Master *"Mathématiques Appliquées, Statistique"* -M1, Université Claude Bernard, Lyon 1
Statistique Paramétrique année 2016-2017

Examen du 18 janvier 2017,
Deux feuilles A4 avec les formules, tables des lois et des fractiles admises,
autres documents interdits.
Appareils connectables interdits. Calculatrice autorisée
Durée 3h. Le sujet est sur 3 pages (recto-verso)

Exercice 1. *(11 points)*
Soit X une variable aléatoire continue et $X_1, ..., X_n$ le n-échantillon associé. La fonction de densité de X est

$$f_\theta(x) = \frac{1+\theta}{1-\theta} \cdot x^{\frac{2\theta}{1-\theta}} \cdot 1\!\!1_{0<x<1}$$

avec le paramètre $\theta \in]0, 1[$.
1) Calculez l'espérance $E[X]$ de la variable aléatoire X. Calculez l'espérance $E[X^2]$ de la variable aléatoire X^2. *(0.5 points)*
2) Trouvez un estimateur par la méthode des moments de θ. Notons cet estimateur par $\hat{\theta}_n^{(1)}$. *(0.5 points)*
3) Etudiez la convergence, le biais pour $\hat{\theta}_n^{(1)}$. Calculez $Var[\hat{\theta}_n^{(1)}]$. *(1 point)*
4) Si on considère $\hat{\theta}_n^{(1)}$ comme estimateur ponctuel pour θ, trouvez un estimateur par intervalle *asymptotique*, de niveau de confiance $1-\alpha$, pour le paramètre θ. *(1.5 points)*
5) Quelles sont les valeurs de la variable aléatoire X, avec une probabilité égale à 1? *(0.5 points)*
6) Trouvez l'estimateur du maximum de vraisemblance de θ. Ne vérifiez pas que c'est un point de maximum pour la vraisemblance. *(1.5 point)*
7) Soit la variable aléatoire $Y = \log X$. Trouvez la fonction de répartition et la densité de la variable aléatoire Y. *(1 point)*
8) Considérons le n-échantillon $Y_1, \cdots, Y_n$ pour la variable aléatoire Y, c'est-à-dire, $Y_i = \log X_i$, pour $i = 1, \cdots, n$. Pour le paramètre θ, considérons l'estimateur suivant : $\hat{\theta}_n^{(2)} = (1 + \bar{Y}_n)/(1 - \bar{Y}_n)$. Etudiez la convergence de cet estimateur. *(1 point)*
9) Si $\theta \in]0, 1[$, quelles sont les valeurs possibles pour la fonction h définie par $h(\theta) = \frac{\theta-1}{1+\theta}$? *(0.5 points)*
10) En utilisant les questions 6) et 9) trouvez l'estimateur du maximum de vraisemblance pour le paramètre ψ défini par $\psi = h(\theta)$. On note cet estimateur par $\hat{\psi}_n$. *(1 point)*
11) Montrez que les densité $f_\theta(x)$ de X est de type exponentiel. Etudiez le biais, la convergence, et l'efficacité de l'estimateur $\hat{\psi}_n$ de ψ. *(1.25 point)*
12) Montrez que la densité suivante $g_\psi(x) = -\frac{1}{\psi} e^{-\frac{x}{\psi}} 1\!\!1_{x<0}$ est aussi de type exponentiel. Déduisez l'exhaustivité de l'estimateur $\hat{\psi}_n$ de ψ. *(0.75 points)*

Exercice 2. *(3 points)*
On considère un n-échantillon $X_i \sim \mathcal{B}(p)$ de loi Bernoulli, $i = 1, \cdots, n$, avec $0 < p < 1$.
1) Soit la fonction h :$]0,1[\to]0,\infty[$ définie par $h(p) = \frac{p}{1-p}$. Etudiez la monotonie de h. *(0.5 points)*
2) Si $p_1 \leq p_2$ comparez $\frac{h(p_2)}{h(p_1)}$ avec 1? *(0.5 points)*
3) On veut tester l'hypothèse nulle $H_0 : p \geq p_0$ contre l'hypothèse alternative $H_1 : p < p_0$, avec p_0 une valeur connue appartenant à l'intervalle $]0,1[$.
Est-ce que le test le plus puissant existe? Justification. (vous utiliserez la question 2)). *(1 point)*
4) En utilisant la question 3), trouvez la statistique de test *asymptotique* de risque $\alpha \in]0,1[$ pour tester H_0 contre H_1. *(1 point)*

