Διάλεξη

29/05/2019

ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

SOS ΑΣΚΗΣΗ

Παράδειγμα: Έστω τ.δ. $X_1,\dots,X_n$ από $Exp(\lambda)$

$(f(x,\lambda)=\lambda e^{-\lambda x},\ x>0,\lambda>0)$

Να κατασκευασθεί ΟΙ τεστ για έλεγχο

$H_0:\lambda=\lambda_0$ ($\lambda_0>0$ γνωστό) έναντι $H_a:\lambda>\lambda_0$

Θεωρώ προς έλεγχο: $H_0:\lambda=\lambda_0$ έναντι της $H_a^*:\lambda=\lambda_a$ ($\lambda_a>\lambda_0$). Η Ι-Κ.Π. για έλεγχο της H_0 έναντι της H_a^* είναι $\frac{L_0}{L_k}\le k$ όπου

$$L_0=\prod_{i=1}^{n} f(x_i,\lambda_0)$$

$$L_a=\prod_{i=1}^{n} f(x_i,\lambda_a)$$

$$\Downarrow$$

$$\frac{\prod_{i=1}^{n}\lambda_0 e^{-\lambda_0 x_i}}{\prod_{i=1}^{n}\lambda_a e^{-\lambda_a x_i}}\le k \Rightarrow \frac{\lambda_0^n e^{-\lambda_0\sum x_i}}{\lambda_a^n e^{-\lambda_a \sum x_i}}\le k \Rightarrow$$

$$\Rightarrow e^{(\lambda_a-\lambda_0)\sum x_i}\le\frac{\lambda_a^n}{\lambda_0^n}k \Rightarrow (\lambda_a-\lambda_0)\sum x_i\le \log\left(\frac{\lambda_a^n}{\lambda_0^n}\right)k$$

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

Επομένως η Ι-ΚΠ για τον έλεγχο της $H_0:\lambda=\lambda_0$ έναντι $H_a^*:\lambda=\lambda_a$ ($\lambda_a>\lambda_0$) είναι $\sum_{i=1}^{n}x_i\le k'$. Παρατηρώ ότι η Ι-ΚΠ δεν εξαρτάται από το λ_a. Άρα το ΟΙ-τεστ για έλεγχο $H_0:\lambda=\lambda_0$ έναντι $H_0:\lambda>\lambda_0$ είναι $\sum_{i=1}^{n}x_i\le k'$.

<u>Υπολογισμός</u> κ' : Πάντα το κ.σ. κ' υπολογίζεται

από το $a = P(\text{Απορ } H_0 \mid H_0 : \text{αληθής})$

$$= P\left(\sum_{i=1}^{n} X_i \le \kappa' \mid X \sim \text{Εκθ}(\lambda_0)\right)$$

Πρέπει να βρω κατανομή $\sum_{i=1}^{n} X_i$, $X_i \sim \text{Εκθ}(\lambda_0)$

Αφού $\underset{\text{ανεξ.}}{X_i} \sim \text{Εκθ}(\lambda_0) \rightsquigarrow \underset{\text{ανεξ.}}{X_i} \sim G\left(1, \frac{1}{\lambda_0}\right)$ γιατί

$$\text{Εκθ}(\lambda) = G\left(1, \frac{1}{\lambda}\right)$$

$$\left(\frac{1}{\beta^{a}\Gamma(a)} x^{a-1} e^{-x/a}\right)$$

Αφού X_i ανεξ. και $X_i \sim G\left(1, \frac{1}{\lambda_0}\right)$ το

$\sum_{i=1}^{n} X_i \sim G\left(n, \frac{1}{\lambda_0}\right)$. (με μέθοδο ροπογεννήτριας)

Άρα $a = P\left(\sum X_i \le \kappa' \mid \sum X_i \sim G\left(n, \frac{1}{\lambda_0}\right)\right)$

$$= \int_0^{\kappa'} f_{G(n, \frac{1}{\lambda_0})}(t)\, dt = \int_0^{\kappa'} \frac{\lambda_0^{n}}{\Gamma(n)} t^{n-1} e^{-\lambda_0 t}\, dt.$$

$\Rightarrow$ Με αριθμητική ολοκλήρωση το κ' προκύπτει από την λύση της εξίσωσης (*)

