CORRECTIONS JANVIER 2015

Problème 0 (1 points) :

Trouver la moyenne et la variance de la statistique

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

où $X_1, X_2, \ldots, X_n$ est un échantillon aléatoire d'une loi normale de moyenne μ et variance σ^2 (μ et σ^2 inconnues).

Comme $X_1, X_2, \ldots, X_n$ sont des variables aléatoires i.i.d. $\mathcal{N}(\mu, \sigma^2)$ on a, d'après le lemme de Fisher,

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

En utilisant les propriétés de la loi χ^2 (espérance et variance) on obtient :

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = (n-1) \Leftrightarrow \frac{(n-1)E[S^2]}{\sigma^2} = (n-1) \Rightarrow E[S^2] = \sigma^2.$$

$$Var\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1) \Leftrightarrow \frac{(n-1)^2 Var[S^2]}{\sigma^4} = 2(n-1) \Rightarrow Var[S^2] = \frac{2\sigma^4}{(n-1)}.$$

Problème 1 (1,5+1,5=3 points) :

On considère une variable aléatoire X dont la loi dépend du paramètre θ de la manière suivante :

$$\mathbb{P}(X=0) = \frac{\theta}{\theta+1} \qquad \mathbb{P}(X=1) = \frac{1}{\theta+1}.$$

Soient $X_1, X_2, \ldots, X_n$ un échantillon i.i.d. de même loi que X. Pour tout $j = 0, 1$ on désigne par n_j le nombre de X_i égaux à j. Un échantillon de taille $n = 50$ de X a donné les observations suivantes $n_0 = 20$, et $n_1 = 30$.

(a) Déterminer un estimateur de θ par la méthode des moments.

On calcule d'abord le moment théorique

$$E[X] = 0 \times \frac{\theta}{\theta+1} + 1 \times \frac{1}{\theta+1} = \frac{1}{\theta+1}$$

et ensuite le moment de l'échantillon

$$\bar{X}_n = \frac{20 \times 0 + 30 \times 1}{50} = \frac{3}{5}.$$

On déduit facilement $\hat{\theta}_{MM} = \frac{2}{3}$.

(b) Déterminer un estimateur de θ par la méthode du maximum de vraisemblance.

La fonction de vraisemblance est :

$$L_n(\theta) = \prod_{i=1}^{50}\mathbb{P}(X = X_i) = \prod_{i=1}^{20}\mathbb{P}(X=0)\prod_{i=1}^{30}\mathbb{P}(X=1)$$
$$= \left(\frac{\theta}{\theta+1}\right)^{20}\left(\frac{1}{\theta+1}\right)^{30} = \frac{\theta^{20}}{(1+\theta)^{50}}$$

1. E. Ursu & A. Lourme, Faculté d'économie, gestion & AES, Université de Bordeaux

La log-vraisemblance est donnée par

$$l_n(\theta) = \ln L_n(\theta) = 20 \ln \theta - 50 \ln(1+\theta),$$

d'où

$$\frac{\partial}{\partial\theta} l_n(\theta) = \frac{20}{\theta} - \frac{50}{1+\theta} = \frac{20-30\theta}{\theta(1+\theta)} = 0$$

avec la solution $\hat{\theta}_{MV} = \frac{2}{3}$.

Problème 2 (2+2+2=6 points) :

La masse en mg d'une fourmi *Crematogaster* est une variable aléatoire normale de moyenne μ et de variance σ^2. On prélève trente et une fourmis de cette espèce au hasard : onze sont des mâles, la masse moyenne l'échantillon est :

$$\bar{X} = \frac{1}{31}\sum_{i=1}^{31} X_i = 4,9 \text{ et la variance des masses est : } S^2 = \frac{1}{30}\sum_{i=1}^{31}(X_i - \bar{X})^2 = 1,1$$

(X_i est la masse de la fourmi i).

$n = 31$ est la taille de l'échantillon, $n_1 = 11$ le nombre de mâles et $n_2 = 20$ le nombre de femelles.

(a) i. Déterminez un intervalle de confiance à 90% de μ lorsque :

A. $\sigma^2 = 1$

$IC_{0,9}(\mu) = \bar{X} \pm z_{0,95} \times \sigma/\sqrt{n} = 4,9 \pm 1,65 \times 1/\sqrt{31} \approx [4,605; 5,195]$.

