Inférence statistique devoir 2 (date limite : le jeudi 15 février)

Wen, Zehai ; Li, Qingyue

2024-02-14 11:48:20-05:00

Exercice 1

Soit $X_1, X_2, \cdots, X_n$ un échantillon de taille n avec densité $f_\theta(x) = \frac{1+\theta x}{2}\chi_{[-1,1]}(x)$, où $\theta \in [-1,1]$ est inconnu. Écrire $Y_i = \chi_{[0,1]}(X_i)$ pour tout $i \in \{1,2,\cdots,n\}$. Estimer θ par la méthode des moments utilisant $\{X_i\}_{i=1}^n$ et calculer l'écart quadratique moyenne. Proposez un autre estimateur de θ qui améliore cet écart quadratique moyenne. Répétez les mêmes tâches utilisant $\{Y_i\}_{i=1}^n$ et pas $\{X_i\}_{i=1}^n$.

Solution Écrire $\hat{\theta}_x$ et $\hat{\theta}_y$ pour les estimateurs de θ utilisant la méthode de moments basés sur $\{X_i\}_{i=1}^n$ et $\{Y_i\}_{i=1}^n$ respectivement. Comme $\mathbb{E}[X_1] = \frac{\theta}{3}$ et $\mathbb{E}[Y_1] = \frac{1}{2} + \frac{\theta}{4}$, on a $\hat{\theta}_x = \frac{3}{n}\sum_{i=1}^n X_i$, $\hat{\theta}_y = \frac{4}{n}\sum_{i=1}^n Y_i - 2$. Alors :

$$\mathbb{E}\left[(\hat{\theta}_x - \theta)^2\right] = \mathbb{E}\left[\frac{9}{n^2}\sum_{i=1}^n X_i^2 + \frac{9\times 2}{n^2}\sum_{i=1}^n\sum_{j=1}^{i-1} X_iX_j + \theta^2 - \frac{6\theta}{n}\sum_{i=1}^n X_i\right]$$
$$= \frac{9}{n}\frac{1}{3} + \frac{18}{n^2}\frac{n(n-1)}{2}\left(\frac{\theta}{3}\right)^2 + \theta^2 - 6\theta\frac{\theta}{3} = \frac{3}{n} - \frac{\theta^2}{n}$$

Ici on utilise le fait que $\mathbb{E}[X_1^2] = \frac{1}{3}$. Ensuite :

$$\mathbb{E}\left[(\hat{\theta}_y - \theta)^2\right] = \mathbb{E}\left[\frac{16}{n^2}\sum_{i=1}^n Y_i^2 + \frac{16\times 2}{n^2}\sum_{i=1}^n\sum_{j=1}^{i-1} Y_iY_j + \theta^2 + 4 - \frac{16}{n}\sum_{i=1}^n Y_i - \frac{8\theta}{n}\sum_{i=1}^n Y_i + 4\theta\right]$$
$$= \frac{16}{n}\left(\frac{1}{2}+\frac{\theta}{4}\right) + \frac{32}{n^2}\frac{n(n-1)}{2}\left(\frac{1}{2}+\frac{\theta}{4}\right)^2 + \theta^2 + 4 - 16\left(\frac{1}{2}+\frac{\theta}{4}\right) - 8\theta\left(\frac{1}{2}+\frac{\theta}{4}\right) + 4\theta = \frac{4}{n} - \frac{\theta^2}{n}$$

Pour améliorer les écarts quadratique moyen, on propose des nouveaux estimateurs $\hat{\theta}_x^{adj}$ et $\hat{\theta}_y^{adj}$ comme suit :

$$\hat{\theta}_x^{adj} = \left(\frac{3}{n}\sum_{i=1}^n X_i\right) X_i\chi_{[-1,1]}\left(\frac{3}{n}\sum_{i=1}^n X_i\right) + \chi_{(1,\infty)}\left(\frac{3}{n}\sum_{i=1}^n X_i\right) - \chi_{(-\infty,-1)}\left(\frac{3}{n}\sum_{i=1}^n X_i\right)$$
$$\hat{\theta}_y^{adj} = \left(\frac{4}{n}\sum_{i=1}^n Y_i - 2\right)\chi_{[-1,1]}\left(\frac{4}{n}\sum_{i=1}^n Y_i - 2\right) + \chi_{(1,\infty)}\left(\frac{4}{n}\sum_{i=1}^n Y_i - 2\right) - \chi_{(-\infty,-1)}\left(\frac{4}{n}\sum_{i=1}^n Y_i - 2\right)$$

