Master Ingénierie Mathématique -M1, Université Claude Bernard, Lyon1
Statistique Paramétrique année 2015-2016

Examen du 13 Janvier 2016
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. *(10.5 points)*
Soit une variable aléatoire continue X, de densité:

$$f_\theta(x) = \begin{cases} b - \frac{\theta}{2} & \text{si } -2 < x \leq 0 \\ \\ b + \frac{\theta}{2} & \text{si } 0 < x < 2 \end{cases}$$

avec b une constante à déterminer plus loin. Le paramètre θ est inconnu et $\theta \in]-\frac{1}{2}, \frac{1}{2}[$.
On considère un n-échantillon $(X_1, ..., X_n)$ pour cette loi (variable).

1) Calculer la valeur de b. *(0.5 points)*
2) Calculez l'espérance de la variable aléatoire X et ensuite déduisez un estimateur pour θ par la méthode des moments. Notons cet estimateur par $\hat{\theta}_n^{(1)}$. *(0.75 points)*
3) Calculez la variance de la variable aléatoire X. Est-elle > 0 pour tout $\theta \in]-\frac{1}{2}, \frac{1}{2}[$? Déduisez $Var[\hat{\theta}_n^{(1)}]$. Etudiez également la convergence et le biais de l'estimateur $\hat{\theta}_n^{(1)}$. *(1.75 points)*
4) Calculez la probabilité $I\!P[0 < X < 2]$. *(0.5 points)*
5) Soit la variable aléatoire $Y = 1\!\!1_{0<X<2}$ et le n-échantillon correspondant $Y_i = 1\!\!1_{0<X_i<2}$ pour $i = 1, \cdots, n$. Quelle est la loi de Y? *(0.5 points)*
6) Calculez la probabilité $I\!P[-2 < X < 2]$ et déduisez quelles sont les valeurs possibles pour la variable aléatoire X. Déduisez la relation entre les variables aléatoires $1\!\!1_{-2<X\leq 0}$ et $1\!\!1_{0<X<2}$. *(1 point)*
7) En utilisant le fait que la densité de la variable aléatoire X peut s'écrire:

$$f_\theta(x) = \left(\frac{1}{4} - \frac{\theta}{2}\right)^{1\!\!1_{-2<x\leq 0}} \cdot \left(\frac{1}{4} + \frac{\theta}{2}\right)^{1\!\!1_{0<x<2}}$$

et la question précédente, écrivez la vraisemblance du n-échantillon $X_1, \cdots, X_n$. Ecrivez cette vraisemblance fonction de $S_n = \sum_{i=1}^n 1\!\!1_{0<X_i<2}$. Trouvez l'estimateur du maximum de vraisemblance de θ. *(1.5 points)*
8) Considérons l'estimateur suivant pour θ: $\hat{\theta}_n^{(2)} = \frac{S_n}{n} - \frac{1}{2}$. Etudiez la convergence, le biais et l'efficacité de $\hat{\theta}_n^{(2)}$. *(2 points)*
9) Tenant compte des questions **3)** et **8)**, entre les estimateurs $\hat{\theta}_n^{(1)}$ et $\hat{\theta}_n^{(2)}$ lequel il faut choisir? Justification. *(0.5 points)*
10) Ecrivez le Théorème Central Limite pour la suite de variables aléatoires $Y_1, \cdots, Y_n$. En utilisant $\hat{\theta}_n^{(2)}$ comme estimateur ponctuel pour θ et les résultats de la question précédente, trouvez un estimateur par intervalle asymptotique de niveau de confiance $1 - \alpha$ pour le paramètre θ. *(1.5 points)*

Exercice 2. *(3 points)*
Soit un n-échantillon $(X_1, ..., X_n)$ avec $X_i \sim \mathcal{N}(m, 2^2)$ de loi Normale d'espérance m (inconnue) et de variance 2^2. On veut tester l'hypothèse nulle $H_0 : m \geq m_0$ contre l'hypothèse alternative $H_1 : m < m_0$, avec m_0 une valeur connue.
Est-ce que le test le plus puissant existe? Justification.
Si oui, trouvez la statistique de test de risque $\alpha \in]0, 1[$.

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. *(1.5 points)*
Une entreprise voudrait remplacer une machine qui fabrique des ampoules avec une autre machine plus performante.
La machine utilisée actuellement produit 3% d'ampoules défectueuses. L'entreprise fait un essai de la machine qui pourrait l'intéresser. Avec cette machine, sur 400 ampoules, 8 ampoules sont défectueuses. Avec un risque de 0.05, est-ce que cette nouvelle machine est de manière significative plus performante que l'actuelle?

Exercice 4. *(5 points)*
Les données traitées dans cet exercice proviennent du package R "car", tableau de données "Wool". On veut modéliser le nombres de cycles jusqu'à la rupture d'un fil de laine. Les variables mésurées sont, dans l'ordre:
- *len*: la longueur du fil de laine étudié (en mm);
- *amp*: l'amplitude du cycle de chargement (en minutes);
- *load*: la charge (en grammes);
- *cycles*: le nombre de cycles jusqu'à la rupture.
Dans l'étude présentée ici, on étudie la variable *cycles* fonction de *amp, load* pour les observations 10 jusqu'à 18.
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 ci-joint (sur la page suivante) le code R et les sorties associées qui vous permettront de répondre aux questions suivantes:
1) Pourquoi on a réalisé un test de Shapiro sur la variable *cycles*? Interpréter les résultats de ce test. *(0.5 points)*
2) Quel type de modèle 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) Testez chaque variable qui intervient dans le modèle. Vous donnez les détails pour une seule variable, pour l'autre vous donnez seulement la conclusion. *(1 point)*
5) Donnez les estimations des paramètres du modèle. Interprétez ces estimations. *(1.25 points)*
6) Quelle est la qualité globale d'ajustement de ce modèle? Interprétation. *(0.25 points)*