Στατ. Σύμπερ

Παράδειγμα

Έστω τ.δ $X_1, \dots, X_n$ από $Εκθ(\lambda)$. Να κατασκευαστεί ΟΙ-τεστ

$H_0: \lambda = \lambda_0$ (λ_0 γνωστό) $H_a: \lambda > \lambda_0$

Λύση

Θεωρώ τον έλεγχο της αληθής $H_0: \lambda = \lambda_0$ έναντι $H_a: \lambda = \lambda_a$ όπου $\lambda_a > \lambda_0$

$$\frac{L_0}{L_a} \le k$$

$$L = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \Rightarrow L = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

Άρα η Ι-κ.π είναι $\frac{L_0}{L_a} \le k \Rightarrow \frac{\lambda_0^n e^{-\lambda_0 \sum_{i=1}^{n} x_i}}{\lambda_a^n e^{-\lambda_a \sum_{i=1}^{n} x_i}} \le k \Rightarrow$

$$\left(\frac{\lambda_0}{\lambda_a}\right)^n e^{-(\lambda_0 - \lambda_a) \sum_{i=1}^{n} x_i} \le k \Rightarrow e^{-(\lambda_0 - \lambda_a) \sum_{i=1}^{n} x_i} \le \left(\frac{\lambda_a}{\lambda_0}\right)^n k \Rightarrow$$

$$(\lambda_a - \lambda_0) \sum_{i=1}^{n} x_i \le \log\left(\frac{\lambda_a}{\lambda_0}\right)^n k \Rightarrow \sum_{i=1}^{n} x_i \le \frac{1}{\lambda_a - \lambda_0} \log\left(\frac{\lambda_a}{\lambda_0}\right)^n k = k'$$

Άρα η Ι-κ.π είναι $\sum_{i=1}^{n} x_i \le k'$

Υπολογισμός κ.σ k'

$$a = P(\text{Απορ } H_0 \mid H_0 \text{ αληθής}) = P\left(\sum_{i=1}^{n} x_i \le k' \mid X_1, \dots X_n \sim Εκθ(\lambda_0)\right)$$

Άρα θέλω να βρω την κατανομή του $\sum_{i=1}^{n} X_i$, $X_i \sim Εκθ(\lambda_0)$

Αφού $X_i \sim Εκθ(\lambda_0)$ $\forall i=1,\dots,n$ και αφού $Εκθ(\lambda_0) \equiv G(1, \frac{1}{\lambda_0})$

Άρα $X_i \sim G(1, \frac{1}{\lambda_0})$ $\forall i=1,\dots,n$

Με την μέθοδο της ροπογεννήτριας αν $X_1,\dots,X_n$ τ.δ από $G(a,b)$ το $\sum_{i=1}^{n} X_i \sim G(na, b)$

Συνολικά αν $X_i \sim Εκθ(\lambda_0)$ $\forall i=1,\dots,n$. τότε $\sum_{i=1}^{n} X_i \sim G(n, \frac{1}{\lambda_0})$

με σ.π.π $\frac{\lambda_0^n}{\Gamma(n)} x^{n-1} e^{-\lambda_0 x}$

$$a = P\left(\sum_{i=1}^{n} X_i \leq k' \,/\, \sum_{i=1}^{n} X_i \sim G\left(n, \frac{1}{\lambda_0}\right)\right) = \int_0^{k'} \frac{\lambda_0^n}{\Gamma(n)} x^{n-1} e^{-\lambda_0 x}\, dx \quad (*)$$

Το ολοκλήρωμα δεν υπολογίζεται σε αναλυτική μορφή

Αριθμητικά λύνεις το ολοκλήρωμα και από την εξίσωση προσδιορίζεται το κρίσιμο σημείο k'

Συγκεντρωτικά: Το Ι-τεστ για τον έλεγχο της $H_0: \lambda = \lambda_0$ έναντι $H_a: \lambda = \lambda_a$ έχει Σ.Σ.Τ $\sum_{i=1}^{n} X_i$ με κατανομή $G(n, \frac{1}{\lambda_0})$ υπό την H_0 και κ.π. $\sum X_i \leq k'$ με k' να υπολογίζεται από την $(*)$

Παρατηρώ ότι το Ι-τεστ δεν εξαρτάται από το λ_a.

