Διάλεξη 4η
22/03/2019

Στατιστική Συμπερασματολογία

(*) Τετάρτη 27/03: αναπλήρωση 18:00 - 21:00

Πρόταση: Έστω οι συνήθεις κανονικότητας ισχύουν για τον αμερόληπτο εκτιμητή $U = U(\underline{X})$ της $g(\theta)$, $\theta \in \Theta$.
Τότε

$$Var(U) = \underset{\substack{\uparrow \\ \text{κάτω φράγμα}}}{K\Phi}_{C\text{-}R} = \frac{[g'(\theta)]^2}{I_{\underline{X}}(\theta)} = \frac{[g'(\theta)]^2}{n \cdot I_X(\theta)} \quad \underline{\text{εαν-ν}}$$

$$U = g(\theta) + \alpha(\theta) W, \quad \text{όπου} \quad W = \frac{\partial}{\partial \theta} \log f(\underline{x}, \theta), \ \forall \theta \in \Theta$$

Απόδ.

(⇐) Έστω $U = g(\theta) + \alpha(\theta) W$ (συνδέονται γραμμικά)

$$\rho(U, W) = \pm 1 \Rightarrow \rho^2(U, W) = 1.$$

$$\Rightarrow Cov^2(U, W) = Cov^2\left(U, \frac{\partial}{\partial \theta} \log f(\underline{x}, \theta)\right)$$

$$= Var(U)\, Var\left(\frac{\partial}{\partial \theta} \log f(\underline{x}, \theta)\right)$$

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}}$$

$\rho(X, Y) = \pm 1$ αν-ν $Y = \alpha X + \beta$

$$\Rightarrow Var(U) = \frac{Cov^2(U, W)}{Var\left(\frac{\partial}{\partial \theta} \log f(\underline{x}, \theta)\right)} \ \underset{\text{ανισ. C-R}}{\overset{\text{απόδειξη}}{=}} \ \frac{[g'(\theta)]^2}{I_{\underline{X}}(\theta)}$$

(⇒) ακολουθούμε αντίστροφη πορεία.

Παρατηρήσεις ① Στην ανισότητα του Cramer-Rao το ίσον επιτυγχάνεται $\underline{\text{αν-ν}}$

$$\frac{\partial}{\partial \theta} \log f(\underline{x}, \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x_i, \theta)$$

$$= K(\theta, n)\,[U - g(\theta)] \quad (*)$$

$$f(\underline{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Άρα αν καταφέρω να γράψω το $\frac{\partial}{\partial \theta} \log(\underline{x}, \theta)$ στην μορφή (*) U ΑΟΕΔ της $g(\theta)$

Παράδειγμα: Έστω τ.δ. $X_1, \dots, X_n$ από $N(\mu, \theta)$

μ γνωστό και $\theta > 0$: άγνωστο. Να βρεθεί ΑΟΕΔ της

$g(\theta) = \theta$.

Λύση

Θυμίζω: $f(x) = f(x_i) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\mu)^2}$

Ανισότητα C-R:

Βήμα 1: $K\Phi_{C-R} = \frac{[g'(\theta)]^2}{n I_X(\theta)} = \frac{1}{n I_X(\theta)}$,

$$I_X(\theta) = E\left(\frac{\partial}{\partial\theta} \log f(x,\theta)\right)^2 = \underbrace{-E\left(\frac{\partial^2}{\partial\theta^2} \log f(X,\theta)\right)}_{\text{πιο εύκολος στην χρήση}}$$

$$\frac{\partial^2}{\partial\theta^2} \log f(x,\theta) = \frac{1}{2\theta^2} - \frac{(x-\mu)^2}{\theta^3}$$

$$I_X(\theta) = -\frac{1}{2\theta^2} + \frac{1}{\theta^3} E(X-\mu)^2 = -\frac{1}{2\theta^2} + \frac{1}{\theta^3} Var(X)$$

$$= -\frac{1}{2\theta^2} + \frac{1}{\theta^3} \cdot \theta = \frac{1}{2\theta^2}.$$

$$K\Phi_{C-R} \quad \frac{1}{n\left(1/2\theta^2\right)} \Rightarrow \boxed{K\Phi_{C-R} = \frac{2\theta^2}{n}}$$

Βήμα 2: Προσπαθούμε να βρούμε στατιστική συνάρτηση (6.6.)

