fonction de densité

$$f(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x} 1_{[0,+\infty[}(x)$$

avec $\lambda > 0$. On dispose d'un échantillon i.i.d $(X_1, \dots, X_n)$ de v.a. parente X.

1. Déterminer l'estimateur du maximum de vraisemblance du paramètre λ, que l'on notera $\hat{\lambda}$.

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

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

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

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

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

La dérivée seconde est donc négative pour $\lambda = \frac{k}{2n} \sum_i x_i^2$ et on obtient

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

2. Déterminer l'information de Fisher associée au paramètre λ. En supposant n grand, en déduire la loi approchée de l'estimateur $\hat{\lambda}$.

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

La loi asymptotique de l'estimateur du maximum de vraisemblance est la loi $\mathcal{N}(\lambda, \frac{1}{I_n(\lambda)})$. On a donc

$$\hat{\lambda} \xrightarrow{L} \mathcal{N}(\lambda, \frac{\lambda^2}{4n}).$$

3. On considère le problème de test suivant :

$$\begin{array}{llcl} H_0: & \lambda & = & \lambda_0 \\ H_1: & \lambda & = & \lambda_1 \quad (\lambda_1 > \lambda_0). \end{array}$$

3.1 En utilisant le théorème de Neyman-Pearson, montrer que le test optimal pour ce problème s'exprime en fonction de la statistique $\hat{\lambda}$. Donner la forme de la région critique de ce test.

D'après le théorème de N-P, la région critique optimale de ce test HS-HS est de la forme

$$W = \left\{ \frac{L_1}{L_0} = \left(\frac{\lambda_1}{\lambda_0}\right)^{4n} e^{-n\bar{x}(\lambda_1 - \lambda_0)} > A \right\}$$

Sachant que $\lambda_1 - \lambda_0$ est positif, le rapport de vraisemblance est une fonction décroissante de $\bar{x}$; on obtient donc une région critique de la forme $\bar{X} < B$, c'est-à-dire

$$\hat{\lambda} > C.$$

3.2 En supposant n grand, déterminer l'expression littérale de la région critique exprimée en fonction de $\hat{\lambda}$, λ_0, n et du niveau de signification α^*.

En utilisant les résultats de la question 2, on a sous H_0 $\hat{\lambda} \sim \mathcal{N}(\lambda_0, \frac{\lambda_0^2}{4n})$. La région critique est alors donnée par l'équation

$$\alpha^* = \mathbb{P}_{H_0}(\hat{\lambda} > C) = \mathbb{P}\left(\frac{\hat{\lambda} - \lambda_0}{\frac{\lambda_0}{2\sqrt{n}}} > \frac{C - \lambda_0}{\frac{\lambda_0}{2\sqrt{n}}}\right) = \Phi\left(\frac{C - \lambda_0}{\frac{\lambda_0}{2\sqrt{n}}}\right)$$

On en déduit

$$\frac{C - \lambda_0}{\frac{\lambda_0}{2\sqrt{n}}} = u_{1-\alpha^*} \quad \text{et donc} \quad C = \lambda_0\left(1 + \frac{u_{1-\alpha^*}}{2\sqrt{n}}\right).$$

Finalement

$$W = \left\{\hat{\lambda} > \lambda_0\left(1 + \frac{u_{1-\alpha^*}}{2\sqrt{n}}\right)\right\}$$

3.3 Déterminer l'expression littérale de la puissance du test en fonction de λ_0, λ_1, n et de α^*.

$$\begin{aligned}
1 - \beta &= \mathbb{P}_{H_1}(\hat{\lambda} > \lambda_0(1 + \frac{u_{1-\alpha^*}}{2\sqrt{n}})) \\
&= \mathbb{P}_{H_1}\left(\frac{\hat{\lambda} - \lambda_1}{\lambda_1/2\sqrt{n}} > \frac{\lambda_0}{\lambda_1}(2\sqrt{n} + u_{1-\alpha^*}) - 2\sqrt{n}\right) \\
&= 1 - \mathbb{P}_{H_1}\left(\frac{\hat{\lambda} - \lambda_1}{\lambda_1/2\sqrt{n}} \leq \frac{\lambda_0}{\lambda_1}(2\sqrt{n} + u_{1-\alpha^*}) - 2\sqrt{n}\right) \\
&= 1 - \Phi\left(\frac{\lambda_0}{\lambda_1}(2\sqrt{n} + u_{1-\alpha^*}) - 2\sqrt{n}\right)
\end{aligned}$$

3.4 Déterminer l'expression littérale, en fonction de λ_0, λ_1, α^* et β^*, de la taille minimale de l'échantillon permettant d'obtenir une puissance au moins égale à $1 - \beta^*$.

$$1 - \Phi\left(\frac{\lambda_0}{\lambda_1}(2\sqrt{n} + u_{1-\alpha^*}) - 2\sqrt{n}\right) \geq 1 - \beta^*$$

$$\Phi\left(2\sqrt{n}(1 - \frac{\lambda_0}{\lambda_1}) - \frac{\lambda_0}{\lambda_1}u_{1-\alpha^*}\right) \geq 1 - \beta^*$$

$$2\sqrt{n}(\frac{\lambda_1 - \lambda_0}{\lambda_1}) - \frac{\lambda_0}{\lambda_1}u_{1-\alpha^*} \geq u_{1-\beta^*}$$

$$2\sqrt{n} \geq \left(u_{1-\beta^*} + \frac{\lambda_0}{\lambda_1}u_{1-\alpha^*}\right)\frac{\lambda_1}{\lambda_1 - \lambda_0}$$

$$n \geq \left(\frac{\lambda_1 u_{1-\beta^*} + \lambda_0 u_{1-\alpha^*}}{2(\lambda_1 - \lambda_0)}\right)^2$$

4 On considère cette fois les hypothèses

$$\begin{array}{lcl} H_0: & \lambda & = & \lambda_0 \\ H_1: & \lambda & \neq & \lambda_0. \end{array}$$

Existe-t-il un test UPP pour ce problème ?

Il faut d'abord traiter le test intermédiaire

$$\begin{array}{lcll} H_0: & \lambda & = & \lambda_0 \\ H_1: & \lambda & = & \lambda_i \quad (\lambda_i \neq \lambda_0). \end{array}$$

En reprenant les calculs menés dans la question 3.1, on obtient une région critique optimale dont la forme est $\hat{\lambda} < C$ ou $\hat{\lambda} > C$ suivant que λ_i est inférieur ou supérieur à λ ; la région critique optimale dépend donc de i et le test n'est donc pas UPP.