M1 – Statistique – 2019/2020

Contrôle continu
Durée : 3h

Tout matériel qui ne ressemblerait pas à un simple stylo ou à une feuille blanche (calculatrice, portable, note de cours...) est interdit. Si vous vous demandez si vous avez le droit d'avoir un objet sur la table, c'est qu'il est certainement interdit.
L'énoncé est composé d'un vrai ou faux et d'un exercice.

1 Vrai ou Faux

Cette section est composée de 10 affirmations dont il faudra dire si elles sont vraies ou fausses. Une mauvaise réponse enlèvera des points. Aucune justification n'est demandée mais vous pouvez mettre votre raisonnement sur la copie ; si la réponse est mauvaise mais que le raisonnement tient la route, il est possible qu'aucun point ne soit enlevé.

1. Soient N un entier naturel supérieur ou égal à 2, U une variable aléatoire uniforme sur $\{1,\dots,N\}$, p un diviseur de N et A_p l'événement "*p divise U*". La probabilité de l'événement A_p est $1/p$.

2. Le modèle défini sur $\mathbb{R}$ muni de la tribu borélienne et de la famille de lois gaussiennes $\mathcal{N}(1,\theta^2)$ avec $\theta \in \mathbb{R}^\star$ est identifiable.

3. Étant donné $\mathbf{X} = (X_1,\dots,X_d)^T$ un vecteur gaussien d'espérance $m = \mathbb{E}[\mathbf{X}]$ et de matrice de variance covariance $\Sigma = \mathbb{V}[\mathbf{X}]$, la loi de $\mathbf{X}$ admet une densité $f_{m,\Sigma}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ valant pour tout $\mathbf{x} \in \mathbb{R}^d$,
$$f_{m,\Sigma}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^d}\frac{1}{\sqrt{|\Sigma|}}e^{-\frac{1}{2}(\mathbf{x}-m)^T\Sigma^{-1}(\mathbf{x}-m)}$$

4. Étant données une fonction de répartition F de densité f et U une variable aléatoire de loi uniforme sur $[0,1]$, la loi de la variable $X = F(U)$ admet pour densité f.

5. Étant donnés deux estimateurs d'un paramètre d'intérêt $g(\theta^\star)$ dont le premier $\widehat{\theta}_n^{(1)}$ est non biaisé asymptotiquement normal et le deuxième $\widehat{\theta}_n^{(2)}$ est biaisé, mais asymptotiquement sans biais, de vitesse de convergence de l'ordre de n^2 ; alors il est préférable d'utiliser le premier estimateur $\widehat{\theta}_n^{(1)}$.

6. Étant donnés un événement A tel que $\mathbb{P}(A) = p \in [0;1]$ et une suite d'événements $(A_n)_{n\in\mathbb{N}}$ telle que pour tout $n \in \mathbb{N}$, $A_n = A$ alors la probabilité de la limite supérieure de la suite $(A_n)_{n\in\mathbb{N}}$ est 1.

7. Étant donné un estimateur $\widehat{g}$ de $g(\theta^\star)$, nous avons la relation suivante :
$$R(\theta^\star) = b(\theta^\star) + \mathbb{V}_{\theta^\star}[\widehat{g}]$$
où R est le risque (quadratique) et b la fonction de biais.
8. Étant donné $\mathbf{X} = (X_1, \ldots, X_d)$ un vecteur aléatoire alors pour tout $(k, \ell) \in \{1, \ldots, d\}$ avec k différent de ℓ, nous avons :
$$X_k \text{ et } X_\ell \text{ sont indépendants} \Leftrightarrow \operatorname{cov}(X_k, X_\ell) = 0.$$
9. Soit X une variable aléatoire représentant le nombre de piles obtenus sur 1000 lancers de pièces équilibrées alors X suit une loi géométrique de paramètre $1/2$.
10. Il est possible qu'un estimateur soit biaisé mais asymptotiquement sans biais.

2 Exercice 1

Si un résultat n'est pas rigoureusement justifié dans cette partie, la totalité des points ne sera pas donnée. ***Si un résultat n'a pas pu être démontré****, il pourra toutefois être admis pour les questions suivantes.*

Nous considérons la famille de densités sur $\mathbb{R}^+$ donnée par
$$f_{\alpha,\beta}(x) = C(\alpha, \beta)\frac{1}{x^{3+\alpha}}\mathbb{1}_{[\beta,+\infty[}(x)$$
avec $\alpha > 0$ et $\beta > 0$.

2.1. Calculer $C(\alpha, \beta)$, $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

2.2. Nous considérons deux cas :
(I) α est connu et β est inconnu.
(II) β est connu et α est inconnu.

Nous répondrons aux deux questions suivantes (2.2.a et 2.2.b) pour les deux cas (I) et (II).

2.2.a. Utiliser $\mathbb{E}[X]$ pour calculer l'estimateur de la méthode des moments avec n observations i.i.d.

2.2.b. Toujours avec n observations i.i.d., calculer l'estimateur du maximum de vraisemblance.

2.3. Dans le cas (I) uniquement et si nous notons $\widehat{\beta}$ l'estimateur du maximum de vraisemblance calculé dans la question 2.b (cas (I)), donner sa fonction de répartition $F_{\widehat{\beta}}$ puis celle de $n\left(\widehat{\beta} - \beta^\star\right)$ où $\beta^\star$ est le vrai paramètre. En déduire la loi limite de $n\left(\widehat{\beta} - \beta^\star\right)$ lorsque n tend vers $+\infty$.

2.4. Montrer que :
$$\mathbb{E}[\ln X] = \frac{1}{2+\alpha} + \ln\beta \text{ et } \mathbb{V}[\ln X] = \frac{1}{(2+\alpha)^2}.$$

2.5. Dans le cas (II) uniquement, étudier la normalité asymptotique des deux estimateurs de α vus précédemment. Lequel préférez-vous ?