ENSIMAG $2^{\text{ème}}$ année Mai 2015

Statistique Inférentielle Avancée

Durée : 3 heures. Tous documents autorisés.

Barème indicatif - Partie 1 : 6 pts, Partie 2 : 9 pts, Partie 3 : 5 pts, Question bonus : 2 pts. Les trois parties sont indépendantes.

Première partie

On considère un échantillon $X_1, \ldots, X_n$ de variables aléatoires indépendantes et de même loi de Poisson $\mathcal{P}(\lambda)$. On rappelle que la statistique $\sum_{i=1}^{n} X_i$ est exhaustive et complète et que l'estimateur de maximum de vraisemblance de λ est $\bar{X}_n$. On souhaite estimer $\theta = e^{-\lambda}$.

1. Donner l'estimateur de maximum de vraisemblance de θ. Calculer son espérance. En déduire qu'il est biaisé mais asymptotiquement sans biais.
2. Soit N le nombre d'observations nulles dans l'échantillon : $N = \sum_{i=1}^{n} \mathbb{1}_{\{X_i=0\}}$. Donner la loi de probabilité de N. Construire un estimateur sans biais de θ à l'aide de N. Expliquer en quoi cet estimateur est naturel.
3. Déterminer l'estimateur sans biais et de variance minimale de θ.

Deuxième partie

Pour estimer un paramètre réel dans un modèle statistique paramétrique, on considère qu'il faut choisir l'estimateur qui minimise l'erreur quadratique moyenne (EQM). Quand l'estimateur est sans biais, son EQM est égale à sa variance et on cherche alors un estimateur sans biais et de variance minimale (ESBVM). Mais on peut imaginer avoir un estimateur biaisé dont l'EQM est inférieure à celle de l'ESBVM. Cet estimateur, bien que biaisé, serait donc meilleur que l'ESBVM au sens de l'EQM. Le but de cet exercice est de construire un tel estimateur dans le cas très simplifié de l'estimation de la moyenne d'une loi normale de variance connue, quand on a une seule observation.

1. Soit U une variable aléatoire de loi normale centrée réduite $\mathcal{N}(0,1)$ et h une fonction continue et dérivable sur $\mathbb{R}$. Montrer que, sous une condition sur h à déterminer, on a : $E\left[U h(U)\right] = E\left[h'(U)\right]$.
2. Pour toute fonction h 2 fois dérivable, montrer que l'on a :

$$\forall x \in \mathbb{R},\ \left[(\ln h(x))'\right]^2 + 2\left[(\ln h(x))''\right] = 4\frac{\sqrt{h(x)}''}{\sqrt{h(x)}}$$

3. Soit X une variable aléatoire de loi normale $\mathcal{N}(m, \sigma^2)$, où σ^2 est connue. On considère X comme un échantillon de taille $n = 1$ de cette loi et on souhaite estimer m au vu de X. Rappeler sans calculs quel est l'ESBVM de m et donner son EQM.

4. On propose d'estimer m par un estimateur de la forme :

$$\hat{m} = X + g(X)$$

   Ecrire l'expression de l'EQM de $\hat{m}$ et montrer que $\hat{m}$ sera un meilleur estimateur que l'ESBVM au sens de l'EQM si et seulement si $A < 0$ où :

$$A = 2E\left[(X - m)g(X)\right] + E\left[g(X)^2\right]$$

5. On pose $U = (X - m)/\sigma$. A l'aide de la question 1, montrer que

$$A = 2\sigma^2 E[g'(X)] + E[g(X)^2]$$

6. Ecrire A quand la fonction g est de la forme $g(x) = \sigma^2[\ln h(x)]'$.

7. En déduire une condition suffisante sur h pour que $A < 0$.

8. Proposer une fonction g et un estimateur $\hat{m}$ dont l'EQM est inférieure à celle de l'ESBVM. Donner l'expression de cette EQM, sans chercher à la calculer explicitement.

Troisième partie

On considère un échantillon $X_1, \ldots, X_n$ de la loi uniforme sur $[0, 1]$.

1. Pour $i \in \{1, \ldots, n\}$, calculer la densité de la $i^{\text{ème}}$ statistique d'ordre X_i^*. En déduire que X_i^* est de loi beta dont on précisera les paramètres. Donner l'expression de l'espérance et de la variance de cette loi. Expliquez pourquoi le résultat sur l'espérance est logique.

2. On admet que, pour $(i, j) \in \{1, \ldots, n\}^2$, avec $i < j$, $Cov\left(X_i^*, X_j^*\right) = \dfrac{i(n - j + 1)}{(n + 1)^2(n + 2)}$. Calculer le coefficient de corrélation linéaire $\rho\left(X_i^*, X_j^*\right)$. En déduire que le minimum et le maximum d'un tel échantillon sont asymptotiquement non corrélés.

3. On considère maintenant un échantillon $X_1, \ldots, X_n$ d'une loi de probabilité dont la fonction de répartition F est continue et inversible. Donner la loi de probabilité de $F(X_i^*)$, $\forall i \in \{1, \ldots, n\}$. Expliquer pourquoi on peut approcher $E(X_i^*)$ par $F^{-1}\left(\dfrac{i}{n + 1}\right)$.

4. *Question bonus* : Montrer le résultat admis dans la question 2 : pour $(i, j) \in \{1, \ldots, n\}^2$, avec $i < j$, $Cov\left(X_i^*, X_j^*\right) = \dfrac{i(n - j + 1)}{(n + 1)^2(n + 2)}$.