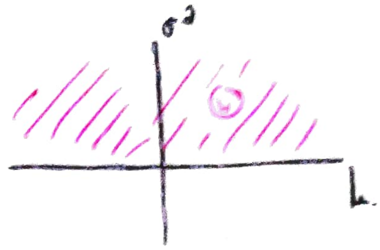
ΤΡΙΤΗ 25 ΜΑΙΟΥ 2021 - ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

Τεστ Πηλίκου Μέγιστων Πιθανοφανειών (ΤΠΜΠ)

Έστω τ.δ. $X_1,\dots,X_n$ από $f(x;\theta)$, $\theta\in\Theta\subseteq\mathbb{R}^k$ ενδιαφέρει ο έλεγχος $H_0: \theta\in\Theta_0$ έναντι της $H_a: \theta\in\Theta_a$. $\Theta_0, \Theta_a \subseteq \Theta$.

Π.χ. $X_1,\dots,X_n \sim N(\mu,\sigma^2)$, μ,σ^2: άγνωστες.

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

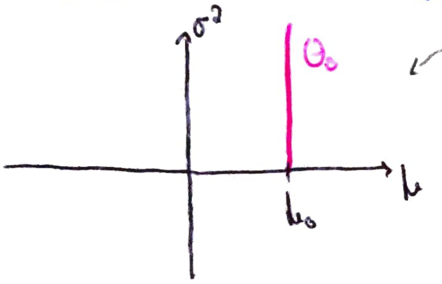


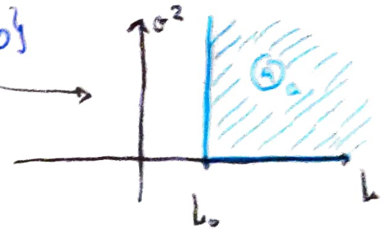
$H_0: \mu=\mu_0$ (μ_0 γνωστό) έναντι $H_a: \mu>\mu_0$

$\Updownarrow$

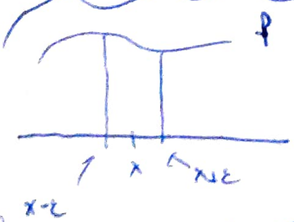
$H_0: (\mu,\sigma^2)\in\Theta_0$ έναντι $H_a: (\mu,\sigma^2)\in\Theta_a$, όπου $\Theta_0=\{(\mu,\sigma^2): \mu=\mu_0, \sigma^2>0\}$

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





Θεωρούμε την πιθανοφάνεια $L(\theta)=\prod_{i=1}^{n} f_{X_i}(x_i,\theta)$. Λαμβάνοντας υπόψη ότι $\prod_{i=1}^{n} f_{X_i}(x_i,\theta) \approx P_\theta(x_1-\varepsilon \le X_1 \le x_1+\varepsilon, \dots, x_n-\varepsilon \le X_n \le x_n+\varepsilon)$ → η πιθανότητα πραγματοποίησης του τ.δ.



$f(x) \approx P(x-\varepsilon \le X \le x+\varepsilon)$

Βήμα 1: Ορίζουμε τον λόγο: $\lambda = \dfrac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \dfrac{L(\hat\theta_0)}{L(\hat\theta)}$

$\hat\theta_0$ ο ΕΜΠ του θ, για $\theta\in\Theta_0$

ή $\hat\theta_0$ ο ΕΜΠ του θ υπό την H_0 (λαμβάνοντας υπόψη την H_0)

$\hat\theta$ ο ΕΜΠ του θ, στο πλήρη παραμετρικό χώρο Θ.

- Προσπαθούμε να εκφράσουμε το λ ως συνάρτηση μιας στατιστικής συνάρτησης με γνωστή κατανομή όταν ισχύει η H_0

Βήμα 2: Μορφή κ.π.: Δηλαδή: Για ποιες τιμές του λ απορ. η H_0?