Έστω $Y=\sum_{i=1}^{n} X_i$ το οποίο $Y \sim G(n, \frac{1}{\lambda_0})$

Θεωρώ την τ.μ. $T = 2\lambda_0 Y = 2\lambda_0 \sum_{i=1}^{n} X_i$

Με αλλαγή μεταβλητών η κατανομή του T είναι $f_T(t,\lambda_0) = \frac{1}{2^n \Gamma(n)} t^{n-1} e^{-t/2}$, $t>0$

$$\Downarrow$$

$$T \sim G(n, 2) \equiv \chi^2_{2n}$$

Έτσι από όλα τα παραπάνω προκύπτει ότι

$a = P\left(2\lambda_0 \sum_{i=1}^{n} X_i \le 2\lambda_0 \cdot k' / 2\lambda_0 \sum X_i \sim \chi^2_{2n}\right)$

$= P(\chi^2_{2n} \le k'') = 1-a \Rightarrow k'' = \chi^2_{2n,1-a}$

Συγκεντρωτικά: Το Ο.Ι. τεστ που ζητείται έχει ΣΣΤ

την $T = 2\lambda_0 \sum_{i=1}^{n} X_i$ με χ^2_{2n} υπό Η0 και Κ.Π.

$T = 2\lambda_0 \sum_{i=1}^{n} X_i \le \chi^2_{2n,1-a}$

Παρατήρηση: Πως σκέφτηκα το $T = 2\lambda_0 \sum_{i=1}^{n} X_i$ ως σ.σ.τ.:

$G(\kappa,\lambda) = \frac{\lambda}{2} \chi^2_{2\kappa} \Rightarrow \chi^2_{2\kappa} = \frac{2}{\lambda} G(\kappa,\lambda)$

Από $\left.\begin{array}{c} \sum X_i \sim G(n, \frac{1}{\lambda_0}) \\ \| \\ \frac{1}{2\lambda_0} \chi^2_{2n} \end{array}\right\} \Rightarrow 2\lambda_0 \sum X_i \sim \chi^2_{2n}$

(αυτός για εξετ.)

Τεστ Πηλίκου Μεγίστων Πιθανοφανειών:

Η θεωρία των Neymann-Pearson εφοδιάζει με άριστα τεστ από άποψη ισχύος, έχει όμως περιορισμένη ευελιξία γιατί δίνει την δυνατότητα ελέγχου απλής προς απλή ή απλή προς σύνθετη

Ερώτημα: Υπάρχουν άλλες; $\longrightarrow$ Ναι

π.χ

τ.δ. $X_1,\dots,X_n$ από $N(\mu,\sigma^2)$, σ^2: άγνωστο και

Έστω: $\underbrace{H_0: \mu=\mu_0 \ (\mu_0: \text{γνωστό})}_{\text{Είναι απλή;}}$, $H_1: \mu>\mu_0$.

↑

Είναι μια περίπτωση που δεν καλύπτεται από Neymann-Pearson.

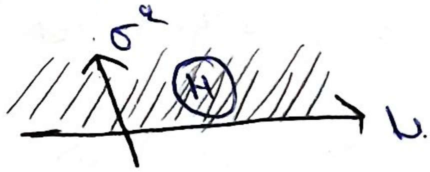
Φορμαλισμός: Έστω τ.δ. $X_1,\dots,X_n$ από $f(x,\theta)$, $\theta\in\Theta\subseteq\mathbb{R}^k$.

Έστω ότι έχω προς έλεγχο

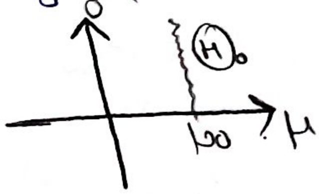
$H_0: \theta\in\Theta_0\subseteq\Theta$ έναντι της $H_1: \theta\in\Theta_1\subseteq\Theta$

π.χ. (σύνδεση με $N(\mu,\sigma^2)$, σ^2: άγνωστο)

$\Theta=\{(\mu,\sigma^2):\mu\in\mathbb{R},\sigma^2>0\}$



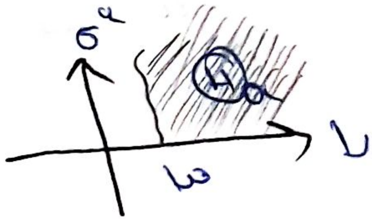
Έστω $H_0: \mu=\mu_0$, μ_0: γνωστό, σ^2: άγνωστο



$H_0: \theta=(\mu,\sigma^2)\in\Theta_0\subseteq\Theta$

$\Theta_0=\{(\mu,\sigma^2), \mu=\mu_0, \sigma^2>0\}$

$H_0: \mu>\mu_0$, σ^2: αγν.



