Statistique Mathématique

Examen — 5 mai 2022

Durée: 3 heures. Documents interdits. Justifiez bien vos réponses et soignez votre présentation. Les deux parties sont indépendantes. Bon courage !

Exercice I

Vous venez d'être recruté dans un laboratoire pharmaceutique et on vous charge de concevoir un test pour détecter une maladie génétique rare. Cette maladie concerne une personne sur 20000. On souhaite que le test sache reconnaître une personne malade avec 95 chances sur 100 et qu'au plus la moitié des personnes positives à ce test soient saines. On pourra utiliser sans justification la formule de Bayes: $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$. On se contentera de donner les résultats numériques sous forme de fraction, sans chercher à calculer des valeurs approchées.

I.1. Formulez l'hypothèse nulle et l'hypothèse alternative de ce test. Justifiez votre réponse.

I.2. Quelle est la probabilité maximale d'être positif à ce test ?

I.3. Quel est le niveau souhaité du test ?

I.4. Quelle est la puissance souhaitée de ce test ?

Concrètement, ce test consiste à estimer la concentration d'un marqueur biologique dans un échantillon. Si la personne est saine, cette concentration vaut exactement μ_0 ; si elle est malade, exactement $\mu_0 + \Delta$ avec μ_0 et $\Delta > 0$ connus. On suppose que la mesure de cette concentration suit une loi normale d'écart-type σ. En investissant davantage pour augmenter la qualité du procédé de mesure, il est possible de réduire cet écart-type autant que vous le souhaitez.

I.5. En admettant que si Z suit la loi normale centrée réduite, alors pour tout $x > 0$, $\mathbb{P}(Z \geq x) \leq \mathrm{e}^{-x^2/2}$, donner un majorant du quantile d'ordre $1 - \alpha$ de la loi normale $\mathcal{N}\left(a, b^2\right)$ en fonction de $a \in \mathbb{R}$, $b > 0$ et $\alpha \in [0, 1/2[$.

I.6. Quelle valeur de σ pouvez-vous fixer pour satisfaire le cahier des charges ?

Exercice II

Dans cet exercice, on s'intéresse à l'estimation des paramètres d'une loi de Cauchy. La densité de cette loi est donnée par

$$\forall x \in \mathbb{R}, \qquad f(x; x_0, \gamma) = \frac{1}{\pi\gamma} \cdot \frac{\gamma^2}{(x - x_0)^2 + \gamma^2},$$

où $x_0 \in \mathbb{R}$ est appelé paramètre de position et $\gamma > 0$ paramètre d'échelle.

II.1. Représenter graphiquement la densité d'une loi de Cauchy de paramètres $x_0 = 0$ et $\gamma \in \{1, 2\}$.

II.2. Expliquez le nom donné aux paramètres x_0 et γ.

On se donne maintenant un n-échantillon i.i.d. $X_1, \ldots, X_n$ suivant la loi de Cauchy de paramètres inconnus x_0 et γ.

II.3. Expliquez pourquoi on ne peut pas mettre en œuvre la méthode des moments ici.

II.4. Écrire la vraisemblance du n-échantillon $X_1, \ldots, X_n$.

II.5. Calculez les dérivées partielles de la log-vraisemblance calculée à la question précédente par rapport à x_0 et γ.

II.6. Montrez que l'on peut maximiser la vraisemblance en fonction de x_0 et γ en résolvant des équations polynomiales dont on précisera le degré.

II.7. On sait qu'on ne peut pas (en géneral) résoudre d'équation polynomiale de degré supérieur à 5. Quelle stratégie pouvez-vous proposer pour résoudre ces équations lorsque n est grand ?