Μικρές τιμές του λ συνηγορούν σε απόρριψη της H_0. Επομένως, η μορφή της κ.π. είναι: $\lambda \le k$, $\forall$ σταθερά k.

Βήμα 3: Υπολογισμός κ.σ. k: Πάντα από την εξίσωση $\alpha = P(\text{απορ } H_0 \mid H_0 \text{ αληθ}) =$
$= P(\lambda \le k \mid H_0 \text{ αληθ})$.

Παράδειγμα (t-test ως εφαρμογή του ΓΠΛΜΠ)

Έστω τ.δ. $X_1, \dots, X_n$ από $N(\mu, \sigma^2)$, μ, σ^2 άγνωστες. Να κατασκευάσετε τεστ για τον έλεγχο της: $H_0: \mu = \mu_0$, μ_0 γνωστό, έναντι $H_a: \mu \ne \mu_0$.

(Δεν μπορώ να εφαρμόσω Neyman-Pearson γιατί παρόλο που έχω απλή προς σύνθετη, η διασπορά είναι άγνωστη!)

Ξεκινάμε με τον λόγο:
$$\lambda = \frac{\sup_{\Theta_0} L(\mu,\sigma^2)}{\sup_{\Theta} L(\mu,\sigma^2)} = \frac{\sup_{\mu=\mu_0,\sigma^2>0} L(\mu,\sigma^2)}{\sup_{\mu,\sigma^2} L(\mu,\sigma^2)} = \frac{\sup_{\sigma^2} L(\mu_0,\sigma^2)}{\sup_{\mu,\sigma^2} L(\mu,\sigma^2)} =$$

$$= \frac{L(\mu_0, \hat\sigma_0^2)}{L(\hat\mu, \hat\sigma^2)}$$
, όπου $\hat\sigma_0^2$ ο ΕΜΠ της σ^2 υπό την H_0, δηλ. για $\mu=\mu_0$, ενώ $\hat\mu, \hat\sigma^2$ οι ΕΜΠ των μ, σ^2.

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

Ο ΕΜΠ της μ βρέθηκε να είναι: $\hat\mu = \bar{x}$, ενώ, ο ΕΜΠ της σ^2 βρέθηκε να είναι ο: $\hat\sigma^2 = \dfrac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$
(από προηγ. μαθ.)

Άρα, $L(\hat\mu, \hat\sigma^2) = \dfrac{1}{(\hat\sigma\sqrt{2\pi})^n} \exp\left\{-\dfrac{1}{2\hat\sigma^2}\sum (x_i-\bar{x})^2\right\} = \dfrac{1}{(\hat\sigma\sqrt{2\pi})^n} \exp\left\{\dfrac{-n}{2\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})^2\right\}$

$$\Rightarrow \boxed{L(\hat\mu, \hat\sigma^2) = \frac{1}{(\hat\sigma\sqrt{2\pi})^n} e^{-n/2}} \quad (1)$$

Εύρεση $\hat{\sigma}_0^2$: Λαμβάνοντας υπόψη την H_0, η $L(\mu_0,\sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2}$

Άρα, το $\hat{\sigma}_0^2$ θα προκύψει από μεγιστοποίηση της $L(\mu_0,\sigma^2)$

$$\frac{\partial}{\partial\sigma^2}\log L(\mu_0,\sigma^2) = 0 \Rightarrow \dots \Rightarrow \sigma^2 = \frac{1}{n}\sum(x_i-\mu_0)^2 . \text{ Άρα } \boxed{\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2}$$

Επομένως : $L(\mu_0,\hat{\sigma}_0^2) = \frac{1}{(\hat{\sigma}_0^2\sqrt{2\pi})^n} e^{-\frac{1}{2\hat{\sigma}_0^2}\sum(x_i-\mu_0)^2} = \frac{1}{(\hat{\sigma}_0\sqrt{2\pi})^n} e^{-\frac{n}{2\cancel{\sum(x_i-\mu_0)^2}}\cancel{\sum(x_i-\mu_0)^2}} =$

