Statistiques appliquées L3 2006/2007

Interrogation n°2

Aucun document n'est autorisé
Les calculatrices programmables ne sont pas autorisées
*Vos téléphonnes portables doivent être **éteints** et **rangés** dans vos sacs*
Durée: 1h20

Justifier soigneusement vos réponses.

Question : Rappeler brièvement la méthodologie de construction d'un intervalle (bilatéral) de confiance pour l'espérance m d'une loi normale $N(m, \sigma^2)$:
- lorsque son écart-type σ est connu
- lorsque son écart-type σ est inconnu

Exercice 1 :
1. On considère un échantillon $(X_1, X_2, .., X_n)$ tiré de façon i.i.d. dans $B(1, p)$.
1.a. Donner les expressions de la vraisemblance et de la log-vraisemblance de cet échantillon.
1.b. En déduire que l'estimateur du maximum de vraisemblance du paramètre p est la proportion empirique $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
1.c. Calculer l'espérance et la variance de l'estimateur $\hat{p}_n$. En déduire que $\hat{p}_n$ est convergent.
1.d. L'estimateur $\hat{p}_n$ est-il efficace?
2. Nous supposons dans cette question que $n = 400$ et que la valeur observée de l'estimateur $\hat{p}_n$ est $0,4$.
2.a. Déterminer un intervalle bilatéral de confiance à 90% pour le paramètre p.
2.b. Déterminer une borne inférieure de confiance à 95% pour p (on obtient ainsi un intervalle unilatéral à droite)
2.c. Déterminer une borne supérieure de confiance à 95% pour p (on obtient ainsi un intervalle unilatéral à gauche).
3. Supposons maintenant que nous disposons d'un autre échantillon $(Y_1, Y_2, .., Y_m)$ tiré de façon i.i.d dans une loi de Bernouilli de paramètre q. Nous supposons en outre que les variables $X_1, X_2, .., X_n, Y_1, Y_2, .., Y_m$ sont indépendantes.
3.a. Déterminer, sous l'hypothèse de validité l'approximation normale da la loi binômiale, l'expression d'un intervalle bilatéral de confiance à 95% pour la différence $p - q$.
3.b. Application : Nous effectuons deux sondages aléatoires indépendants, l'un auprès de la population masculine et l'autre auprès de la population féminine pour déterminer la proportion de fumeurs p dans l'ensemble de la population masculine et la proportion de fumeuses q dans l'ensemble de la population féminine. Le premier sondage porte sur un échantillon de 300 hommes dont 105 déclarent fumer et le second porte sur un échantillon de 200 femmes dont 48 déclarent fumer. Déterminer un intervalle de confiance bilatéral à 95% pour la différence $p - q$.

Exercice 2 :
Considérons le modèle linéaire :

$$y_i = \beta_1 + \beta_2 \; x_i + \varepsilon_i \qquad i = 1, .., N$$

sous les hypothèses standards $E(\varepsilon_i) = 0$; $V(\varepsilon_i) = \sigma^2$ et $cov\,(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$.
1. Que minimise-t-on lorsqu'on utilise la méthode des moindres carrés? En déduire les équations normales qui caractérisent les estimateurs des MCO $\widehat{\beta_1}$ et $\widehat{\beta_2}$ des paramètres β_1 et β_2. Résoudre ces équations et retrouver les expressions classiques des estimateurs $\widehat{\beta_1}$ et $\widehat{\beta_2}$.

2. Rappeler les propriétés des estimateurs des MCO.
Bonus : Vérifiez-les pour $\widehat{\beta_1}$ et $\widehat{\beta_2}$.
2. Donner l'expression d'un estimateur sans biais de σ^2.
3. Posons, pour tout $i = 1, ..N$,

$$q_i = y_i - \bar{y}; \quad p_i = x_i - \bar{x}; \quad u_i = \varepsilon_i - \bar{\varepsilon}$$

3.a. Montrer que pour tout $i = 1, ..N$,

$$q_i = \beta_2 \; p_i + u_i$$

3.b. Montrer que $E(u_i) = 0$ pour tout $i = 1, ..N$
3.c. Montrer que $V(u_i) = \sigma^2 \left(1 + \frac{1}{N}\right)$ et $cov(u_i, u_j) = 0$ pour $i \neq j$.