ΜΤΣ (MSE): $E\left[(U-g(\theta))^2\right]$

Μπορούμε να γράψουμε:

$$E\left[(U-g(\theta))^2\right] = E\left[(U-E(U)+E(U)-g(\theta))^2\right] =$$

$$= E\left[(U-E(U))^2\right] + E\left[(g(\theta)-E(U))^2\right] -$$

$$-2E\left[(U-E(U))\cdot\underbrace{(g(\theta)-E(U))}_{0}\right] =$$

$$= V(U) + \left[g(\theta)-E(U)\right]^2$$

$$= V(U) + b^2(U)$$

όπου: $b^2(U) = \left[g(\theta)-E(U)\right]^2$ μεροληψία (bias)

Συμπερασματικά: $ΜΤΣ(U) = V(U) + b^2(U)$

(*) Αν θέλουμε λοιπόν σε α.ε. τη συνθήκη η διασπορά να είναι όσο το δυνατό μικρότερη, τότε . . . .

βάσει των πιο πάνω, το ΜΤΣ γίνεται μικρό, αφού

$$E\left[(U-g(\theta))^2\right] = V(U).$$

P2

Δευτέρα, 18 Οκτωβρίου 2021 1:47 μμ

Π(χ) Έστω $X_1, X_2, \ldots, X_\nu$ τ.δ. από
$f(x; \theta_1 = \mu, \theta_2 = \sigma^2)$, με μ, σ^2 - άγνωστα.

1) ΝΔΟ η κενητική δειγ. ροπή 2ης τάξης
$$M_2 = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} (X_i - \bar{X})^2$$
δεν είναι α.ε. του σ^2.

2) Να βρεθεί το σφάλμα του
3) Να βρεθεί α.ε. του σ^2.

Έστω:

1) Δείξαμε ότι: $E\left[\sum_{i=1}^{\nu} (X_i - \bar{X})^2\right] = (\nu - 1)\sigma^2$

Άρα: $E(M_2) = E\left[\frac{1}{\nu} \cdot \sum (X_i - \bar{X})^2\right] = \frac{\nu - 1}{\nu} \sigma^2$

η οποία δεν είναι α.ε. του σ^2.

2) Σφάλμα εκτιμήτριας:
$$b(M_2) = E(M_2) - \sigma^2 = \frac{\nu - 1}{\nu} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{\nu}$$

3) Αφού: $E\left[\sum (X_i - \bar{X})^2\right] = (\nu - 1)\sigma^2 \Rightarrow$

$$\Rightarrow E\left[\frac{1}{\nu-1}\sum(x_i-\bar{x})^2\right]=\sigma^2$$

Άρα: $S^2=\frac{1}{\nu-1}\sum_{i=1}^{\nu}(x_i-\bar{x})^2$ είναι α.ε. του σ^2

ΕΠΑΡΚΕΙΑ (Sufficiency)

Αν έχουμε παρατηρήσεις $X_1, X_2, \dots, X_\nu$ από οικογένεια $\mathcal{F} = \{F(x;\theta);\ \theta \in \Theta\}$, προσπαθούμε να "συμπυκνώσουμε" την πληροφορία του δείγματος $\underline{X}$ για την παράμετρο θ σε μία σ.σ. $T = T(\underline{X})$ η οποία για το λόγο αυτό λέγεται επαρκής σ.σ.

Ορισμός: Μια σ.σ. $T = T(\underline{X})$ ονομάζεται επαρκής για την παράμετρο $\underline{\theta} \in \Theta$ όταν η πιθανότητα:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_\nu = x_\nu \mid T = t)$$ δεν εξαρτάται από το $\underline{\theta}$.

Πχ) Έστω η οικ. κατανομών Poisson $\mathcal{F} = \{F(x;\lambda),\ \lambda > 0\}$ και έστω $X_1, X_2, \dots, X_\nu$ τ.δ. από Poisson(λ). ΝΔΟ η $T = \sum_{i=1}^{\nu} X_i$ είναι επαρκής σ.σ. για το λ.