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

Άρα $\boxed{L(\mu_0,\hat{\sigma}_0^2) = \frac{1}{(\hat{\sigma}_0\sqrt{2\pi})^n} e^{-n/2}}$ (2)

Από (1), (2) : $\lambda = \dfrac{\frac{1}{(\hat{\sigma}_0\sqrt{2\pi})^n}\cancel{e^{-n/2}}}{\frac{1}{(\hat{\sigma}\sqrt{2\pi})^n}\cancel{e^{-n/2}}} = \left(\frac{\hat{\sigma}}{\hat{\sigma}_0}\right)^n = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\mu_0)^2}\right)^{n/2}$

Άρα, $\lambda = \left(\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\mu_0)^2}\right)^{n/2}$.

$$\sum(x_i-\mu_0)^2 = \sum(x_i-\bar{x}+\bar{x}-\mu_0)^2 = \sum(x_i-\bar{x})^2 + 2\sum(x_i-\bar{x})(\bar{x}-\mu_0) + \sum(\bar{x}-\mu_0)^2 =$$

$$= \sum(x_i-\bar{x})^2 + 2(\bar{x}-\mu_0)(\sum x_i - \sum\bar{x}) + n(\bar{x}-\mu_0)^2 =$$

$$= \sum(x_i-\bar{x})^2 + 2(\bar{x}-\mu_0)(\cancel{n\bar{x}-n\bar{x}}) + n(\bar{x}-\mu_0)^2 = \sum(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2.$$

Άρα: $\lambda = \left(\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2+n(\bar{x}-\mu_0)^2}\right)^{n/2} = \left(\frac{1}{1+\frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}}\right)^{n/2} =$

$$= \left(\frac{1}{1+\frac{n(\bar{x}-\mu_0)^2}{(n-1)s^2}}\right)^{n/2} = \left(\frac{1}{1+\frac{1}{n-1}t^2}\right)^{n/2}, \text{ όπου } t^2 = \left(\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right)^2$$

Βήμα 2: κ.π. $\lambda \le k \Rightarrow \left(\dfrac{1}{1+\frac{1}{n-1}t^2}\right)^{n/2} \le k \Rightarrow \dfrac{1}{1+\frac{1}{n-1}t^2} \le k^{2/n} \Rightarrow$

$\Rightarrow 1 + \dfrac{1}{n-1}t^2 \ge k^{2/n} \Rightarrow \dfrac{1}{n-1}t^2 \ge k^{2/n} - 1 \Rightarrow t^2 \ge (n-1)(k^{2/n}-1)$

$\Rightarrow |t| \ge \left((n-1)(k^{2/n}-1)\right)^{1/2} = k'$. Άρα κ.π.: $|t| \ge k'$

Βήμα 3: εύρεση του κ.σ. k'. Πάντα: από: $\alpha = P(\text{Απορ } H_0 \mid H_0 \text{ αληθ.}) =$

$= P\left(|t| \ge k' \mid \bar{X} \sim N(\mu_0, \sigma^2/n) \text{ ή } t = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}\right) = P(|t_{n-1}| \ge k') =$

$= P(|t_{n-1}| \ge k') = P(t_{n-1} \ge k' \text{ ή } t_{n-1} \le -k') = 2P(t_{n-1} \ge k') \Rightarrow$

$\Rightarrow P(t_{n-1} \ge k') = \dfrac{\alpha}{2} \Rightarrow \boxed{k' = t_{n-1,\alpha/2}}$

Για τον έλεγχο αυτόν η ΣΣΤ είναι: $t = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$ με κατανομή t_{n-1}, υπό την H_0 και κ.π. μεγέθους α, $|t| \ge t_{n-1,\alpha/2}$.

Παρατήρηση: $-2\log\lambda \xrightarrow[n\to\infty]{d} \chi^2_{r-s}$

διάσταση Θ_0

διάσταση του Θ