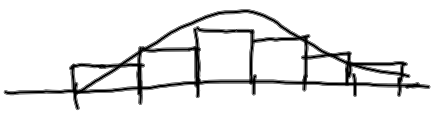
17ο 6/12/2017

Στατιστική.

Τυχαίο Δείγμα: ν ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, \ldots, X_\nu \sim F(x)$

ν: μέγεθος δείγματος.

π.χ. $X_1 = x_1, \ldots, X_\nu = x_\nu$ είναι ν μετρήσεις



Δεκ 6-5:04 μμ

<u>Πρόβλημα:</u> Με βάση τις μετρήσεις $x_1, \dots, x_n$ να εξάγω συμπεράσματα για την <u>άγνωστη σ.κ. F.</u>

(A) Μη παραμετρική Στατιστική: Η F είναι απολύτως άγνωστη.

(B) Παραμετρική Στατιστική: Η F είναι άγνωστη αλλά ανήκει σε μία γνωστή παραμετρική οικογένεια κατανομών.

<u>Παράδειγμα (Β)</u>: Τα δεδομένα προέρχονται από κάποια κατανομή Poisson με λ άγνωστο.

ή προέρχονται από $b(p)$, $p \in (0,1)$, άγνωστο.

Δεκ 6-5:24 μμ

Δεδομένα από $\exp(\lambda)$, $\lambda>0$ άγνωστο

Δεδομένα από $\Gamma(\alpha,\lambda)$, $\alpha>0, \lambda>0$ άγνωστα

Δεδομένα από $N(\mu,\sigma^2)$, μ, σ^2 άγνωστα (φυσικά $\sigma^2>0$).

Γενικό πλαίσιο:

$$X_1,\dots,X_\nu \sim f(x|\theta), \quad \theta \text{ άγνωστο}$$

ή για δύο παραμέτρους

$$X_1,\dots,X_\nu \sim f(x|\theta_1,\theta_2), \quad \theta_1,\theta_2 \text{ άγνωστα.}$$

Θέλουμε να "εκτιμήσουμε" την άγνωστη παράμετρο θ (ή τις άγνωστες παραμέτρους).

<u>Ορισμός:</u> Μία συνάρτηση $T = T(X_1, .., X_\nu)$ (που δεν περιέχει άγνωστες παραμέτρους) καλείται εκτιμήτρια.

π.χ. $T(X_1, X_2, .., X_\nu) = \frac{X_1 - X_2}{2}$

$T = \bar{X} = \frac{X_1 + \dots + X_\nu}{\nu}$ δειγμ. μέσος

$T = 2$

$T = \sum_{i=1}^{\nu} (X_i - \bar{X})^2 / (\nu - 1)$ δειγμ. διασπορά

όχι $T = \frac{\Sigma X_i}{\nu \cdot \theta}$ ΟΧΙ ΕΚΤΙΜΗΤΡΙΑ

ή $T = X_1 e^{\theta}$ όχι

Δεκ 6-5:36 μμ

Κριτήριο Αμεροληψίας. Η T καλείται αμερόληπτη για την άγνωστη παράμετρο θ αν $E(T) = \theta \quad \forall \theta$.

Παράδειγμα: $X_1, \dots, X_\nu \sim b(\theta)$

$$T = \bar{X} \qquad E(T) = E\left(\frac{X_1 + \dots + X_\nu}{\nu}\right) = \frac{1}{\nu} E(X_1 + \dots + X_\nu)$$

$$= \frac{1}{\nu}\left(E(X_1) + \dots + E(X_\nu)\right) = \frac{1}{\nu} \cdot (\nu \cdot \theta) = \theta$$

$$T_1 = X_1 - X_2 + X_3 \qquad E(T_1) = E(X_1) - E(X_2) + E(X_3) = \theta - \theta + \theta = \theta$$

$$T_2 = X_1 \qquad E(T_2) = \theta$$

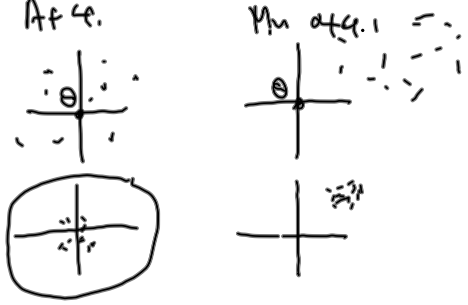
$T_3 = X_1 + X_2 + X_3$ δεν είναι αμερόληπτη αφού

$$E(T_3) = 3\theta \neq \theta$$

$$T_4 = X_1 \cdot X_2 \qquad E(T_4) = \theta^2 \qquad \cancel{T_5 = \theta} \qquad T_6 = \frac{1}{2}$$