B. σ^2 est inconnu.

$IC_{0,9}(\mu) = \bar{X} \pm t_{n-1;0,95} \times S/\sqrt{n} = 4,9 \pm 1,70 \times \sqrt{1,1}/\sqrt{31} \approx [4,580; 5,220]$.

ii. Se peut-il dans le cas i. ou ii. que $\mu = 5,2$?

Au seuil de 10%, on rejette $\mathscr{H}_0 : \mu = 5,2$ au profit de $\mathscr{H}_1 : \mu \neq 5,2$ lorsque $\sigma = 1$ puisque $5,2 \notin [4,605; 5,195]$; lorsque σ est inconnu, on ne rejette pas $\mathscr{H}_0$ puisque $5,2 \in [4,580; 5,220]$.

(b) i. Déterminez un intervalle de confiance à 90% de σ^2.

$IC_{0,9}(\sigma^2) = \left[(n-1)S^2/\chi^2_{n-1;0,95}; (n-1)S^2/\chi^2_{n-1;0,05}\right] = [30 \times 1,1/43,77; 30 \times 1,1/18,49] \approx [0,75; 1,78]$.

ii. Se peut-il que $\sigma^2 = 1$?

Au seuil de 10%, on ne rejette pas $\mathscr{H}_0 : \sigma^2 = 1$ au profit de $\mathscr{H}_1 : \sigma^2 \neq 1$ puisque $1 \in [0,75; 1,78]$.

(c) On cherche à tester la proportion p de mâles dans l'espèce.

i. Déterminez un intervalle de confiance $1-\alpha$ $(0 < \alpha < 1)$ de p.

$IC_{1-\alpha}(p) = \hat{p} \pm z_{1-\alpha/2} \times \sqrt{\hat{p}(1-\hat{p})/n} = 11/31 \pm z_{1-\alpha/2} \times \sqrt{(11/31) \times (1 - 11/31)/31}$.

ii. Quelle valeur maximale de α permet d'accepter l'hypothèse : il y a autant de mâles que de femelles dans l'espèce.

On cherche α de sorte que : $0,5 \in IC_{1-\alpha}(p)$ ou encore : $0,5 \leq 11/31 + z_{1-\alpha/2} \times \sqrt{(11/31) \times (1-11/31)/31}$; d'où : $\alpha \leq 0,091$. Le seuil doit être de 9,1% au plus pour que l'on puisse accepter l'hypothèse : $p = 0,5$.

Problème 3 (1+1+2=4 points) :

Deux fabricants de piles pour des ordinateurs portables prétendent que leurs piles peuvent faire fonctionner un ordinateur pendant au moins trois heures avant de devoir être rechargées. Des piles de ces deux fabricants ont été testées avec dix ordinateurs différents. Les observations sont les suivantes :

Ordinateur	1	2	3	4	5	6	7	8	9	10
T_1	2,9	2,8	2,9	3,2	3,0	3,1	2,7	2,9	2,7	2,9
T_2	3,1	3,2	3,3	3,0	2,9	2,9	3,1	3,2	2,8	3,2

où T_i est la durée de fonctionnement obtenue avec une pile de marque i, pour $i = 1, 2$. On suppose que T_i suit une loi gaussienne.

On trouve :

$$\begin{array}{rcl} \bar{X}_{T_1} = 2,91 & ; & \bar{X}_{T_2} = 3,07 \\ S_1^2 = 0,025444 & ; & S_2^2 = 0,026777 \\ S_1 = 0,15951 & ; & S_2 = 0,16363 \end{array}$$

(a) On croit que l'écart-type de la variable aléatoire T_1 est supérieur à $0,2$. Vérifier cette hypothèse au seuil $\alpha = 0,05$.

L'hypothèse statistique à tester est :

$$\mathscr{H}_0 : \sigma^2 = (0,2)^2 \quad \text{versus} \quad \mathscr{H}_1 : \sigma^2 > (0,2)^2.$$

La statistique qui convient pour ce test est :

$$W_0 = \frac{(n-1)S_1^2}{\sigma_0^2}.$$

Lorsque H_0 est vraie, W_0 suit une loi du khi-deux à $(n-1)$ degrés de liberté. Puisque $n = 10$ et $\alpha = 0,05$, la valeur critique est $\chi^2_{0,95;9} = 16,919$. On adopte la règle de décision suivante : rejeter $\mathscr{H}_0$ si

$$W_0 > 16,919.$$

Puisque $S_1^2 = 0,02544$, $\sigma_0^2 = 0,04$ on obtient

$$W_0 = \frac{9 \times 0,02544}{0,04} = 5,725.$$

Puisque $W_0 = 5,725 < 16,919$ on ne peut pas rejeter $\mathscr{H}_0$.

