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

Wen, Zehai ; Li, Qingyue

2024-02-13 17:45:50-05:00

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 L'espérance de X_1, $E[X_1]$, s'obtient en intégrant le produit de X_1 par la densité de probabilité sur l'intervalle de définition de X_1, ce qui nous donne :

$$E[X_1] = \int_{-1}^{1} x_1 \cdot \frac{1 + \theta x_1}{2}\, dx_1$$

Cette intégration aboutit à :

$$E[X_1] = \frac{\theta}{3}$$

$$E[X_1^2] = \frac{1}{3}$$

En utilisant la méthode des moments, on égalise le moment théorique, ici l'espérance calculée $E[X_1]$, avec le moment empirique, la moyenne d'échantillon $\bar{X}$, ce qui nous donne l'équation :

$$\bar{X} = \frac{\theta}{3}$$

En résolvant cette équation pour θ, nous obtenons l'estimateur de θ par la méthode des moments,$\hat{\theta_x}$:

$$\hat{\theta_x} = 3\bar{X}$$

$$\hat{\theta_x} = 3\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

L'écart quadratique moyen d'un estimateur $\hat{\theta_x}$ est défini par :

$$EQM(\hat{\theta_x}) = E[(\hat{\theta_x} - \theta)^2] = E\left[9\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - 6\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + \theta^2\right]$$

On procède maintenant à la substitution de $E[X_1]$ et $E[X_1^2]$ dans notre expression précédente pour simplifier le calcul.

$$EQM(\hat{\theta_x}) = \frac{3 - \theta^2}{n}$$

La contribution attendue de X_1 sous cette condition, ou l'espérance de Y_1, est calculée comme suit:

$$E[Y_1] = \int_0^1 \frac{1 + \theta x_1}{2} dx_1 = \frac{1}{2} + \frac{\theta}{4}$$

Comme Y_1 est une fonction indicatrice dont les valeurs peuvent uniquement être 0 ou 1, la valeur de Y_1^2 est en réalité identique à celle de Y_1. Cela est dû au fait que $1^2 = 1$ et $0^2 = 0$. Par conséquent, nous avons $E[Y_1^2] = E[Y_1]$.

Par la méthode des moments, l'alignement de l'espérance théorique $E[Y_1]$ avec la moyenne empirique des observations de Y_1 donne :

$$E[Y_1] = \frac{1}{2} + \frac{\theta}{4} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Résoudre cette équation pour θ, on introduis l'estimateur $\hat{\theta}_y$, qui est défini et obtenu par :

$$\hat{\theta}_y = 4\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) - 2,$$

offrant une estimation directe de θ basée sur la moyenne d'échantillon des Y_i.

L'écart quadratique moyen d'un estimateur $\hat{\theta}_y$ est défini par :

$$EQM(\hat{\theta}_y) = E[(\hat{\theta}_y - \theta)^2] = E\left[16\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 - 8\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)(2+\theta) + (2+\theta)^2\right]$$

On procède maintenant à la substitution de $E[Y_1]$ et $E[Y_1^2]$ dans notre expression précédente pour simplifier le calcul:

$$EQM(\hat{\theta}_y) = \frac{4-\theta^2}{n}$$

Considérant l'estimateur $\hat{\theta}_x = 3\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$, avec $X_i \in [-1,1]$, et $\theta \in [-1,1]$, l'espace image de $\hat{\theta}_x$ est $[-3,3]$, couvre et dépasse l'intervalle de θ. Cela indique que $\hat{\theta}_x$ peut générer des estimations au-delà des valeurs réelles possibles de θ.

Définissons la valeur ajustée $\hat{\theta}_x^{adj}$ comme suit :

$$\hat{\theta}_x^{adj} = \hat{\theta}_x \cdot I(|\hat{\theta}_x| \leq 1) + 1 \cdot I(\hat{\theta}_x > 1) - 1 \cdot I(\hat{\theta}_x < -1)$$

On découvre que, grâce à l'ajustement post-traitement, l'erreur quadratique moyenne de l'estimateur ajusté est réduite par rapport à l'estimateur initial. Cela peut être exprimé mathématiquement comme suit :

$$E\left[(\hat{\theta}_x^{adj} - \theta)^2\right] \leq E\left[(\hat{\theta}_x - \theta)^2\right],$$

ce qui indique que cette méthode réussit à améliorer les écarts quadratiques moyens.

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 La fonction de vraisemblance pour deux variables aléatoires indépendantes X_1, X_2 est :

$$L(\theta|x_1,x_2) = f(x_1|\theta)\cdot f(x_2|\theta) = \left(\frac{1}{\pi}\frac{1}{1+(x_1-\theta)^2}\right)\left(\frac{1}{\pi}\frac{1}{1+(x_2-\theta)^2}\right)$$

Introduisons le nouveau paramètre $\eta = \theta - \frac{x_1+x_2}{2}$ et $y = \frac{x_1}{2} - \frac{x_2}{2}$, ce qui implique :

En remplaçant θ par η dans la fonction de vraisemblance :

$$L(\eta) = \frac{1}{\pi^2}\cdot\frac{1}{1+(y-\eta)^2}\cdot\frac{1}{1+(y+\eta)^2} = \frac{1}{\pi^2\left[(\eta^2+1-y^2)^2+4y^2\right]}$$

Trouver la valeur minimale de $\eta^2 + 1 - y^2$ permet de déterminer la valeur maximale de $L(\eta)$. On conclut que lorsque $\eta = \pm\sqrt{y^2-1}$, $L(\theta)$ atteint sa valeur maximale. Donc on peut calculer l'extimateur de θ, $\hat{\theta}$:

$$\hat{\theta} = \pm\sqrt{\left[\frac{(x_1-x_2)^2-4}{4}\right]} + \frac{x_1+x_2}{2}$$

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 L'estimateur du maximum de vraisemblance θ_{MLE} est la valeur de θ qui maximise la fonction de vraisemblance $L(\theta) = \prod_{i=1}^{n} g(|x_i - \theta|)$, avec g étant une fonction continue et décroissante. La propriété de décroissance de g entraîne que $L(\theta)$ est maximisée lorsque θ se situe au plus proche des x_i, et en considérant l'ensemble des observations dans l'intervalle compact $[x_{(1)}, x_{(n)}]$, $L(\theta)$ atteindra nécessairement son maximum dans cet intervalle d'après le théorème des valeurs extrêmes. Par conséquent, θ_{MLE} doit exister à l'intérieur de $[x_{(1)}, x_{(n)}]$.