$T = T(X_1, \dots, X_n)$ τ.ω. $\begin{cases} E(T) = \theta \\ Var(T) = K\Phi_{C-R}. \end{cases}$

Προσοχή! Ξεκινώ από το $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$.

Ξέρω ότι $E(S^2) = \theta$.

$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$

αντί για $\bar{X}$ έχω μ που το ξέρω

③

Δοκιμάζω το $T = \sum (X_i - \mu)^2$.

↑ ΔΕΝ ΙΣΧΥΕΙ ΠΑΝΤΑ. ΜΠΟΡΕΙ ΝΑ ΥΠΑΡΞΟΥΝ ΦΟΡΕΣ ΠΟΥ ΔΕΝ ΘΑ ΒΟΗΘΑΕΙ.

$E(T) = E\left(\sum (X_i - \mu)^2\right) = \sum E(X_i - \mu)^2$

$= \sum Var(X_i) = \sum \theta = n\theta$.

(*) Δεν είναι αμερόληπτος όμως αν διαιρέσω με n τότε $E\left(\frac{T}{n}\right) = \theta$.

Άρα θεωρώ $T = \frac{1}{n} \sum (X_i - \mu)^2$.

και $E(T) = \theta$, δηλ. αμερόληπτος.

$Var(T) = Var\left(\frac{1}{n} \sum (X_i - \mu)^2\right)$

$= \frac{1}{n^2} \sum Var(X_i - \mu)^2$. Θέλω το $Var(X_i - \mu)^2$

$\Rightarrow Var(X_i - \mu)^2 = 2\theta^2$.

Είναι $X_i \sim N(0,1)$

Άρα $\frac{X_i - \mu}{\sqrt{\theta}} \sim N(0,1)$, $\frac{(X_i - \mu)^2}{\theta} \sim N(0,1)^2 \equiv X_1^2$.

$\Rightarrow Var\left(\frac{(X_i - \mu)^2}{\theta}\right) = Var(X_1^2) = 2 \cdot 1 = 2$.

$\frac{1}{\theta^2} Var(X_i - \mu)^2 = 2$.

$Var(T) = \frac{1}{n^2} \sum (X_i - \mu)^2 = \frac{1}{n^2} \sum 2\theta^2 = \frac{1}{n^2} \cdot 2n\theta^2 = \frac{2\theta^2}{n}$

$\Rightarrow Var(T) = \frac{2\theta^2}{n} = K\Phi_{C-R}$.

Άρα $T = \frac{1}{n} \sum (X_i - \mu)^2$ ΑΟΕΔ της θ.

<u>Ανάλυση της παρατήρησης ①:</u>

$\frac{\partial}{\partial \theta} \log f(\underline{x}, \theta) = \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} f(x_i, \theta)$

$= \frac{\partial}{\partial \theta} \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x_i - \mu)^2}$

$$= \frac{\partial}{\partial\theta} \log\left(\frac{1}{(2\pi\theta)} e^{-\frac{1}{2\theta}\sum(x_i-\mu)^2}\right)$$

$$= \frac{\partial}{\partial\theta}\left[-\frac{n}{2}\log 2\pi\theta - \frac{1}{2\theta}\sum(x_i-\mu)^2\right)$$

$$= -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum(x_i-\mu)^2$$

$$= -\frac{n\theta}{2\theta^2} + \frac{n}{2n\theta^2}\sum(x_i-\mu)^2 = \frac{n}{2\theta^2}\left[\frac{1}{n}\sum(x_i-\mu)^2-\theta\right]$$

Άρα $\frac{1}{n}\sum(x_i-\mu)^2$: ΑΟΕΔ.

Παρατήρηση ②: Η δειγματική διακύμανση

$S^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$ είναι αμερόληπτη της θ.