Έχουμε:

$$P(X_1 = x_1, \dots, X_\nu = x_\nu \mid T = t) = \begin{cases} \dfrac{P(X_1 = x_1, \dots, X_\nu = x_\nu, T = t)}{P(T(\underline{x}) = t)}, & \sum_{i=1}^{\nu} x_i = t \\ 0, & \text{αλλού} \end{cases}$$

Αν: $\sum_{i=1}^{\nu} X_i = t$, τότε:

$$P(X_1=x_1, \dots, X_\nu = x_\nu, T=t) = P(X_1 = x_1, \dots, X_\nu = x_\nu) =$$

$$= P(X_1=x_1)\cdot P(X_2=x_2) \cdots P(X_\nu=x_\nu)$$

Συνεπώς:

$$P(X_1=x_1, \dots, X_\nu=x_\nu \mid T=t) = \begin{cases} \dfrac{P(X_1=x_1)\cdots P(X_\nu=x_\nu)}{P(T=t)}, & \sum x_i = t \\ 0, & \text{αλλιώς} \end{cases}$$

Αφού $X_1, \dots, X_\nu$ ακολουθούν Poisson(λ), τότε:

$$T = \sum_{i=1}^{\nu} X_i \sim \text{Poisson}(\nu\lambda)$$

Οπότε:
$$\frac{P(X_1=x_1)\cdots P(X_\nu=x_\nu)}{P(T=t)} = \frac{e^{-\lambda}\frac{\lambda^{x_1}}{x_1!} \cdots e^{-\lambda}\frac{\lambda^{x_\nu}}{x_\nu!}}{e^{-\nu\lambda}\cdot\frac{(\nu\lambda)^{\sum x_i}}{\sum x_i !}} =$$

$$= \frac{\cancel{e^{-\nu\lambda}}\,\cancel{\lambda^{\sum_{i=1}^{\nu} x_i}} \Big/ \prod_{i=1}^{\nu} x_i!}{\cancel{e^{-\nu\lambda}}\,\nu^{\sum_{i=1}^{\nu} x_i}\cdot \cancel{\lambda^{\sum_{i=1}^{\nu} x_i}} \Big/ \left(\sum_{i=1}^{\nu} x_i\right)!} = \frac{\left(\sum_{i=1}^{\nu} x_i\right)!}{\nu^{\sum_{i=1}^{\nu} x_i}\cdot \prod_{i=1}^{\nu} x_i!}$$

το οποίο και δεν εξαρτάται από το λ. $\boxed{}$

ΠΑΡΑΤΗΡΗΣΗ: Οι σ.σ. $T_1 = \left(X_1, \sum_{i=2}^{\nu} X_i\right) = T_1(\underline{X})$,
$T_2 = \left(X_1, X_2, \sum_{i=3}^{\nu} X_i\right), \dots$ είναι επίσης επαρκείς σ.σ. για το λ.

$$P(X_1=x_1, \dots, X_\nu=x_\nu, T_1 = \underline{t}_1) \quad \underline{t}_1 = (x_1, t_0)$$

σ.σ. για το λ.

Έστω:

$$P(X_1=x_1,\dots,X_\nu=x_\nu \mid T_1(\underline{X})=\underline{t}_1)=\begin{cases}\dfrac{P(X_1=x_1,\dots,X_\nu=x_\nu,\ \underline{T}_1=\underline{t}_1)}{P(\underline{T}_1=\underline{t}_1)}, & \underline{T}_1=\underline{t}_1=\left(x_1,\sum_{i=2}^{\nu}X_i\right)=(x_1,t_0)\\[2ex] 0, & \text{αλλού}\end{cases}$$

Έχουμε:

$$P(X_1=x_1,\dots,X_\nu=x_\nu \mid T_1=t_1)=\frac{\cancel{P(X_1=x_1)}\cdot\ldots\cdot P(X_\nu=x_\nu)}{\cancel{P(X_1=x_1)}\cdot P\left(\sum_{i=2}^{\nu}X_i=t_0\right)}$$

$$=\frac{\cancel{e^{-(\nu-1)\lambda}}\cdot\dfrac{\cancel{\lambda^{\sum_{i=2}^{\nu}x_i}}}{\prod_{i=2}^{\nu}x_i!}}{\cancel{e^{-(\nu-1)\lambda}}\,\cancel{\left((\nu-1)\lambda\right)^{\sum_{i=2}^{\nu}x_i}}\Big/\left(\sum_{i=2}^{\nu}x_i\right)!}=\frac{\left(\sum_{i=2}^{\nu}x_i\right)!}{(\nu-1)^{\sum_{i=2}^{\nu}x_i}\cdot\prod_{i=2}^{\nu}x_i!}$$

επαρκής, αφού δεν εξαρτάται από το λ.