Exercice 3. *(1.5 points)*
On mesure la longueur (en mm) des oeufs d'une certaine espèce d'oiseaux:

20, 22, 23, 19, 15, 20, 22, 21, 19

On suppose que la longueur des oeufs suit une loi Normale. Peut-on dire, avec un niveau de confiance de 0.95 que la longueur moyenne d'un oeuf est de 20 mm?
On suppose connu que:

$$(20+22+23+19+15+20+22+21+19)/9 = 20.1, \qquad (20-20.1)^2+(22-20.1)^2+...+(19-20.1)^2 = 44.89$$

Note: Ecrire la variable aléatoire, sa loi, les hypothèses à tester, la statistique de test, sa loi, la zone de rejet et la conclusion.

Exercice 4. *(4.5 points)*
Les données proviennent du site "Center for Machine Learning and Intelligent Systems at the University of California":

http://archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test

de l'Université de Californie, Etats Unis.
Il s'agit de l'étude sur l'écoulement du béton qui n'est pas déterminé seulement par la teneur en eau, mais il est également influencé par d'autres ingrédients du béton.
On modélise ainsi la variable *FLOW* (l'écoulement du ciment) fonction des variables:
Slag: scories
Fly: Cendres volantes
Water: eau
Coarse_A: agrégat à gros grain
Fine_A: agrégat à fin grain

Note: Les tests sont à faire pour un seuil $\alpha = 0.05$. Pour chaque test, écrire les hypothèses à tester, les modèles correspondants, les statistiques de test et leur loi, valeur de la statistique de test, la p-value.

Vous trouvez plus en bas le code R et les sorties associées qui vous permettront de répondre aux questions suivantes:

1) Sur combien d'observations on a réalisé le modèle statistique correspondant à la fonction "lm"? *(0.5 points)*

2) Quel type de modèle a-t-on utilisé ? Ecrivez le modèle statistique correspondant à la fonction R "*lm*". *(0.75 points)*

3) Testez si le modèle écrit à la question 2) est significatif. *(0.75 points)*

4) Testez chaque variable qui intervient dans le modèle. Vous donnez les détails pour une seule variable, pour les autres vous donnez seulement la conclusion. *(1 point)*

4) Donnez les estimations des paramètres du modèle. Interprétez ces estimations. *(1 point)*

5) Quelle est la qualité globale d'ajustement de ce modèle? Interprétation. *(0.5 points)*

```
###################################  CODE  R ################################"
cem=read.table("exam2016.txt",col.names=c("no","Cement", "Slag","Fly","Water","SP","Coarse_A",
"Fine_A","SLUMP","FLOW","MPA"))
attach(cem)
m1=lm(FLOW~Slag+Fly+Water+Coarse_A+Fine_A)
summary(m1)

############################################SORTIES du code R ###########################"

Call:
lm(formula = FLOW ~ Slag + Fly + Water + Coarse_A + Fine_A)

Residuals:
    Min      1Q  Median      3Q     Max
-31.663 -10.127   1.692   9.604  23.470

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -86.168185  59.292365  -1.453   0.1494
Slag         -0.078227   0.028592  -2.736   0.0074 **
Fly           0.007146   0.018857   0.379   0.7055
Water         0.566227   0.095527   5.927 4.71e-08 ***
Coarse_A      0.009760   0.026116   0.374   0.7094
Fine_A        0.027777   0.030274   0.918   0.3611
---
Signif. codes:  0 □***□ 0.001 □**□ 0.01 □*□ 0.05 □.□ 0.1 □ □ 1

Residual standard error: 12.73 on 97 degrees of freedom
Multiple R-squared:  0.5009,    Adjusted R-squared:  0.4751
F-statistic: 19.47 on 5 and 97 DF,  p-value: 2.225e-13
```