Άρα αυτό είναι και το ΟΙ για τον έλεγχο της $H_0: \lambda=\lambda_0$ $\quad H_a: \lambda>\lambda_0$

Β' τρόπος

Θεωρώ $T=2\lambda_0 \sum_{i=1}^{n} X_i$ $\quad X_i \sim Εκθ(\lambda_0) \equiv G\left(1, \frac{1}{\lambda_0}\right)$ και $\sum_{i=1}^{n} X_i \sim G\left(n, \frac{1}{\lambda_0}\right)$

Αν $Y = \sum_{i=1}^{n} X_i$ τότε $Y \sim G\left(n, \frac{1}{\lambda_0}\right)$

$$f_Y(y) = \frac{\lambda_0}{\Gamma(n)} y^{n-1} e^{-\lambda_0 y}, \quad y>0$$

Με αλλαγή μεταβλητών η κατανομή του $T=2\lambda_0 Y$ είναι

$$f_T(t) = \frac{1}{2^n \Gamma(n)} t^{n-1} e^{-t/2}, \quad t>0$$

Άρα $T = 2\lambda_0 Y = 2\lambda_0 \sum_{i=1}^{n} X_i \sim G(n, 2) \equiv \chi^2_{2n}$

$$a = P\left(\sum X_i \le k' / X_i \sim Εκθ(\lambda_0)\right) = P\left(2\lambda_0 \sum X_i \le 2\lambda_0 k' = k'' / X_i \sim Εκθ(\lambda_0)\right) =$$

$$= P\left(T \le k'' / T \sim \chi^2_{2n}\right) = P\left(\chi^2_{2n} \le k''\right) \Rightarrow$$

$$\Rightarrow a = 1 - P\left(\chi^2_{2n} \ge k''\right) \Rightarrow P\left(\chi^2_{2n} \ge k''\right) = 1-a \xrightarrow{\text{(ορισμός εκατοστιαίων σημείων } \chi^2_{2n})}$$

$$k'' \equiv \chi^2_{2n, 1-a}$$

Συγκεντρωτικά το Ι-τεστ για έλεγχο της $H_0: \lambda=\lambda_0$, $H_a: \lambda=\lambda_a$ έχει Σ.Σ.Τ την $2\lambda_0 \sum X_i \sim \chi^2_{2n}$ υπό H_0 και

Κ.Π. $2\lambda_0 \sum X_i \le \chi^2_{2n, 1-a}$

Μέχρι εδώ για τους μαθηματικούς

Μεθοδολογία Τεστ πηλίκου μέγιστων Πιθανοφανειών (ΤΠΜΠ)

Η θεωρία Neymann-Pearson δίνει άριστα τεστ από άποψη ισχύος είναι όμως περιορισμένης εμβέλειας

H_0 : απλή έναντι H_a : απλής $\longrightarrow$ I-τεστ

ή

H_0 : απλή έναντι H_a : σύνθετη $\longrightarrow$ ΟΙ-τεστ

Παράδειγμα

Έστω τ.δ $X_1, \dots, X_n$ από $N(\mu, \sigma^2)$ σ^2 άγνωστη. Θέλω να ελέγξω

$H_0 : \mu = \mu_0$ έναντι $H_a : \mu > \mu_0$ (μ_0 γνωστή)

(αφού σ^2 άγνωστη δεν προσδιορίζεται η κατανομή $\Rightarrow$ όχι απλή)

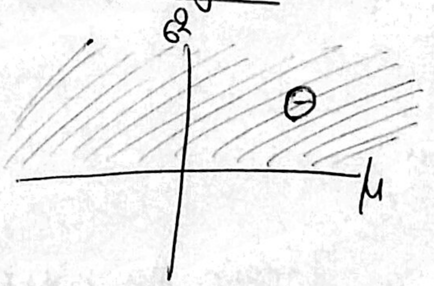
Γενικός φορμαλισμός

έστω τ.δ $X_1, \dots, X_n$ από $f(x,\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$.