(b) On veut ensuite tester l'hypothèse qu'il n'y a pas de différence significative entre la moyenne de T_1 et celle de T_2 (seuil de signification $\alpha = 0,05$).

Nous sommes en présence d'**échantillons dépendants ou appariés**.

Les calculs préliminaires conduisent à

$$\bar{D} = \bar{X}_{T_1} - \bar{X}_{T_2} = -0,16, \quad S_d^2 = 0,06044.$$

L'hypothèse statistique à tester est :

$$\mathscr{H}_0 : \mu_d = 0 \quad \text{versus} \quad \mathscr{H}_1 : \mu_d \neq 0.$$

La statistique qui convient pour ce test est :

$$T_0 = \frac{\bar{D}}{S_d/\sqrt{n}}.$$

Sous $\mathscr{H}_0$ on sait que T_0 suit une loi de Student à $n-1$ degrés de liberté.
Au seuil $\alpha = 0,05$ avec 9 degrés de liberté la valeur critique est $t_{0,975;9} = 2,2621$. La règle de décision est donc la suivante : rejeter $\mathscr{H}_0$ si $|T_0| > 2,2621$. On obtient

$$T_0 = \frac{-0,16}{\sqrt{0,06044/10}} = -2,0579.$$

Puisque $|T_0| = 2,0579 < 2,2621$ on ne peut pas rejeter $\mathscr{H}_0$.

(c) Sur la base d'un échantillon de taille $n = 23$, on trouve que l'écart-type de l'échantillon aléatoire T_2 est (environ) égal à $0,1636$. On veut tester les hypothèses suivantes :

$$\mathscr{H}_0 : \mu_{T_2} = 3 \quad \text{versus} \quad \mathscr{H}_1 : \mu_{T_2} > 3.$$

On a $n = 23$ observations avec une variance inconnue. L'hypothèse statistique à tester est :

$$\mathscr{H}_0 : \mu_{T_2} = 3 \quad \text{versus} \quad \mathscr{H}_1 : \mu_{T_2} > 3.$$

On rejette $\mathscr{H}_0$ si

$$T_0 = \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} > t_{1-\alpha;n-1}.$$

On a $\mu_0 = 3$, $n = 23$, $t_{1-\alpha;n-1} = 1,7171$, $S = 0,1636$.

i. Trouver la région de rejection pour ce test (seuil $\alpha = 0,05$).
La region de rejection est obtenue à partir de

$$\frac{\bar{X}_n - 3}{0,1636/\sqrt{23}} > 1,7171,$$

d'où on trouve $\bar{X}_n > 3,058577$.

ii. Quelle est la probabilité de commettre une erreur de deuxième espèce selon l'hypothèse alternative $\mathscr{H}_1 : \mu_{T_2} = 3,1$ que l'on suppose vraie (seuil de signification $\alpha = 0,05$) ?
On veut calculer l'erreur de type II

$$\begin{aligned}
\beta &= \mathbb{P}(\text{ne pas rejeter } H_0 | H_0 \text{ est fausse }) = \mathbb{P}(\bar{X}_n \leq 3,058577 | \mu = 3,1) \\
&= \mathbb{P}\left(\frac{\bar{X}_n - 3,1}{S/\sqrt{n}} \leq \frac{3,058577 - 3,1}{0,1636/\sqrt{23}}\right) = \mathbb{P}(T \leq -1,214),
\end{aligned}$$

où $T \sim$ Student$_{22}$. On trouve $0,10 \leq \beta \leq 0,15$.

Problème 4 (2+2=4 points) :

Les salaires mensuels (en k€) des français sont distribués selon $\mathcal{N}(\mu_1, \sigma^2 = 1)$ pour les hommes et $\mathcal{N}(\mu_2, \sigma^2 = 1)$ pour les femmes.

$\bar{X}_1$ désignant le salaire mensuel moyen d'un échantillon aléatoire de n_1 français et $\bar{X}_2$ celui de n_2 françaises, on teste l'hypothèse

$$H_0 : \mu_1 = \mu_2 \quad \text{contre} \quad H_1 : \mu_1 \neq \mu_2$$

grâce à la statistique $Z = (\bar{X}_1 - \bar{X}_2)/(\sigma \times \sqrt{1/n_1 + 1/n_2})$.