Κριτήριο Ελάχιστης Διασποράς: Αν T_1, T_2 είναι δύο αμερόληπτες εκτιμήτριες για το θ, τότε προτιμάμε την T_1 σε σχέση με την T_2 όταν

$$Var(T_1) \le Var(T_2) \quad \forall \theta.$$



Μεγ. Διασπορά

Μικρή Διασπορά

$T = \overline{X}$ αμερόληπτη για το θ στην $b(\theta)$

και

$T_1 = X_1 - X_2 + X_3$ αμερόληπτη

$$Var(T) = Var\left(\frac{1}{\nu} \sum_{i=1}^{\nu} X_i\right) = \left(\frac{1}{\nu}\right)^2 \sum_{i=1}^{\nu} Var(X_i) = \frac{\nu\,\theta(1-\theta)}{\nu^2}$$

$$= \frac{\theta(1-\theta)}{\nu}$$

$$Var(T_1) = Var(X_1 - X_2 + X_3) = Var(X_1) + Var(-X_2) + Var(X_3)$$

$$= \theta(1-\theta) + (-1)^2\theta(1-\theta) + \theta(1-\theta)$$

$$= 3\theta(1-\theta) \geq \frac{\theta(1-\theta)}{\nu}$$

Αφού $Var(\overline{X}) \leq Var(T_1)$, προτιμώ την $T = \overline{X}$.

Ορισμός: Μέσο τετραγωνικό σφάλμα μιας εκτιμήτριας T για το θ καλείται η ποσότητα

$$MSE(T) = E\left[(T-\theta)^2\right].$$

Η T_1 είναι καλύτερη από την T_2 αν

$$MSE(T_1) \leq MSE(T_2) \quad \forall\ \theta.$$

<u>Παρατήρηση:</u> Στη γενική περίπτωση που η T δεν είναι αμερόληπτη θα ισχύει $E(T)=\mu=\mu(\theta)\neq\theta$ για κάποια θ. Τότε

$$MSE(T) = E\left[(T-\theta)^2\right] = E\left[(T-\mu+\mu-\theta)^2\right] = E(T-\mu)^2 + (\mu-\theta)^2$$
$$+\,2(\mu-\theta)\,E(\cancelto{0}{T-\mu}) = Var(T) + b(\theta)^2 \qquad b(\theta)=\mu(\theta)-\theta = E(T)-\theta$$

Ορισμός: Αν έχω μία ακολουθία εκτιμητριών T_ν (δηλ. $T_\nu = T_\nu(X_1, \dots, X_\nu)$) τότε η T_ν καλείται συνεπής εάν ισχύει το εξής:

$$\boxed{\forall \varepsilon > 0, \quad \lim_{\nu \to \infty} P(|T_\nu - \theta| \geq \varepsilon) = 0} \qquad (*)$$

Η (*) λέει $T_\nu \xrightarrow{P} \theta \quad (\nu \to \infty)$.

Θεώρημα Αν T_ν είναι ακολουθία εκτιμητριών και

(1) $E(T_\nu) \to \theta \quad (\nu \to \infty) \quad \forall \theta$

(2) $Var(T_\nu) \to 0 \quad (\nu \to \infty) \quad \forall \theta$

τότε η T_ν είναι συνεπής εκτιμήτρια του θ.

Ανισότητες Markov-Chebyshev.

Markov: Αν $X \geq 0$ και $E(X)=\mu \Rightarrow P(X \geq \alpha) \leq \frac{\mu}{\alpha}$ $\forall \alpha>0$.

Απόδ: $X = X\,I(X<\alpha) + X\,I(X\geq\alpha) \geq X\,I(X\geq\alpha) \geq \alpha\,I(X\geq\alpha)$

$X \geq \alpha\,I(X\geq\alpha)$

$\Rightarrow E(X) \geq \alpha\,E\big(I(X\geq\alpha)\big) = \alpha \cdot P(X\geq\alpha)$

(η $I(X\geq\alpha)$ είναι bernoulli)

Chebyshev: Αν X μια τ.μ. με $E(X)=\mu$, $Var(X)=\sigma^2 \geq 0$

τότε $\boxed{P(|X-\mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2} \quad \forall \alpha>0.}$

$$P(|X-\mu|\geq\alpha) = P(|X-\mu|^2 \geq \alpha^2) \overset{Markov}{\leq} \frac{E(X-\mu)^2}{\alpha^2} = \frac{\sigma^2}{\alpha^2}.$$