Έστω $H_0 : \theta \in \Theta_0 \subseteq \Theta$ έναντι $H_a : \theta \in \Theta_a \subseteq \Theta$

Σύνδεση γενικού φορμαλισμού με το παράδειγμα

$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < +\infty,\ \sigma^2 > 0\}$



$H_0 : \mu = \mu_0$ έναντι $H_a : \mu > \mu_0$ (μ_0 γνωστό

$\Downarrow$ $\qquad\qquad\qquad\qquad$ $\Downarrow$

$H_0 : \theta \in \Theta_0 \subseteq \Theta$ $\qquad$ $H_a : \theta \in \Theta_a$

$\Theta_0 = \{(\mu_0, \sigma^2) : \mu_0 \text{ γνωστό}, \sigma^2 > 0\}$

$\Theta_a = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$

Εφαρμογή μέθοδος Πηλίκου Μέγιστων Πιθανοφανειών

1) Στηρίζεται στο λόγο (πηλίκο) $\lambda = \dfrac{\sup_{\Theta_0} L(\theta)}{\sup_{\theta} L(\theta)} \xrightarrow[\text{ΕΜΠ}]{\text{από ορισμό}} \dfrac{L(\hat{\theta}_0)}{L(\hat{\theta})}$

όπου $\hat{\theta}_0$ ο ΕΜΠ της θ υπό την H_0 (ΕΜΠ της θ στον Θ_0)

ενώ $\hat{\theta}$ ο ΕΜΠ της θ στον πλήρη παραμετρικό χώρο Θ

2) Μορφή κ.π

Μικρές τιμές λ συνηγορούν απόρριψη H_0 (Άρα κ.π $\lambda \leq k$)

$$\frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{P_{\hat{\theta}_0}(x_1-\varepsilon_1 \leq X_1 \leq x_1+\varepsilon_1, \ldots, x_n-\varepsilon_n \leq X_n \leq x_n+\varepsilon_n)}{P_{\hat{\theta}}(x_1-\varepsilon_1 \leq X_1 \leq x_1+\varepsilon_1, \ldots, x_n-\varepsilon_n \leq X_n \leq x_n+\varepsilon_n)}$$

$\lambda \leq k$ αν $P_{\hat{\theta}_0} \ll P_{\hat{\theta}}$

Παράδειγμα (t-τεστ)

Έστω $X_1, \ldots, X_n$ τ.δ από $N(\mu, \sigma^2)$, $\sigma^2 > 0$ άγνωστο. Να κατασκευαστεί ΤΠΜΠ $H_0 : \mu = \mu_0$ έναντι $H_a : \mu \neq \mu_0$ (μ_0 γνωστό).

Λύση

$$\lambda = \frac{\sup_{\mu=\mu_0, \sigma^2>0} L(\mu, \sigma^2)}{\sup_{\mu, \sigma^2} L(\mu, \sigma^2)} = \frac{L(\mu_0, \hat{\sigma}_0^2)}{L(\hat{\mu}, \hat{\sigma}^2)}$$

όπου $\hat{\mu}, \hat{\sigma}^2$ οι ΕΜΠ των μ, σ^2 στον πλήρη παραμετρικό χώρο

ενώ $\mu_0, \hat{\sigma}_0^2$ ο ΕΜΠ της σ^2 όταν $\mu = \mu_0$

Οι ΕΜΠ των μ και σ^2 είναι $\hat{\mu} = \bar{x}$ και $\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

Εύρεση $\hat{\sigma}_0^2$ της σ^2 όταν $\mu = \mu_0$

Η πιθανοφάνεια είναι: $L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} =$

$= \frac{1}{(\sigma\sqrt{2})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2}$

όταν $\mu = \mu_0$ $\quad L(\mu_0, \sigma^2) = \frac{1}{(\sigma\sqrt{2})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2}$

$\frac{\partial \log L(\mu_0, \sigma^2)}{\partial \sigma^2} = 0 \Rightarrow \ldots \Rightarrow \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2$ ο ΕΜΠ της σ^2 όταν $\mu = \mu_0$

Επομένως $\lambda = \frac{L(\mu_0, \hat{\sigma}_0^2)}{L(\hat{\mu}, \hat{\sigma}^2)} = \frac{\frac{1}{(\hat{\sigma}_0\sqrt{2\pi})^n} e^{-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^{n}(x_i - \mu_0)^2}}{\frac{1}{(\hat{\sigma}\sqrt{2\pi})^n} e^{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \bar{x})^2}} =$