(a) i. Montrez que sous H_0 : $Z \sim \mathcal{N}(0,1)$.

Si l'on admet que les salaires des hommes de l'échantillon sont indépendants, $\bar{X}_1$ est distribuée selon $\mathscr{N}(\mu_1, 1/n_1)$; sous la même hypothèse pour les femmes : $\bar{X}_2 \sim \mathscr{N}(\mu_2, 1/n_2)$. Comme $\bar{X}_1$ et $\bar{X}_2$ sont indépendantes, $\Delta = \bar{X}_1 - \bar{X}_2$ est distribuée selon une loi normale de moyenne $\mu_1 - \mu_2$ et de variance $1/n_1 + 1/n_2$. Ainsi, $Z = \Delta/\sqrt{1/n_1 + 1/n_2}$ est normale, réduite et sa moyenne $(\mu_1 - \mu_2)/\sqrt{1/n_1 + 1/n_2}$ est nulle sous $\mathscr{H}_0$.

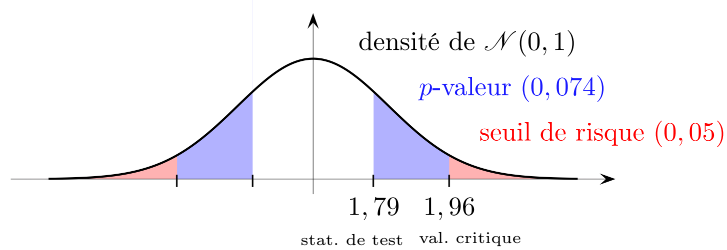
ii. Quelle est la zone de rejet de H_0 au seuil de 5% ?

Au seuil de 5%, on rejette $\mathscr{H}_0$ si : $|Z| > z_{0,975} = 1,96$.

Dans un groupe de vingt personnes dont quatre hommes on observe : $\bar{X}_1 = 3$, $\bar{X}_2 = 2$.

2. i. Quelle est la p-valeur du test ?

La valeur prise par la statistique de test est : $Z = (3-2)/\sqrt{1/4 + 1/16} = 8/\sqrt{20} \approx 1,79$. Comme le test est bilatéral sa p-valeur est $p = \mathbb{P}(|Z| > 1,79|\mathscr{H}_0) = 2 \times \mathbb{P}(\mathscr{N}(0,1) > 1,79) = 2 \times (1 - \phi(1,79)) \approx 0,074$.



ii. Doit on rejeter $\mathscr{H}_0$ au seuil de 5% ?

On ne rejette pas $\mathscr{H}_0$ au seuil de 5% puisque $p > 0,05$.

Problème 5 (3 points) :

Un sociologue s'intéresse à la relation entre la couleur de la peau et la mobilité professionnelle. Il prélève un échantillon de 94 personnes de peau pâle, 175 personnes de peau moyenne, et 80 personnes de peau brune. Il construit une mesure de mobilité à l'aide de quoi il classifie ses sujets selon leur mobilité. Voici les résultats.

Mobilité \ Couleur de la peau	pale	moyenne	brune	total
grande	35	84	51	170
faible	59	91	29	179
total	94	175	80	$n = 349$

Tester l'hypothèse qu'il n'y a aucune relation entre la couleur de la peau et la mobilité professionnelle avec $\alpha = 0,05$.

C'est un test d'indépendance.
Hypothèse $(\mathscr{H}_0)$: les deux caractères sont indépendants.
Hypothèse $(\mathscr{H}_1)$: les deux caractères ne sont pas indépendants.
Les effectifs théoriques sont donnés par le tableau suivant :

Mobilité \ Couleur de la peau	pale	moyenne	brune	total
grande	45,79	85,24	38,97	170
faible	48,21	89,76	41,03	179
total	94	175	80	$n = 349$

Tous les effectifs théoriques sont supérieurs à 5. La variable

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{ij}-T_{ij})^2}{T_{ij}}$$

suit une loi $\chi^2_{(3-1)(2-1)}$.
Puisque $X^2 = 12,23$ est supérieur à la valeur $c = \chi^2_{0,95;2} = 5,9915$ on rejete $\mathscr{H}_0$ au seuil de signification de $\alpha = 0,05$.