Ας βρω $Var(S^2)$

Αλλά $\frac{(n-1)S^2}{\theta} \sim \chi^2_{n-1}$.

$$Var\left(\frac{(n-1)S^2}{\theta}\right) = Var(\chi^2_{n-1}) = 2(n-1)$$

$$\Rightarrow \frac{(n-1)^2}{\theta^2} Var(S^2) = 2(n-1) \Rightarrow Var(S^2) = \frac{2\theta^2}{n-1}.$$

Παρατηρώ $Var(S^2) = \frac{2\theta^2}{n-1} > \frac{2\theta^2}{n} = Var(T)$

αναμενόμενο γιατί

T : ΑΟΕΔ και ο ΑΟΕΔ είναι

βασικός

Ερώτημα: Πόσο αποτελεσματικός είναι ο S^2 σε σχέση με τον T;

Απάντηση:

Η έννοια της σχετικής αποτελεσματικότητας:

Αποτελεσματικός εκτιμητής:

Ορισμός: Ο εκτιμητής $U = U(\underline{X})$ της $g(\theta)$ λέγεται αποτελεσματικός αν ο U είναι ο ΑΟΕΔ της $g(\theta)$.

Σχετική Αποτελεσματικότητα: Έστω U αποτελεσματικός εκτιμητής της $g(\theta)$ και έστω $T = T(X)$ αμερόληπτος της $g(\theta)$.
Η σχετική αποτελεσματικότητα του T σε σχέση με τον U ορίζεται από το πηλίκο $\dfrac{Var(U)}{Var(T)}$.

Παρατήρηση ③:

$$0 < \frac{Var(U)}{Var(T)} \leq 1 \quad \{= 1 \text{ αν } T \equiv U\}$$

Παράδειγμα: Έστω τ.δ. X_1, X_2, X_3 από Poisson(θ). Νδο η $U = \bar{X}$ είναι αποτελεσματικός εκτιμητής της θ και να βρεθεί η σχετική αποτελεσματικότητα του $T = \dfrac{X_1 + 2X_2 + 3X_3}{6}$ ως προς U.

Λύση

(*) Για Poisson κατανομή ο $\bar{X}$ είναι ΑΟΕΔ.
Από προηγ. άσκηση ο $\bar{X}$: ΑΟΕΔ. Άρα ο $U = \bar{X}$ είναι αποτελεσματικός.
Είναι ο T αμερόληπτος;

$$E(T) = \frac{1}{6}E(X_1) + \frac{2}{6}E(X_2) + \frac{3}{6}E(X_3) = \theta.$$

$$Var(U) = Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{9}\left(Var(X_1) + Var(X_2) + Var(X_3)\right) = \frac{3\theta}{9} = \frac{\theta}{3}$$

$$Var(T) = Var\left(\frac{1}{6}X_1 + \frac{2}{6}X_2 + \frac{3}{6}X_3\right) = \frac{1}{36}\theta + \frac{4}{36}\theta + \frac{9}{36}\theta = \frac{14}{36}\theta.$$

Εκθετική οικογένεια κατανομών (ΕΟΚ)

Ορισμός: Η κατανομή του τ.δ. $\underline{X} = (X_1, \ldots, X_n), n \geq 1$ ανήκει στην μονοπαραμετρική ΕΟΚ αν για τη σ.π.π. του $\underline{X}$ ισχύουν:

α) Το $S = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : f(\underline{x}, \theta) > 0\}$ ανεξάρτητο της $\theta \in \Theta \subseteq \mathbb{R}$.

β) Αν η $f(\underline{x}, \theta)$ έχει τη μορφή

$$f(\underline{x}, \theta) = e^{A(\theta) + B(\underline{x}) + C(\theta) D(\underline{x})}, \quad \underline{x} \in S_n, \ \theta \in \Theta \subseteq \mathbb{R}.$$

Παρατήρηση (1): i) Αν X είναι τ.μ. η μονοπαραμετρική ΕΟΚ ορίζεται: $f(x, \theta) = e^{A^*(\theta) + B^*(x) + C^*(\theta) D^*(x)}$, $x \in S_1 = \{x : f(x, \theta) > 0\}$ και το S_1 ανεξ. της θ. (*)

ii) Η $N(\theta, \sigma^2)$, σ^2: γνωστό, $N(\mu, \theta)$ μ: γνωστό, η $B(n, \theta)$, η $P(\theta)$, η Εκθ(θ) ανήκουν στην ΕΟΚ. Διότι οι κατανομές αυτές μπορούν να γραφούν στην μορφή (*).

Η $U(0, \theta)$ (ομοιόμορφη) $\nexists$ στην ΕΟΚ, διότι το S_1 εξαρτάται από το θ.

