Master Ingénierie Mathématique -M1, Université Claude Bernard, Lyon1
Statistique Paramétrique année 2014-2015

Examen du 14 Janvier 2015

Deux feuilles A4 avec les formules, tables des lois et des fractiles admises, autres documents interdits.
Téléphones portables interdits. Calculatrice autorisée
Durée 3h. Le sujet est sur 4 pages (recto-verso)

Remarque: pour les Exercices 1 et 2, si ce n'est pas demandé expressément, vous pouvez utiliser comme connus les résultats obtenus(prouvés) en cours et TD.

Exercice 1. *(11 points)*

Soit une variable aléatoire continue X, de densité:

$$f_\theta(x) = \left(\frac{1}{\theta} - 1\right) e^{-\left(\frac{1}{\theta}-1\right)x} 1\!\!1_{x>0}$$

avec $\theta \in]0,1[$ un paramètre, inconnu. On considère un n-échantillon $(X_1, ..., X_n)$ pour cette loi (variable).

Partie I.

1) Calculez l'espérance et la variance de la variable aléatoire X. *(1 point)*
2) Trouvez un estimateur pour le paramètre θ par la méthode des moments. Etudiez la convergence de cet estimateur. *(1 point)*
3) La loi est-elle de type exponentiel? Donnez une statistique exhaustive pour θ. *(1 point)*

Partie II.

4) Calculez la fonction de répartition, notée par F, de la variable aléatoire X. Calculez $\mathbb{P}[X \geq 1]$, probabilité qu'on va noter avec p. *(1 point)*
5) Donnez la loi de la variable aléatoire $Y = 1\!\!1_{X\geq 1}$. *(1 point)*

Sur la base du n-échantillon $(X_1, ..., X_n)$ on a le n-échantillon $(Y_1, ..., Y_n)$, avec $Y_i = 1\!\!1_{X_i\geq 1}$, pour $i = 1, \cdots, n$. Donc, si on connaît la variable aléatoire X_i on connaît aussi la variable aléatoire Y_i, pour tout $i = 1, \cdots, n$.

6) On note $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Calculez l'espérance et la variance de $\bar{Y}_n$ (les écrire fonction du paramètre p). *(0.5 points)*
7) Sur la base du n-échantillon $(Y_1, ..., Y_n)$, trouvez (justifiez, sans faire les calculs) l'estimateur du maximum de vraisemblance du paramètre p. On note cet estimateur avec $\hat{p}_n$. *(0.5 points)*
8) La loi de la variable aléatoire Y est-elle de type exponentiel par rapport au paramètre p? Est-ce que l'estimateur $\hat{p}_n$ de p est un estimateur convergent, sans biais, efficace, exhaustif? *(1.5 points)*
9) En utilisant les questions **7)** et **8)**, trouvez (donnez la définition de l'estimateur par intervalle et ensuite donnez les détails de calcul) l'estimateur par intervalle asymptotique pour le paramètre p, de niveau de confiance $1 - \alpha$, avec $\alpha \in]0,1[$ connu. *(1.5 points)*
10) En utilisant les questions **6)** et **8)**, montrez que $\hat{\theta}_n = \dfrac{1}{1 - \log(\bar{Y}_n)}$ est un estimateur fortement

consistant pour θ. *(0.5 points)*

Partie III.
Pour la variable aléatoire $Y = \mathbb{1}_{X\geq 1}$ de la Partie II., et le n-échantillon correspondant $Y_i = \mathbb{1}_{X_i\geq 1}$, $i = 1, \cdots, n$, nous considérons un nouveau paramètre $\mu = p^2$. On rappelle que $p = \mathbb{P}[X \geq 1]$.
11) Trouvez l'estimateur du maximum de vraisemblance du paramètre μ. On note cet estimateur avec $\hat{\mu}_n$. *(0.75 points)*
12) L'estimateur $\hat{\mu}_n$ est-il convergent, sans biais? *(0.75 points)*

Exercice 2. *(2.5 points)*
Soit une variable aléatoire continue X, de densité:

$$f_\theta(x) = \left(\frac{1}{\theta} - 1\right) e^{-\left(\frac{1}{\theta}-1\right)x}\mathbb{1}_{x>0}$$

avec $\theta \in]0,1[$ un paramètre, inconnu. On considère un n-échantillon $(X_1, ..., X_n)$ pour cette loi (variable).
1) On rappelle la définition des lois de type gamma. La fonction de densité d'une loi gamma $\gamma(s, \lambda)$ est donnée par

$$g(x) = \frac{\lambda^s}{\Gamma(s)} e^{-\lambda x} x^{s-1}\mathbb{1}_{x\geq 0}, \qquad \lambda, s > 0$$

avec

$$\Gamma(s) = \int_0^{+\infty} x^{s-1} e^{-x} dx$$

On rappelle que la somme de deux lois gamma indépendantes $\gamma(s, \lambda)$ et $\gamma(t, \lambda)$, est une loi gamma $\gamma(s + t, \lambda)$ et que $\Gamma(1) = 1$.
La loi de la variable aléatoire X est-elle de type gamma? Si oui, laquelle? Quelle est la loi de $\displaystyle\sum_{i=1}^{n} X_i$? *(0.5 points)*
2) On suppose n fixé, donc connu. En utilisant la question **1)** de cet exercice, pour un seuil $\alpha \in]0,1[$ fixé, trouvez le test uniformément le plus puissant pour tester l'hypothèse $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$, pour un $\theta_0 > 0$ connu. *(2 points)*

