Master Pro M1, MAIM, Université Claude Bernard, Lyon1
Techniques Probabilites et Statistiques année 2007-2008

Examen du 27 Mai 2008
Une feuille A4 avec les formules, tables des lois admises,
autres documents interdits.
Téléphones portables interdits. Calculatrice autorisée
Durée 3h (le sujet est sur 3 pages)

Exercice 1. *(5 points)*
On considère la famille paramétrique de loi de densité:

$$f_\alpha(x) = \alpha(\alpha+1)x^{\alpha-1}(1-x)1\!\!1_{0\le x\le 1}$$

On observe un n-échantillon $X_1, ..., X_n$ de densité marginale f_α.
1) Montrer que par la méthode des moments d'ordre 1 on obtient l'estimateur $\hat{\alpha}_n = 2\bar{X}_n/(1-\bar{X}_n)$ de α. *(1.5 points)*
2) Montrer que $\hat{\alpha}_n$ est un estimateur fortement consistant de α. *(1.5 points)*
3) Etablir le TLC suivant: $\sqrt{n}(\hat{\alpha}_n - \alpha) \to^{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ avec σ^2 à déterminer. *(2 points)*

Exercice 2. *(4 points)*
On considère un n-échantillon $X_1, ...X_n$ de loi exponnetielle $\mathcal{E}(\mu)$, $\mu > 0$, de densité:

$$f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}1\!\!1_{x>0}$$

1) Montrer que la loi de X est de type exponentiel. *(1 point)*
2) Pour un seuil α fixé et n fixé, trouver le test le plus puissant pour tester l'hypothèse $H_0 : \mu = \mu_0$ contre $H_1 : \mu > \mu_0$ pour un μ_0 connu.*(1.5 points)*
3) Pour un seuil α fixé et n très grand, trouver le test le plus puissant **asymptotique** pour tester l'hypothèse $H_0 : \mu = \mu_0$ contre $H_1 : \mu > \mu_0$ pour un μ_0 connu. *(1.5 points)*
Note: Rappellons que la loi exponentielle de paramètre μ: $\mathcal{E}(\mu)$ est une loi Gamma $\Gamma(1, \mu^{-1})$. Si la variable aléatoire $Y \sim \Gamma(1, \mu^{-1})$ alors $\frac{1}{\mu}Y \sim \Gamma(1, 1)$. Si $Y_1 \sim \Gamma(1, \mu^{-1})$ et $Y_2 \sim \Gamma(1, \mu^{-1})$ sont deux variables aléatoires indépendantes alors $Y_1 + Y_2 \sim \Gamma(2, \mu^{-1})$.

Exercice 3. (*Modèle logistique*) *(6 points)*
On considère les variables: X déterministe et Y aléatoire. La variable aléatoire Y prend deux valeurs 0 et 1. On aimerait modéliser Y fonction de X, mais parce que Y est une variable discrète on va modéliser la variable $l(\pi) := \log(\pi/(1-\pi))$ avec $\pi := I\!P[Y = 1|X, \alpha, \beta]$ la valeur de la probabilité que Y vaut 1 en connaissant les valeurs de X, α et β. On fait n observations pour Y et X. Le modèle considéré est donc le suivant:

$$l(\pi_i) = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1, ...n \tag{1}$$

avec $l(\pi_i) = \log(\pi_i/(1-\pi_i))$ et $\pi_i = I\!P[Y_i = 1|X_i, \alpha, \beta]$, $I\!E(\varepsilon_i) = 0$, $I\!E(\varepsilon_i^2) = \sigma^2$. Les variables aléatoires ε_i sont indépendantes et de même loi.
Pour résumer, pour le modèle (1) seules les valeurs de Y_i (donc aussi de π_i, $l(\pi_i)$) et de X_i sont connues. Les paramètres sont α, β.
1) Donnez l'expression de π_i fonction de α, β et X_i. *(0.5 points)*
2) Pour α, β et X_i donnés, quelle est la loi de la variable aléatoire Y_i? *(1 point)*
3) Montrez que $\frac{\partial \pi_i}{\partial \alpha} = \pi_i(1-\pi_i)$ et $\frac{\partial \pi_i}{\partial \beta} = x_i \pi_i(1-\pi_i)$. *(0.5 points)*
4) En utilisant les questions précédentes, écrivez les équations de vraisemblance pour (α, β), fonction de X_i, Y_i et π_i. *(2 points)*
5) Donnez la matrice de l'information de Fisher (notée $I_i(\alpha, \beta)$) pour une variable aléatoire Y_i fonction de X_i et π_i. Donnez la matrice de l'information de Fisher pour la vecteur $(Y_1, ..., Y_n)$. *(2 points)*

Exercice 4. *(5 points)*
On modélise la variable quantitative SR fonction d'autres variables quantitatives: SB, LC, IC et AGE. Vous disposez des sorties de la fonction *lm* sous R:

1) Quel est le nombre d'observations considérées? *(0.5 points)*
2) Ecrivez le modèle statistique correspondant. Testez le modèle. *(1.5 points)*
3) Donnez les estimations des paramètres du modèle. *(1 point)*
4) Testez chaque variable du modèle. Conclusion.*(1.5 points)*
5) Interprétez le coefficient de détermination R^2. *(0.5 points)*

Note: Les tests sont à faire pour un seuil $\alpha = 0.05$. *Ecrire les hypothèses à tester, les modèles correspondants, les statistiques de test et leurs loi.*