Final SY02 Automne 2010

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. Aucun document, à l'exception du recueil de tables, n'est autorisé. Les calculettes sont autorisées à condition qu'elles ne contiennent aucune information relative au cours de sy02.

Exercice 1

On dispose d'un échantillon i.i.d $(X_1, \ldots, X_n)$ de v.a. parente X dont la fonction de densité vérifie

$$f_\lambda(x) = \lambda\, x\, \exp\left(-\frac{\lambda x^2}{2}\right)\, 1_{[0,+\infty[}(x).$$

1. Montrer qu'il existe un estimateur efficace T de $\frac{1}{\lambda}$?

Le support de X est R^+ et ne dépend pas de λ : les conditions de CR sont vérifiées.

$$L(\lambda, x_1, \ldots, x_n) = \lambda^n \prod x_i e^{-\frac{\lambda}{2}\sum_i x_i^2}$$

$$\ln L(\lambda, x_1, \ldots, x_n) = n \ln \lambda - \frac{\lambda}{2}\sum_i x_i^2 + cste$$

$$\frac{\partial \ln L}{\partial \lambda}(\lambda, x_1, \ldots, x_n) = \frac{n}{\lambda} - \frac{1}{2}\sum_i x_i^2$$

La factorisation

$$\frac{\partial \ln L}{\partial \lambda}(\lambda, X_1, \ldots, X_n) = -n(\frac{1}{2n}\sum_i X_i^2 - \frac{1}{\lambda})$$

montre que $T = \frac{\sum X_i^2}{2n}$ est un estimateur efficace de $\frac{1}{\lambda}$.

2. Donner son espérance et sa variance.

T étant un estimateur efficace de $\frac{1}{\lambda}$, est donc un estimateur sans biais de $\frac{1}{\lambda}$:

$$E(T) = \frac{1}{\lambda}.$$

Le corollaire de la CNS d'efficicacité entraîne par ailleurs que

$$var(T) = \frac{-\frac{1}{\lambda^2}}{-n} = \frac{1}{n\lambda^2}.$$

3 En déduire la borne de Fréchet $B_F\left(\frac{1}{\lambda}\right)$ et l'information de Fisher $I_n(\lambda)$.

La variance de l'estimateur efficace d'un paramètre est égale à la borne de Fréchet associée à ce même paramètre ; on a donc

$$B_F(\frac{1}{\lambda}) = \frac{1}{n\lambda^2}.$$

Sachant que pour tout paramètre θ, on a $B_F(u(\theta)) = \frac{u'(\theta)^2}{I_n(\theta)}$, on peut en déduire

$$I_n(\lambda) = \frac{(-\frac{1}{\lambda^2})^2}{\frac{1}{n\lambda^2}} = \frac{n}{\lambda^2}.$$

4. Quel est l'estimateur $\widehat{\lambda}_{MV}$ du maximum de vraisemblance de λ ?

T estimateur efficace de $\frac{1}{\lambda}$ est aussi l'unique estimateur du maximum de vraisemblance de $\frac{1}{\lambda}$.
La propriété d'invariance fonctionnelle des EMV entraîne donc que $\widehat{\lambda}_{MV} = \frac{1}{T}$ est l'unique EMV de λ.

5. En déduire une fonction asymptotiquement pivotale pour λ et un intervalle de confiance bilatéral approché sur λ au niveau de confiance $1-\alpha$.

Fonction asymptotiquement pivotale : Les propriétés asymptotiques des EMVs entraînent que

$$\widehat{\lambda}_{MV} \overset{\text{ap.}}{\sim} \mathcal{N}(\lambda, \frac{1}{I_n(\lambda)}).$$

Sachant que $I_n(\lambda) = \frac{n}{\lambda^2}$, on peut en déduire que

$$\frac{\widehat{\lambda}_{MV} - \lambda}{\frac{\lambda}{\sqrt{n}}} = \sqrt{n}\left(\frac{\widehat{\lambda}_{MV}}{\lambda} - 1\right) \overset{\text{ap.}}{\sim} \mathcal{N}(0,1)$$

Intervalle de confiance :
En utilisant la fonction as. pivotale précédente, on obtient

$$P\left(-u_{1-\alpha/2} < \sqrt{n}\left(\frac{\widehat{\lambda}_{MV}}{\lambda} - 1\right) < u_{1-\alpha/2}\right) = 1-\alpha$$

d'où l'on peut tirer

$$P\left(\frac{\widehat{\lambda}_{MV}}{1 + \frac{u_{1-\alpha/2}}{\sqrt{n}}} < \lambda < \frac{\widehat{\lambda}_{MV}}{1 - \frac{u_{1-\alpha/2}}{\sqrt{n}}}\right) = 1-\alpha.$$

