UNIVERSITE PARIS-DAUPHINE
Département LSO- DUGEAD
UV 44 - Statistiques II
Groupes 19

Contrôle continu n$^{\text{o}}$ 2.

20-01-10. **Durée:1h20**

Eléments de correction

Calculatrices non-programmables autorisées. Téléphones portables interdits. Aucun document autorisé.

Problème

Soit X une v.a. admettant pour densité, connue à une constante $k > 0$ près,

$$f(x;\theta) = k \exp(-\theta|x|) \qquad \text{pour tout } x \in \mathbb{R}$$

où $\theta > 0$ est un paramètre inconnu à estimer et $X_1, \dots, X_n$ un n-échantillon de X.

1. Vérifier que nécessairement $k = \theta/2$ pour que f soit bien une densité.

   On vérifie bien que f est positive, continue (presque partou)t et que
   $$\int_{-\infty}^{\infty} \exp(-\theta|x|)dx = 2\int_0^{\infty} \exp(-\theta x)dx = \frac{2}{\theta}$$
   donc $k = \theta/2$ entraîne que $\int f(x)dx = 1$ et que f est une densité.

2. Montrer que $\mathbb{E}(X) = 0$.

   Sa densité f est paire, donc la v.a. X est symétrique donc $\mathbb{E}(X) = 0$.

3. En déduire que $\mathbb{E}(X^2) = \mathrm{Var}(X) = 2/\theta^2$. Huygens donne $\mathbb{E}(X^2) = \mathrm{Var}(X)$ et on calcule
   $$\mathbb{E}(X^2) = \frac{\theta}{2}\int_{-\infty}^{\infty} x^2 \exp(-\theta|x|)dx = \theta \int_0^{\infty} x^2 \exp(-\theta x)dx.$$
   Par IPP, $x^2 \mapsto 2x$ et $\theta \exp(-\theta x) \mapsto -\exp(-\theta x)$ on trouve
   $$\mathbb{E}(X^2) = 2\int_0^{\infty} x \exp(-\theta x)dx.$$
   On conclut par une nouvelle IPP $2x \mapsto 2$ et $\exp(-\theta x) \mapsto -\theta^{-1}\exp(-\theta x)$:
   $$\mathbb{E}(X^2) = 2\theta^{-1}\int_0^{\infty} \exp(-\theta x)dx = 2\theta^{-1}[-\theta^{-1}\exp(-\theta x)]_0^{\infty} = \frac{2}{\theta^2}.$$

4. Déterminer un estimateur T_n de θ par la méthode des moments.

   La méthode des moments donne directement
$$T_n = \sqrt{\frac{2n}{\sum_{i=1}^n X_i^2}}.$$

5. Donner les propriétés de convergence de T_n.

   Etant donné qu'il est obtenu par la méthode des moments, T_n est convergent en probabilité.

6. Déterminer l'estimateur W_n de θ par la méthode du maximum de vraisemblance.

   On calcule la vraisemblance
$$L_n(\theta) = \left(\frac{\theta}{2}\right)^n \exp\left(-\theta \sum_{i=1}^n |x_i|\right)$$
   son logarithme
$$\ln L_n(\theta) = n\ln(\theta) - n\ln(2) - \theta \sum_{i=1}^n |x_i|$$
   la dérivée de son logarithme
$$(\ln L_n(\theta))' = \frac{n}{\theta} - \sum_{i=1}^n |x_i|$$
   et la dérivée seconde
$$(\ln L_n(\theta))'' = -\frac{n}{\theta^2} < 0$$
   La CN est directement vérifiée quelque soit $\theta > 0$ et la CS donne
$$\frac{n}{\hat{\theta}^{MV}} - \sum_{i=1}^n |x_i| = 0 \Longleftrightarrow \hat{\theta}^{MV} = \frac{n}{\sum_{i=1}^n |x_i|}$$

7. Calculer la fonction de répartition de $|X|$ définie par $\mathbb{P}(|X| \leq x)$ pour $x > 0$.

   On a $\mathbb{P}(|X| \leq x) = \mathbb{P}(-x \leq X \leq x)$ pour $x > 0$ donc
$$F_{|X|}(x) = \int_{-x}^{x} \frac{\theta}{2} \exp(-\theta|u|)du = \theta \int_0^x \exp(-\theta u)du = [-\exp(-\theta u)]_0^x = 1 - \exp(-\theta x).$$

8. En déduire que $Y = |X|$ suit une loi exponentielle $\mathcal{E}(\theta)$.

   On reconnaît dans l'expression de $F_{|X|}(x)$ la fonction de répartition de $\mathcal{E}(\theta)$ donc $Y = |X|$ suit une loi exponentielle $\mathcal{E}(\theta)$.