$H_0: \theta(\mu,\sigma^2)\in\Theta_0\subseteq\Theta$

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

Περιγραφή TEST Πηλίκου Μεγίστων Πιθανοφανειών

Στοιχείο 1: Ο λόγος $\lambda = \dfrac{\sup_{\Theta_0} L(\theta)}{\sup_{\Theta} L(\theta)} = \dfrac{L(\hat{\theta}_0)}{L(\hat{\theta})}$

Όπου $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$ η πιθανοφάνεια και όπου

$\hat{\theta}_0$ ο ΕΜΠ της θ από την $H_0: \theta \in \Theta_0$

$\hat{\theta}$ ο ΕΜΠ της θ στον πλήρη παρ. χώρο Θ

Πως φανταστήκα το λ;

Από το $\dfrac{L_0}{L_a}$ της θεωρίας Neyman-Pearson: $\dfrac{L_0}{L_a} = \dfrac{\prod_{i=1}^{n} f(x_i,\theta_0)}{\prod_{i=1}^{n} f(x_i,\theta_a)}$

Στοιχείο 2: (Μορφή κρ. περιοχής)

$$\lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{P_{\hat{\theta}_0}(x_1-\varepsilon_1 \le X_1 \le x_1+\varepsilon_1, \dots, x_n-\varepsilon_n \le X_n \le x_n+\varepsilon_n)}{P_{\hat{\theta}}(\dots\dots \quad -||- \quad \dots\dots)}$$

Μικρές τιμές του στατιστικού συμβαίνει όταν

$$\boxed{P_{\hat{\theta}_0} << P_{\hat{\theta}}}$$

Άρα κ.π. μικρές τιμές του $\lambda \le k$. Το κρίσιμο σημείο

k υπολογίζεται έτσι ώστε: $a = P(\text{Απορ } H_0 \mid H_0: \text{αληθής})$

Εφαρμογή t-test:

Έστω τ.δ. $X_1, \dots, X_n$ από $N(\mu, \sigma^2)$ με σ^2: άγνωστο.
Να κατασκευάσετε τεστ πηλίκου μέγιστων πιθανοφανειών για έλεγχο:

$$H_0: \mu = \mu_0 \quad v \quad H_0: \mu \neq \mu_0$$

Θέτουμε τον λόγο μέγιστων πιθανοφανειών

$$\lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}, \qquad L = L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

$$= \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{L(\mu_0, \hat{\sigma}_0^2)}{L(\hat{\mu}, \hat{\sigma}^2)}$$

ΕΜΠ της σ^2 όταν $\mu = \mu_0$

ΕΜΠ σ'όλο τον παραμετρικό χώρο $\Theta = \{(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$

Ξεκινάω από παρατήρηση και έχουμε βρει ότι

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Αρκεί να βρω ΕΜΠ της σ^2 υπό την $H_0: \mu = \mu_0$.

Υπό την $H_0: \mu = \mu_0$ η $L = \prod_{i=1}^{n} f(x_i, \mu_0, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i - \mu_0)^2}$

$$\frac{\partial}{\partial \sigma^2} \log L = 0 \Rightarrow \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

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

Υπολογ. του κ.σ κ^2:

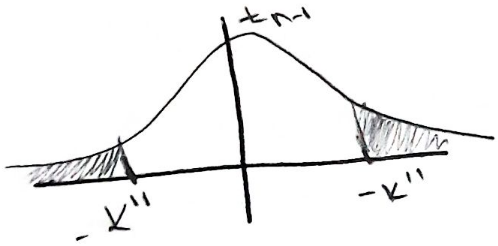
$$\alpha = P(\text{Απορρ } H_0 \mid H_0 : \text{αληθ}) = P\left(t^2 \geq k' \mid X \sim N(\mu_0, \sigma^2)\right)$$

$$= P\left(t^2 \geq k' \mid t = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}\right)$$

$$= P\left(t \geq k'' \text{ ή } t \leq -k'' \mid t \sim t_{n-1}\right)$$

$$= P(t \geq k'' \mid t \sim t_{n-1}) + P(t \leq -k'' \mid t \sim t_{n-1})$$

$$= 2P(t \geq k'' \mid t \sim t_{n-1})$$



$$= P(t \geq k'' \mid t \sim t_{n-1}) = \alpha/2 \Rightarrow k'' = t_{n-1,\alpha/2}$$

Συγκεντρωτικά:

Για τον έλεγχο της $H_0: \mu = \mu_0$ v $H_0: \mu \neq \mu_0$ όταν σ^2: άγνωστο η ΣΣΤ είναι $t = \dfrac{\bar{X} - \mu_0}{S / \sqrt{n}}$ με κατανομή t_{n-1} από H_0 και κ.π. $|t| \geq t_{n-1,\alpha/2}$.