Exercice 2

On dispose d'un échantillon de taille n=10 de v.a. parente X normale de moyenne 0 et de variance σ^2 inconnue. On notera dans tout ce problème T la statistique $\sum_{i=1}^{n} X_i^2$.
On veut effectuer le test suivant :

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 \quad (= 1) \\ H_1 : \sigma^2 &= \sigma_1^2 \quad (= 2). \end{aligned}$$

1. Déterminer la région critique optimale pour $\alpha^* = 0.05$.

On est en présence d'un test entre 2 hypothèses simples. Il suffit donc d'appliquer le théorème de Neyman Pearson : la forme de la région critique est fournie par la relation

$$\frac{L(x_1, \ldots, x_n, \sigma_1^2)}{L(x_1, \ldots, x_n, \sigma_0^2)} > constante.$$

On en déduit

$$\frac{\sigma_0^2}{\sigma_1^2} \exp(-\frac{1}{2}(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}) \sum x_i^2 > constante$$

et, en utilisant la propriété $\sigma_1^2 > \sigma_0^2$, on obtient

$$\sum x_i^2 > A \qquad \text{c'est-à-dire} \qquad t > A.$$

La valeur de A est alors obtenue par la relation

$$P_{H_0}(T > A) = \alpha^*.$$

Sous l'hypothèse H_0, la statistique $\frac{T}{\sigma_0^2}$ est la somme de lois normales centrées-reduites ; elle suit donc une loi de χ_N^2. On obtient

$$\alpha = P_{H_0}(\frac{T}{\sigma_0^2} > \frac{A}{\sigma_0^2}).$$

D'où

$$A = \sigma_0^2 \chi_{n,1-\alpha^*}^2$$

et la région critique est finalement

$$T > \sigma_0^2 \chi_{n,1-\alpha^*}^2.$$

A.N. $T > 18,3$

2. Calculer la puissance du test.

On a

$$1 - \beta = P(T > \sigma_0^2 \chi_{n,1-\alpha}^2 / H_1)$$

$$1 - \beta = P(\frac{T}{\sigma_1^2} > \frac{\sigma_0^2}{\sigma_1^2} \chi_{n,1-\alpha}^2 / H_1)$$

$$\chi_\beta^2 = \frac{\sigma_0^2}{\sigma_1^2} \chi_{n,1-\alpha}^2$$

A.N. $\chi_\beta^2 = \frac{1}{2} 18,3 = 9,15$ et donc $1 - \beta = 0,5$

On veut maintenant effectuer le test suivant :

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 \quad (= 1) \\ H_1 : \sigma^2 &> \sigma_0^2. \end{aligned}$$

3. Déterminer la région critique du test UPP pour $\alpha^* = 0.05$.

On commence par effectuer le test intermédiaire

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_0^2 (= 1) \\ h_1 : \sigma^2 &= \sigma_1^2 (> \sigma_0^2) \end{aligned}$$

On retrouve un test semblable à celui de la question précédente qui conduit à la région critique

$$T > \sigma_0^2 \chi^2_{n,1-\alpha^*}$$

qui ne dépend pas de σ_1^2. Le test est donc UPP avec comme région critique

$$T > \sigma_0^2 \chi^2_{n,1-\alpha^*}.$$

4. Calculer la puissance du test pour les valeurs $\sigma^2 = 1, 2, 3, 4, 5$ (approximativement) et tracer la courbe de puissance.

On a

$$\chi^2_\beta = \frac{\sigma_0^2}{\sigma_1^2} \chi^2_{n,1-\alpha}$$

D'où les valeurs

σ^2	χ^2_β	β	$1-\beta$
1	18, 3	0, 95	0, 05
2	9, 15	0, 50	0, 50
3	6, 1	0, 25	0, 75
4	4, 6	0, 10	0, 90
5	3, 7	0, 05	0, 95

Exercice 3

On désire savoir si, chez les individus qui consomment régulièrement de l'huile d'olive, le risque cardio-vasculaire est diminué. On utilise pour cela le logarithme du dosage en d-dimères. Sur un échantillon de 9 individus consommant de l'huile d'arachide, on a observé une moyenne empirique $\overline{x}$ de -0.78, avec une variance empirique corrigée s_X^{*2} de 0.0729. Sur un échantillon de 13 individus consommant de l'huile d'olive, on a observé une moyenne empirique $\overline{y}$ de -0.97, avec une variance empirique corrigée s_Y^{*2} de 0.1024. On admettra que les 2 populations suivent des distributions normales.