Συνεπώς, αναζητούμε την ελάχιστη επαρκή σ.σ. (minimal sufficient statistic).

P4

Παρ) Έστω $X_1, X_2, \ldots, X_\nu$ τ.δ. από Bernoulli(θ)

ΝΔΟ η $T = \sum_{i=1}^{\nu} X_i$ είναι επαρκής για το θ.

Επίσης: $T = \sum_{i=1}^{\nu} X_i \sim Bin(\nu, \theta)$.

Οπότε:

$$P(X_1 = x_1, \ldots, X_\nu = x_\nu \mid T = t) = \begin{cases} \dfrac{P(X_1 = x_1) \cdots P(X_\nu = x_\nu)}{P(T = t)}, & \sum x_i = t \\ 0, & \text{αλλιώς} \end{cases}$$

και:

$$\frac{P(X_1 = x_1) \cdots P(X_\nu = x_\nu)}{P(T = t)} = \frac{\theta^{x_1} \cdot (1-\theta)^{1-x_1} \cdots \theta^{x_\nu} \cdot (1-\theta)^{1-x_\nu}}{\binom{\nu}{t} \theta^t \cdot (1-\theta)^{\nu-t}} =$$

$$= \frac{\theta^t \cdot (1-\theta)^{\nu-t}}{\binom{\nu}{t} \theta^t \cdot (1-\theta)^{\nu-t}} = \binom{\nu}{t}^{-1}$$

δεν εξαρτάται από το θ.

Πχ) Έστω σ.σ. που <u>δεν</u> είναι επαρκής.

Έστω $X_1, X_2, X_3 \sim Bern(\theta)$

Είναι η $T = T(\underline{X}) = X_1 + 2X_2 + X_3$ επαρκής σ.σ. για το θ?

Έχουμε:

$$P(X_1=x_1, X_2=x_2, X_3=x_3 \mid T=t) =$$

$$= \frac{P(X_1=x_1) \cdot P(X_2=x_2) \cdot P(X_3=x_3)}{P(T=t)}$$

Έστω συγκεκριμένες τιμές: $X_1=1,\ X_2=0,\ X_3=1$

Οπότε: $T = X_1 + 2X_2 + X_3 = 1 + 0 + 1 = 2$

$$\frac{P(X_1=1) \cdot P(X_2=0)\, P(X_3=1)}{P(T=2)} =$$

$$= \frac{P(X_1=1) \cdot P(X_2=0) \cdot P(X_3=1)}{P(X_1=1) \cdot P(X_2=0) \cdot P(X_3=1) + P(X_1=0) \cdot P(X_2=1) \cdot P(X_3=0)} =$$

$$\frac{\theta \cdot (1-\theta) \cdot \theta}{\qquad} = \frac{\theta^2 \cdot \cancel{(1-\theta)}}{\qquad} = /$$

$$= \frac{\theta\cdot(1-\theta)\cdot\theta}{\theta\cdot(1-\theta)\cdot\theta+(1-\theta)\cdot\theta\,(1-\theta)} = \frac{\theta^{2}\cdot\cancel{(1-\theta)}}{\theta^{\cancel{2}}\cdot\cancel{(1-\theta)}+\cancel{\theta}\cdot(1-\theta)^{\cancel{2}}}$$

$$= \frac{\theta}{\theta+(1-\theta)} = \theta$$ εξαρτάται από το θ.

ΠΑΡΑΤΗΡΗΣΕΙΣ

→ Ο ορισμός της επάρκειας δεν βοηθάει στην επιλογή του T.

→ Διαλέγουμε ένα τυχαίο T και ελέγχουμε την επάρκεια με τον ορισμό.

  → Υπάρχουν δυσκολίες!

→ Εύκολος τρόπος είναι το παραγοντικό κριτήριο του Neyman.