$= \frac{\frac{1}{(\hat{\sigma}_0\sqrt{2\pi})^n} e^{-n/2}}{\frac{1}{(\hat{\sigma}\sqrt{2\pi})^n} e^{-n/2}} \Rightarrow \lambda = \left(\frac{\hat{\sigma}}{\hat{\sigma}_0}\right)^n = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} \Rightarrow$

$\Rightarrow \lambda = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2}\right)^{n/2} = \left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \mu_0)^2}\right)^{n/2}$

Αλλά $\sum_{i=1}^{n}(x_i - \mu_0)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu_0)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$

Καταλήγω ότι $\lambda = \left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2 + n(\bar{X}-\mu_0)^2} \right)^{n/2}$ διαιρώ με $\sum_{i=1}^{n}(X_i-\bar{X})^2$ $\Longrightarrow$ (4)

$$\left.\begin{array}{l} \lambda = \left(\dfrac{1}{1+\dfrac{n(\bar{X}-\mu_0)^2}{\sum(X_i-\bar{X})^2}} \right)^{n/2} \\ \\ S^2 = \dfrac{1}{n-1}\sum(X_i-\bar{X})^2 \end{array}\right\} \Rightarrow \lambda = \left(\frac{1}{1+\dfrac{n(\bar{X}-\mu_0)^2}{n-1\, S^2}} \right)^{n/2} \Rightarrow$$

$$\Rightarrow \lambda = \left(\frac{1}{1+\frac{1}{n-1}t^2} \right)^{n/2}$$ όπου $t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ αφού $X_i \sim N(\mu_0, \sigma^2)$ υπό H_0

Κρίσιμη περιοχή $\lambda \leq k$ $\Rightarrow \frac{1}{1+\frac{1}{n-1}t^2} \leq k^{2/n} \Rightarrow$

$$\Rightarrow 1 + \frac{1}{n-1}t^2 \geq \frac{1}{k^{2/n}} \Rightarrow t^2 \geq (n-1)\left(\frac{1}{k^{2/n}} - 1\right) = k''$$

Άρα κ.π. $t^2 \geq k''$ όπου $t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ υπό H_0

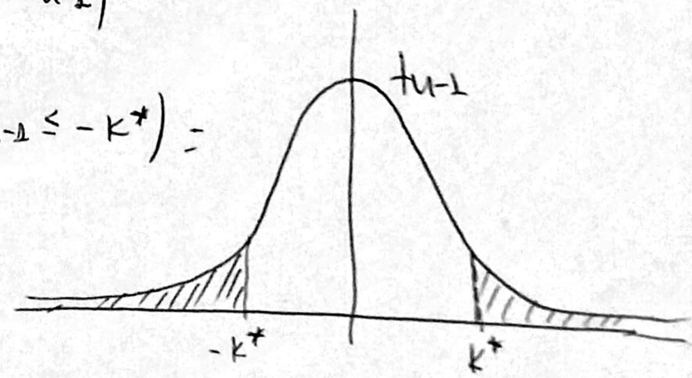
ή $|t| \geq (k'')^{1/2} = k^*$

Προσδιορισμός κ.σ k^*

$$\alpha = P(\text{Απορ } H_0 / H_0 \text{ αληθ}) = P\left(|t| \geq k^* \,\middle|\, t \sim t_{n-1}\right) =$$

$$= P(|t_{n-1}| \geq k^*) = P(t_{n-1} \geq k^* \text{ ή } t_{n-1} \leq -k^*) =$$

$$= P(t_{n-1} \geq k^*) + P(t_{n-1} \leq -k^*) =$$



$$= 2P(t_{n-1} \geq k^*) \Rightarrow P(t_{n-1} \geq k^*) = \frac{a}{2} \xrightarrow[\text{σημείου της } t]{\text{ορισμός εκατοστιαίου}}$$

$$\Rightarrow k^* = t_{n-1, \frac{a}{2}}$$

<u>Συγκεντρωτικά</u>: Για τον έλεγχο $H_0: \mu = \mu_0$ έναντι $H_a: \mu \neq \mu_0$ η Σ.Σ.Τ είναι $t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$ με κατανομή t_{n-1} υπό H_0 και κ.π με ε.σ a $|t| \geq t_{n-1, a/2}$ (t-τεστ)