Alors :

$$\mathbb{E}\left[(\hat{\theta}_x^{adj} - \theta)^2\right] = \mathbb{E}\left[(\hat{\theta}_x - \theta)^2; \left\{\hat{\theta}_x \in [-1,1]\right\}\right] + \mathbb{E}\left[(1-\theta)^2; \left\{\hat{\theta}_x > 1\right\}\right] + \mathbb{E}\left[(-1-\theta)^2; \left\{\hat{\theta}_x > 1\right\}\right]$$
$$< \mathbb{E}\left[(\hat{\theta}_x - \theta)^2; \left\{\hat{\theta}_x \in [-1,1]\right\}\right] + \mathbb{E}\left[(\hat{\theta}_x - \theta)^2; \left\{\hat{\theta}_x > 1\right\}\right] + \mathbb{E}\left[(\hat{\theta}_x - \theta)^2; \left\{\hat{\theta}_x > 1\right\}\right] = \mathbb{E}\left[(\hat{\theta}_x - \theta)^2\right]$$

On souligne que l'inégalité est stricte parce que les ensembles $\left\{\hat{\theta}_x > 1\right\}$ et $\left\{\hat{\theta}_x < -1\right\}$ ont mesures positives. Alors, $\hat{\theta}_x^{adj}$ améliore l'écart quadratique moyen de $\hat{\theta}_x$. Changer tout x à y, on montre que $\hat{\theta}_y^{adj}$ améliore l'écart quadratique moyen de $\hat{\theta}_y$. Les démonstrations sont complètes. ////

Exercice 2

On a un échantillon de taille 2 de Cauchy$(\theta, 1)$, où le centre $\theta \in \mathbb{R}$ est inconnu. Estimer θ selon la méthode du maximum de vraisemblance.

Solution Étant observé x_1 et x_2, la fonction de vraisemblance $L : \mathbb{R} \to \mathbb{R}$ définie par, pour tout $\theta \in \mathbb{R}$:

$$L(\theta) = \frac{1}{\pi^2} \frac{1}{1 + (x_1 - \theta)^2} \frac{1}{1 + (x_2 - \theta)^2}$$

Définir $\eta = \theta - \frac{x_1+x_2}{2}$ et $y = \frac{x_1}{2} - \frac{x_2}{2}$. Il s'agit de minimiser le dénominateur parmi tout $\eta \in \mathbb{R}$, qui est :

$$\pi^2 \left[(\eta^2 + 1 - y^2)^2 - (1 - y^2)^2 + (y^2 + 1)^2\right]^2$$

Si $y^2 \geqslant 1$, on voit immédiatement que $\eta = \pm\sqrt{y^2 - 1}$ est le minimum global. Si $y^2 < 1$, le développement de dénominateur est un polynôme de η^2 avec des coefficients positifs et $\eta = 0$ est le minimum global. Alors, l'estimateur de maximum de vraisemblance de θ, $\hat{\theta}$, est :

$$\hat{\theta} = \begin{cases} \frac{x_1+x_2}{2} & \left(\frac{x_1-x_2}{2}\right)^2 \leqslant 1 \\ \sqrt{\left[\frac{x_1-x_2}{2}\right]^2 - 1} + \frac{x_1+x_2}{2} \text{ou} - \sqrt{\left[\frac{x_1-x_2}{2}\right]^2 - 1} + \frac{x_1+x_2}{2} & \left(\frac{x_1-x_2}{2}\right)^2 > 1 \end{cases}$$

Le calcul est complet. ////

Exercice 3

Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction continue et décroissante. On suppose que toutes les fonctions $f_\theta : \mathbb{R} \to \mathbb{R}$, où $\theta \in \mathbb{R}$, définies par $f_\theta(x) = g(|x - \theta|)$ sont densités de probabilité. Soit $x_1, x_2, \cdots, x_n$ un échantillon de taille n avec densité f_θ, où $\theta \in \mathbb{R}$ est inconnu. Écrire $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$ pour les statistiques d'ordre. Montrez que l'estimateur de maximum de vraisemblance de θ existe et se trouve dans l'intervalle $[x_{(1)}, x_{(n)}]$.

Solution Écrire $L(\theta) = \prod_{i=1}^n g(|x_i - \theta|)$. Comme $[x_{(1)}, x_{(n)}]$ est compacte et compositions de fonctions continues sont continues, on voit que $\arg\max_{\theta \in [x_{(1)}, x_{(n)}]} L(\theta) \neq \emptyset$. Tout $\hat{\theta} \in \arg\max_{\theta \in [x_{(1)}, x_{(n)}]} L(\theta)$ est un estimateur de maximum de vraisemblance de θ parce que $L(\theta) \leqslant L(x_{(1)})$ pour tout $\theta < x_{(1)}$ et $L(\theta) \leqslant L(x_{(n)})$ pour tout $\theta > x_{(n)}$. ////