Πρόταση: Η ΕΟΚ ικανοποιεί τις συνθήκες κανονικότητας της ανισότητας Cramer-Rao αν η $C(\theta)$ ή η $C^*(\theta)$ έχει συνεχή και μη μηδενική παράγωγο $\forall \theta \in \Theta$ (δεν θα ασχοληθούμε παραπάνω μ' αυτό)

↑ΠΑΝΤΑ ΘΑ ΕΧΕΙ ΣΥΝΘΗΚΕΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ

Παράδειγμα 1: Έστω τ.δ $X_1, \ldots, X_n$ από διωνυμική $B(m,\theta)$, $0<\theta<1$. Να βρεθεί επαρκή σ.σ. για την θ.

Λύση

$$B(m,\theta) \rightsquigarrow p(x,\theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x}, \quad 0<\theta<1, \quad x=0,\ldots,m$$

$$f(\underline{x},\theta) = \prod_{i=1}^{n} p(x_i,\theta) = \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i}$$

$$= \left(\prod_{i=1}^{n} \binom{m}{x_i}\right) \theta^{\sum x_i} (1-\theta)^{n\cdot m - \sum x_i}$$

$$= \theta^{\sum x_i} (1-\theta)^{nm} (1-\theta)^{-\sum x_i} \left(\prod_{i=1}^{n} \binom{m}{x_i}\right)$$

$$= \underbrace{(1-\theta)^{nm} \left(\frac{\theta}{1-\theta}\right)^{\sum x_i}}_{g(T(\underline{x}),\theta)} \underbrace{\left(\prod_{i=1}^{n} \binom{m}{x_i}\right)}_{h(\underline{x})}$$

Άρα $T(\underline{x}) = \sum_{i=1}^{n} x_i$ επαρκές (από παραγοντικό θεώρ. Neyman-Fisher)

Παράδειγμα 2: Έστω τ.δ. $X_1, \ldots, X_n$ από Poisson(θ), $\theta>0$

$$p(x,\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad \theta>0, \quad x=0,1,\ldots$$

$$f(\underline{x},\theta) = \prod_{i=1}^{n} p(x_i,\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!} = \underbrace{e^{-n\theta}\theta^{\sum x_i}}_{g(T(\underline{x}),\theta)} \underbrace{\left(\prod_{i=1}^{n} x_i!\right)^{-1}}_{h(\underline{x})}$$

$\Rightarrow$ Το $T(\underline{x}) = \sum_{i=1}^{n} x_i$: επαρκές //.

Παράδειγμα 3: Έστω τ.δ $X_1, \ldots, X_n$ από κανονική κατανομή $N(\theta, \sigma^2)$, σ^2: γνωστό. Να βρεθεί επαρκής για την θ.

Λύση

$$f(x,\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

$$f(\underline{x},\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (x_i-\theta)^2}$$

$$=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i^2-2\theta x_i+\theta^2)}$$

$$=\underbrace{e^{\frac{\theta}{\sigma^2}\sum_{i=1}^{n}x_i-n\frac{\theta^2}{2\sigma^2}}}_{g(T(\underline{x}),\theta)}\underbrace{\left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\cdot e^{-\frac{1}{2\sigma^2}\cdot\sum x_i^2}\right)}_{h(\underline{x})}$$

$\uparrow$

$\sum x_i$

$\Rightarrow$ Το $T(\underline{x})=\sum x_i$: επαρκής.

<u>Παράδειγμα</u>: Έστω τ.δ. $x_1,\dots,x_n$ από κανονική κατανομή $N(b,\theta)$, b: γνωστό. Να βρεθεί επαρκής για την θ.

<u>Λύση</u>

$$f(x,\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-b)^2}{2\theta}}\quad,-\infty<x<\infty,\ b\in\mathbb{R},\ \theta>0$$

$$f(\underline{x},\theta)=\prod_{i=1}^{n}f(x_i,\theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x_i-b)^2}{2\theta}}$$

$$=\underbrace{\left(\frac{1}{2\pi\theta}\right)^{n/2}e^{-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-b)^2}}_{g(T(\underline{x}),\theta)}\cdot\underbrace{1}_{h(\underline{x})}$$

Άρα $T(\underline{x})=\sum_{i=1}^{n}(x_i-b)^2$: επαρκής.