1. Tester l'hypothèse d'égalité des variances pour $\alpha^* = 0.05$.

Test de Fisher :

$$W = \left\{ \frac{s_X^{*2}}{s_Y^{*2}} < F_{n_X-1,n_Y-1;\frac{\alpha^*}{2}} \text{ ou } \frac{s_X^{*2}}{s_Y^{*2}} > F_{n_X-1,n_Y-1;1-\frac{\alpha^*}{2}} \right\}$$

$$F_{n_X-1,n_Y-1;\frac{\alpha^*}{2}} = F_{8,12,0.025} = \frac{1}{F_{12,8,0.975}} = \frac{1}{4.20} = 0.24$$

$$F_{n_X-1,n_Y-1;1-\frac{\alpha^*}{2}} = F_{8,12,0.975} = 3.51$$

$$\frac{s_X^{*2}}{s_Y^{*2}} = \frac{0.0729}{0.1024} = 0.71$$

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

2. Quel test proposez-vous pour décider si l'huile d'olive abaisse significativement le risque cardio-vasculaire, c'est-à-dire le dosage en d-dimères ? Quelle est votre conclusion ? On prendra $\alpha^* = 0.05$.

On peut proposer le test de Student d'égalité des moyennes avec l'alternative $H_1 : \mu_X > \mu_Y$

$$W = \left\{ \frac{\overline{x} - \overline{y}}{s^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} > t_{n_X+n_Y-2,1-\alpha^*} \right\}$$

$$s* = \frac{(n_X-1)s_X^{*2} + (n_Y-1)s_Y^{*2}}{n_X + n_Y - 2} = 0.906$$

$$t_{n_X+n_Y-2,1-\alpha^*} = t_{20;0.95} = 1.725$$

$$\frac{\overline{x} - \overline{y}}{s^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} = \frac{-0.78 - (-0.97)}{\sqrt{0.906 \times (1/9 + 1/13)}} = 1.46$$

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

Exercice 4

Le centre de transfusion sanguine de Pau a observé la répartition suivante sur 5000 donneurs.

Groupe / *Facteur*	O	A	B	AB
Rhesus+	2291	1631	282	79
Rhesus−	325	332	48	12

Peut-on considérer, au niveau de signification de 5 %, que le groupe sanguin et le rhésus sont indépendants ?

Pour tester l'indépendance des deux critères, on peut effectuer le test du χ^2 de contingence :
La région critique est $W : \{d^2 > \chi^2_{(r-1)(s-1);1-\alpha^*}\}$, où r et s sont les nombres de modalités des deux facteurs et

$$D^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n}$$

Groupe / *Facteur*	O	A	B	AB	
Rhesus+	2291(2240.9)	1631(1681.5)	282(282.7)	79(77.9)	4283
Rhesus−	325(375.1)	332(281.5)	48(47.3)	12(13.1)	717
Total	2616	1963	330	91	

D'où $d^2 = 18.5$ et $\chi^2_{3;0.95} = 7.81$.
Par conséquent, on rejette l'hypothèse H_0 d'indépendance entre les deux facteurs.

Exercice 5

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

$$9.1 \quad 7.4 \quad 17.2 \quad 10.7 \quad 15.5$$

Peut-on admettre au niveau $\alpha^* = 0.05$ que X suit une loi normale d'espérance 10 et de variance 4 ? (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	$\widehat{F}(x_i)$	$F_0(x_i) = \phi(z_i)$	$\lvert\widehat{F}(x_i) - F_0(x_i)\rvert$	$\lvert\widehat{F}(x_i^-) - F_0(x_i)\rvert$
7.4	-1.30	0.2	0.10	0.10	0.10
9.1	-0.45	0.4	0.33	0.07	0.13
10.7	0.35	0.6	0.64	0.04	0.24
15.5	2.75	0.8	1	0.20	**0.40**
17.2	3.60	1	1	0.00	0.20

On a donc $d_n^* = 0.40$.
Sachant que $d_{5,0.95} = 0.563$, on ne rejette pas l'hypothèse H_0.