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

Wen, Zehai ; Li, Qingyue

2024-02-13 14:38:06-05:00

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.

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

Y_i est une fonction indicatrice pour chaque i, définie par:

$$Y_i = \begin{cases} 1 & \text{si } X_i \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

Cela indique que Y_i mesure la présence de X_i dans l'intervalle $[0,1]$.

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:

$$EQM(\hat{\theta}_y) =$$

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 :

$$\hat{\theta}_x^{adj} = \max(\min(\hat{\theta}_x, 1), -1)$$

Cette formule, par l'application d'abord du minimum entre $\hat{\theta}_x$ et 1, puis du maximum entre ce résultat et -1, garantit que quelle que soit la valeur initiale de $\hat{\theta}_x$, la valeur ajustée $\hat{\theta}_x^{adj}$ restera toujours dans l'intervalle autorisé $[-1, 1]$.

Examinons les différents cas où cette méthode d'ajustement s'avère nécessaire :

Cas 1 : Si $\hat{\theta}_x > 1$, cela signifie que la valeur estimée dépasse la limite supérieure de l'intervalle autorisé. Dans ce cas, l'ajustement est réalisé en fixant $\hat{\theta}_x^{adj}$ à 1, la limite supérieure, pour assurer que la valeur ajustée ne dépasse pas cette limite.

Cas 2 : Si $\hat{\theta}_x < -1$, la valeur estimée est inférieure à la limite inférieure de l'intervalle autorisé. L'ajustement consiste alors à définir $\hat{\theta}_x^{adj}$ à -1, garantissant ainsi que la valeur ajustée ne tombe pas en dessous de cette limite.

Cas 3 : Lorsque $\hat{\theta}_x$ se trouve déjà dans l'intervalle $[-1, 1]$, aucun ajustement n'est nécessaire, et la valeur de $\hat{\theta}_x^{adj}$ est simplement égale à $\hat{\theta}_x$, reflétant fidèlement l'estimation initiale sans modification.

En résumé, cette approche d'ajustement garantit que les estimations demeurent toujours dans les bornes définies par l'espace paramétrique, assurant ainsi la cohérence et la validité des valeurs estimées par rapport aux limites prédéfinies du modèle.

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 :

$$L(\eta) = \frac{1}{\pi^2} \cdot \frac{1}{1+(y-\eta)^2} \cdot \frac{1}{1+(y+\eta)^2}$$

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 La fonction de vraisemblance est définie par $L(\theta) = \prod_{i=1}^n g(|x_i - \theta|)$.

Premièrement, examinons la continuité de $L(\theta)$. Étant donné que g est continue et décroissante, pour tout x_i et θ, la fonction $g(|x_i - \theta|)$ est également continue par rapport à θ. Par conséquent, comme $L(\theta)$ est le produit des fonctions continues $g(|x_i - \theta|)$ pour i allant de 1 à n, $L(\theta)$ est elle-même continue par rapport à θ.

Considérer θ dans l'intervalle compact $[x_{(1)}, x_{(n)}]$ est justifié par le fait que toutes les observations x_i se situent dans cet intervalle. La décroissance de la fonction g implique que la valeur de $L(\theta)$ diminue hors de cet intervalle, rendant l'estimation de θ optimale uniquement à l'intérieur de $[x_{(1)}, x_{(n)}]$.

La continuité de $L(\theta)$ sur cet intervalle compact assure, selon le théorème des valeurs extrêmes, que $L(\theta)$ atteint son maximum dans $[x_{(1)}, x_{(n)}]$. Cela confirme l'existence de l'estimateur du maximum de vraisemblance dans cet intervalle, exploitant ainsi directement les propriétés des ensembles compacts.

En conclusion, la combinaison de la continuité de $L(\theta)$ et des propriétés de l'intervalle compact $[x_{(1)}, x_{(n)}]$ permet d'affirmer que l'estimation du maximum de vraisemblance existe et est confinée à l'intérieur de cet intervalle. Cela illustre l'importance de la continuité et des intervalles compacts dans la détermination de l'estimation du maximum de vraisemblance, tout en mettant en évidence la propriété de maximisation de la fonction de vraisemblance dans l'intervalle donné.