15 décembre 2017, examen terminal de statistique

Durée de l'épreuve : 2 heures. Notes manuscrites sur une feuille A4 recto-verso, tables statistiques et calculatrice autorisées. La qualité de la rédaction sera prise en compte pour l'évaluation. Le barême donné est approximatif.

Exercice 1 : (8.5 pts)
Quand les fréquences de gènes sont en équilibre, les génotypes AA, Aa et aa se manifestent dans une population avec probabilités $(1-\theta)^2$, $2\theta(1-\theta)$ et θ^2 respectivement, où θ est un paramètre inconnu. Plato *et al.* (1964) ont publié les données suivantes sur le type de haptoglobine dans un échantillon de 190 personnes :

Type de haptoglobine	Hp-AA	Hp-Aa	Hp-aa
effectifs	10	68	112

On propose de modéliser ce problème par une famille de n variables aléatoires $X_1, \dots, X_n$ indépendantes identiquement distribuées à valeurs dans $\{\text{AA}, \text{Aa}, \text{aa}\}$ telles que

$$\mathbb{P}(X = \text{AA}) = (1-\theta)^2 \,,\ \mathbb{P}(X = \text{Aa}) = 2\theta(1-\theta) \text{ et } \mathbb{P}(X = \text{aa}) = \theta^2.$$

1. (1.5 pts) Comment interpréter le paramètre θ ? Donnez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ. Calculez le sur les observations.
   Indication : on pourra introduire les quantités $N_n(\text{AA})$ (*resp.* $N_n(\text{Aa})$, $N_n(\text{aa})$) le nombre de génotypes AA (*resp.* Aa, aa) dans l'échantillon $X_1, \dots, X_n$.

2. (2 pts) Est-ce que l'estimateur est biaisé ? consistant ? Vous argumenterez votre réponse.

3. (1 pt) Donnez la loi asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta)$.

4. (1.5 pts) Proposez un intervalle de confiance de niveau asymptotique 95% pour θ. Evaluez cet intervalle de confiance sur les observations.

5. (2.5 pts) Proposez un test asymptotique de l'hypothèse nulle $\mathcal{H}_0 \ : \ \theta = 0.8$ contre l'hypothèse alternative $\mathcal{H}_1 \ : \ \theta \neq 0.8$. Mettez ce test en œuvre sur les observations, au seuil de 5%.

Exercice 2 : (4 pts)
Soient $X_1, \dots, X_n$ indépendantes identiquement distribuées selon la densité de probabilité $f_\theta(x) = \theta x^{-\theta-1}\mathbb{1}_{x\geq 1}$, où $\theta > 0$.

1. (1pt) On suppose $\theta > 1$. Estimez θ par la méthode des moments.

2. (1 pt) Estimez θ par maximum de vraisemblance.

3. (1 pt) Calculez l'information de Fisher de ce modèle.

4. (1 pt) Donnez la variance minimale qu'un estimateur sans biais de θ peut atteindre.

Exercice 3 : (7.5 pts)
Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes identiquement distribuées selon la loi uniforme sur $[a, b]$ où a et b sont des paramètres réels inconnus tels que $a < b$.

1. **Estimateur par la méthode des moments.**
   (a) (1 pt) Soit X une variable de loi uniforme sur $[a, b]$. Calculez $\mathbb{E}X$ et $\mathbb{E}X^2$ en fonction de a et de b.
   (b) (2 pts) Proposez un estimateur $(\hat{a}, \hat{b})$ par la méthode des moments.
   Indication : passez par la résolution d'un système de deux équations à deux inconnues (les paramètres a et b).
   (c) (1 pt) Etudiez la consistance des estimateurs $\hat{a}$ et $\hat{b}$.

2. **Estimateur par la méthode du maximum de vraisemblance.**
   (a) (1 pt) Donnez l'estimateur $(\tilde{a}, \tilde{b})$ maximisant la vraisemblance en (a, b).
   (b) (1 pt) Soit $\varepsilon > 0$. Majorez $\mathbb{P}_{(a,b)}(|\tilde{a} - a| > \varepsilon)$. Déduisez-en la consistance de $\tilde{a}$.
   (c) (1 pt) Etudiez la loi limite de $n(\tilde{a} - a)$.
   (d) (0.5 pt) Comparez les estimateurs $\tilde{a}$ et $\hat{a}$.