9. Vérifier, par le calcul, que $\mathbb{E}(Y) = \theta^{-1}$.

   On fait une IPP $x \mapsto 1$ et $\theta^{-1}\exp(-\theta y) \mapsto \exp(-\theta y)$: $\mathbb{E}(Y) = \theta^{-1}\int_0^\infty y\exp(-\theta y)dy = \int_0^\infty \exp(-\theta y)dy$

10. En déduire que l'information de Fisher s'écrit sous la forme $I_n(\theta) = n\,\mathrm{Var}(Y)$.

    Par définition $I_n(\theta) = n\mathbb{E}((\theta^{-1} - |X|)^2) = n\mathbb{E}((Y - \mathbb{E}(Y))^2) = n\,\mathrm{Var}(Y)$

11. Donner (sans calcul) $\mathrm{Var}(Y)$ et en déduire que $\sqrt{n}(W_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta)$.

    On a $\mathrm{Var}(Y) = \theta^{-2}$ d'après le cours et d'après l'approximation normale de Fischer on a
    $$W_n - \theta \sim \mathcal{N}(0, \sigma_n)$$
    où $\sigma_n^2 = I_n^{-1} = \theta^2/n$. Le résultat vient de la multiplication par $\sqrt{n}$ à droite et à gauche de l'approximation et en faisant tendre n vers l'infini.

12. En déduire que $\sqrt{n}(W_n/\theta - 1)$ est une statistique pivotale asymptotique pour θ dont on précisera la loi asymptotique.

    On réduit le résultat trouvé précédemment en divisant par l'écart type θ.

13. Construire un intervalle de confiance de niveau asymptotique 95% pour le paramètre inconnu θ.

    Asymptotiquement, on a
    $$\begin{aligned}\mathbb{P}\left(\left|\sqrt{n}(W_n/\theta - 1)\right| \leq 1,96\right) = 0,95 &\iff \\ \mathbb{P}\left(1 - 1,96/\sqrt{n} \leq W_n/\theta \leq 1 + 1,96/\sqrt{n}\right) = 0,95 &\iff \\ \mathbb{P}\left((1 + 1,96/\sqrt{n})^{-1} \leq \theta/W_n \leq (1 - 1,96/\sqrt{n})^{-1}\right) = 0,95 &\end{aligned}$$
    d'où
    $$IC_{95\%} = \left[\frac{W_n}{1 + 1,96/\sqrt{n}}; \frac{W_n}{1 - 1,96/\sqrt{n}}\right]$$

14. En déduire une fourchette d'estimation pour θ de risque asymptotique égal à 5% sachant qu'on observe 100 observations dont la somme des valeurs négatives est -4723 et la somme des valeurs positives est 5178.

    AN : On trouve la fourchette d'estimation $0,84\% \leq \theta \leq 1,26\%$

15. La fourchette d'estimation est-elle centrée en l'estimation $\hat{\theta}^{MV}$ obtenue par la méthode du Maximum de Vraisemblance?

    On calcule $w_n = 0,01 = 1\%$ et la fourchette n'est pas centrée en 1%, $0,84$ étant plus proche de 1 que $1,26$.

16. En supposant l'estimation $\hat{\theta}^{MV}$ fixée, quel est le nombre d'observations nécessaire pour doubler la précision de la fourchette (i.e. diviser par deux sa longueur)?

La longueur trouvée vaut $0,41\%$. On recherche le plus petit n tel que

$$w_n \left(\frac{1}{1 - 1,96/\sqrt{n}} - \frac{1}{1 + 1,96/\sqrt{n}} \right) \leq 0,2\%.$$

En remplaçant w_n par sa valeur 1% supposé constante, en réduisant au même dénominateur, on trouve

$$\frac{2 * 1,96\sqrt{n}}{n - (1,96)^2} \leq 0,41$$

soit, en multipliant par $n - (1,96)^2/0,2$ à droite et à gauche :

$$n - 19,6\sqrt{n} - 3,84 \geq 0$$

On obtient un polynôme en $\sqrt{n}$ dont on calcule le discriminant $\Delta = 400$ d'où la racine positive $(19,6 + 20)/2 = 19,8$. On en déduit finalement $n = 393$ le plus petit entier plus grand que $19,8^2 = 392,04$.