Απόδειξη Θεωρήματος. Υποθέτω ότι $E(T_\nu)=\theta$ και $Var(T_\nu)\to 0$. Τότε $\forall\, \varepsilon>0$,

$$P(|T_\nu-\theta|\geq\varepsilon) \overset{\text{Cheb.}}{\leq} \frac{Var(T_\nu)}{\varepsilon^2} \xrightarrow{\nu\to\infty} 0 \Rightarrow P(|T_\nu-\theta|\geq\varepsilon)\to 0.$$

<u>Θεώρημα:</u> Η ακολουθία $\overline{X}=\overline{X}_\nu=\dfrac{X_1+\dots+X_\nu}{\nu}$ είναι συνεπής εκτιμήτρια του άγνωστου μέσου μ ενός πληθυσμού. (Εφ' όσον η διασπορά είναι πεπερασμένη).

<u>Απόδ:</u> $E(\overline{X}_\nu)=\mu \quad (\forall\, \nu=1,2,\dots)$

$$Var(\overline{X}_\nu)=\left(\frac{1}{\nu}\right)^2 Var\left(\sum_{i=1}^{\nu}X_i\right)=\frac{\nu\sigma^2}{\nu^2}=\frac{\sigma^2}{\nu}\to 0 \quad (\nu\to\infty)$$

Δεκ 6-6:35 μμ

Θεώρημα: Η δειγματική διασπορά

$$S_\nu^2 = \frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i - \bar{X}_\nu)^2$$

είναι συνεπής εκτιμήτρια της άγνωστης διασποράς σ^2 του πληθυσμού.

Πρόταση: $E(S_\nu^2) = \sigma^2 \quad (\forall\ \nu = 2, 3, \ldots)$

Επίσης μπορεί να δειχθεί ότι $Var(S_\nu^2) \to 0$ (όταν $E(X_i^4) < \infty$).

$$\sum_{i=1}^{\nu}(X_i - \bar{X}_\nu)^2 = \sum_{i=1}^{\nu}\left\{X_i^2 - 2X_i\bar{X}_\nu + \bar{X}_\nu^2\right\}$$

$$= \sum_{i=1}^{\nu}X_i^2 - 2\bar{X}_\nu \overbrace{\sum_{i=1}^{\nu}X_i}^{=\nu\bar{X}_\nu} + \nu\cdot\bar{X}_\nu^2 = \sum_{i=1}^{\nu}X_i^2 - 2\nu\bar{X}_\nu^2 + \nu\bar{X}_\nu^2$$

$$\boxed{\sum_{i=1}^{\nu}(X_i-\bar{X}_\nu)^2 = \sum_{i=1}^{\nu}X_i^2 - \nu\,\bar{X}_\nu^2} \qquad (**)$$

Λαμβάνω μέσες τιμές:

$$E\left\{\sum_{i=1}^{\nu}(X_i-\bar{X}_\nu)^2\right\} = E\left\{\sum_{i=1}^{\nu}X_i^2 - \nu\,\bar{X}_\nu^2\right\}$$

$$= \sum_{i=1}^{\nu}E(X_i^2) - \nu\,E(\bar{X}_\nu^2) \overset{***}{=} \text{γρ αμ.}$$

$$E(X_i)=\mu \quad Var(X_i)=\sigma^2 \Rightarrow E(X_i^2)=\mu^2+\sigma^2 \quad (i=1,2,\dots,\nu)$$

$$E(\bar{X}_\nu^2)=\left[E(\bar{X}_\nu)\right]^2+Var(\bar{X}_\nu)=\mu^2+\frac{\sigma^2}{\nu}$$

$$\overset{***}{=} \sum_{i=1}^{\nu}(\mu^2+\sigma^2)-\nu\left(\mu^2+\frac{\sigma^2}{\nu}\right)=\nu(\mu^2+\sigma^2)-\nu\left(\mu^2+\frac{\sigma^2}{\nu}\right)$$

$$=(\nu-1)\sigma^2$$

$$E\left\{\sum_{i=1}^{\nu}(X_i-\overline{X}_\nu)^2\right\}=(\nu-1)\,\sigma^2$$

$$\frac{1}{\nu-1}\,E\left\{\sum_{i=1}^{\nu}(X_i-\overline{X}_\nu)^2\right\}=\sigma^2$$

$$\Rightarrow \boxed{S_\nu^2=\frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i-\overline{X}_\nu)^2}$$

Δεκ 6-6:54 μμ

Για άλλες μεθόδους σημειακής εκτίμησης

θα μιλήσουμε στο άλλο μάθημα.