Note: Pour les tests d'hypothèse de l'exercice suivant, écrire la variable aléatoire, sa loi, les hypothèses à tester, la statistique de test et sa loi, la zone de rejet et la conclusion.
Exercice 3. *(2.5 points)*
On veut étudier la population de poissons d'un étang. Pour ceci, un échantillon (jugé représentatif) de 1600 poissons est pêché. Sur les 1600 poissons pêchés, 211 sont des Carpes.
1) Peut-on dire que, avec un risque de 0.05, les Carpes représentent une cinquième de la population de poissons de l'étang? *(1.25 points)*
2) On analyse maintenant le poids des 211 Carpes pêchées. On suppose que le poids d'une Carpe suit une loi Normale. Nous notons par x_1 le poids de la première Carpe, x_2 le poids de la deuxième Carpe, ..., x_{211} le poids de la deux-cent-onzième Carpe. Sachant que $\displaystyle\frac{1}{211}\sum_{i=1}^{211} x_i = 950$g et $\displaystyle\frac{1}{210}\sum_{i=1}^{211}(x_i - 950)^2 = 1600$, testez l'affirmation que (risque 0.05) le poids moyen d'une Carpe de l'étang est d'au moins 1kg. *(1.25 points)*

Exercice 4. *(4 points)*
Les données traitées dans cet exercice se trouvent sur le site de l'Université de Californie, Etats Unis: *http://archive.ics.uci.edu/ml/datasets/Yacht+Hydrodynamics*. Le but est d'étudier la résistance au déplacement des yachts fonction des dimensions de la coque et de la vitesse du bateau.
Les variables lues dans le fichier *"exam_2014.txt"* sont, dans l'ordre:
- *LP*: Position longitudinale du centre de flottabilité;
- *PC*: Coefficient prismatique;
- *LDR*: Rapport longueur/déplacement;
- *BDR*: Rapport poutre/tirant d'eau;
- *LBR*: Rapport longueur/poutre;
- *NRFROUDE*: Nombre de Froude (de l'eau);
- *RESIST*: résistance au déplacement du bateau;

Dans l'étude présentée ici, on étudie la variable *LRESIST* fonction de *LP, PC, LDR, BDR, LBR, NRFROUDE*. On a considéré la variable *LRESIST=log(RESIST)* pour avoir une variable expliquée de loi Normale.

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 leurs loi.*
Vous trouvez ci-joint (sur la page suivante) le code R et les sorties associées qui vous permettront de répondre aux questions suivantes:
1) Sur combien d'observations l'étude a-t-elle été réalisée? Justification. *(0.25 points)*
2) Quel type de modèle linéaire a-t-on utilisé? Ecrivez le modèle statistique correspondant à la fonction R *"lm"*. *(1 point)*
3) Testez si le modèle écrit à la question 2) est significatif. *(1 point)*
4) Donnez les estimations des paramètres du modèle. *(0.5 points)*
5) Testez chaque variable qui intervient dans le modèle. Vous donnez les détails pour une seule variable, pour les autres vous donnez que la conclusion. *(1 point)*
6) Quelle est la qualité globale d'ajustement de ce modèle? Interprétation. *(0.25 points)*

code R et sorties

code R

```
sink('exam1_2014.sink')
yacht=read.table("exam_2014.txt",col.names=c("LP","PC","LDR","BDR","LBR","NRFROU
DE","RESIST"))
attach(yacht)
LRESIST=log(RESIST)
m1=lm(LRESIST~LP+PC+LDR+BDR+NRFROUDE)
print(summary(m1))
sink()
```

exam1_2014.sink

```
Call:
lm(formula = LRESIST ~ LP + PC + LDR + BDR + NRFROUDE)

Residuals:
     Min       1Q   Median       3Q      Max
-2.72371 -0.10910  0.01487  0.18628  0.54714

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.484558   0.599483  -5.813 1.56e-08 ***
LP           0.021274   0.011897   1.788  0.07475 .
PC          -1.836268   0.838689  -2.189  0.02933 *
LDR          0.007665   0.078380   0.098  0.92216
BDR          0.110513   0.038425   2.876  0.00431 **
NRFROUDE    18.030816   0.178329 101.110  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3154 on 302 degrees of freedom
Multiple R-squared:  0.9713,    Adjusted R-squared:  0.9709
F-statistic:  2048 on 5 and 302 DF,  p-value: < 2.2e-16
```