

2/12/2016

Πρόταση: Όταν n αβλ. είναι ανεξάρτητες τότε:

- $Y_1 = F_X(X, \theta) \sim U(0, 1)$
- $Y_2 = 1 - F_X(X, \theta) \sim U(0, 1)$
- $Y_3 = -\log Y_1 \sim \text{Exp}(1)$
- $Y_4 = -\log Y_2 \sim \text{Exp}(1)$

Άρα $2 \sum_{i=1}^n Y_{3i} \sim \chi^2_{2n}$, $2 \sum_{i=1}^n Y_{4i} \sim \chi^2_{2n}$

Απόδειξη: $F_{Y_1}(y_1) = P(Y_1 \leq y_1) = P[F_X(X, \theta) \leq y_1] = P[X \leq F_X^{-1}(y_1)] =$
 $= F_X(F_X^{-1}(y_1)) = y_1, \quad 0 < y_1 < 1$

• $F_{Y_2}(y_2) = P(Y_2 \leq y_2) = P(1 - Y_1 \leq y_2) = P(Y_1 \geq 1 - y_2) = 1 - P(Y_1 < 1 - y_2) =$
 $= 1 - F_{Y_1}(1 - y_2) = 1 - (1 - y_2) = y_2, \quad 0 < y_2 < 1$

• $F_{Y_3}(y_3) = P(Y_3 \leq y_3) = P(-\log Y_1 \leq y_3) = P(\log Y_1 \geq -y_3) = P(Y_1 \geq e^{-y_3}) =$
 $= 1 - P(Y_1 < e^{-y_3}) = 1 - F_{Y_1}(e^{-y_3}) = 1 - e^{-y_3}, \quad y_3 > 0$
 $\text{Exp}(1)$

$\left. \begin{array}{l} Y_{3i} \sim \text{Exp}(1) \equiv \text{Gamma}(1, 1) \\ Y_{4i} \sim \text{Exp}(1) \equiv \text{Gamma}(1, 1) \end{array} \right\} \quad \begin{array}{l} \sum_{i=1}^n Y_{3i} \sim \text{Gamma}(n, 1) \\ \sum_{i=1}^n Y_{4i} \sim \text{Gamma}(n, 1) \end{array}$

$2 \sum Y_{3i} \sim \text{Gamma}(n, 2)$

$2 \sum Y_{4i} \sim \text{Gamma}(n, 2)$

Άρα έστω ότι:

$$\begin{array}{l} -2 \sum \log f_X(x_i, \theta) \sim \chi^2_{2n} \\ \text{και} \\ -2 \sum \log [1 - f_X(x_i, \theta)] \sim \chi^2_{2n} \end{array}$$

→ με αυτά δεν μπορούμε να έρω δ.ε. ελ.π.κ.α. Μόνο ίσως οφείν.

↓ (η) βλεχθεία με ασφάλεια από.

(ΠΚ) $X_1, \dots, X_n$ ε.δ. $f(x_i, \theta) = \frac{1}{\theta} e^{-x/\theta}$, $x \geq 0$
 Βρείτε δ. ε. π. για την $f(x, \theta)$

λύση: Δεν ξεκινώ με τους δύο παραγόμενους τύπους που αναφέρεται γιατί εκείνοι δίνουν μόνο $\delta \in$ ίδιων αρίθ. Ενώ έχω ίδιους μπορεί να βρω με άλλη κατανομή $\delta \in$ ελ. π. κ.

Αρα ξεκινώ με επάρκεια: $\prod_{i=1}^n f(x_i, \theta) = \frac{1}{\theta^n} e^{-\sum x_i / \theta}$

$T = \sum x_i$ επαρκής

$X_i \sim \text{Exp}(\frac{1}{\theta}) \equiv \text{Gamma}(1, \theta)$

$\rightarrow \sum x_i \sim \text{Gamma}(n, \theta)$ (αποδείξη με πολλαπλασιασμό)

Δεν μου κάνει γιατί στο $\sum x_i$ δεν έχω θ και στο $\text{Gamma}(n, \theta)$ έχω θ , το οποίο πρέπει να φύγει

Αρα έχω: $\frac{2}{\theta} \sum x_i \sim \text{Gamma}(n, 2)$ (μολ/βα με το 2 γιατί βέρω
 ότι το $\text{Gamma}(n, 2) = \chi_{2n}^2$
 Ενώ αν έχω $\text{Gamma}(n, 1)$ δεν θα ισοβαί με κάτι άλλο.

$$P(q_1 \leq \overset{Q}{\frac{2}{\theta} \sum x_i} \leq q_2) = 1 - \alpha \Rightarrow$$

$$\Rightarrow P\left(\frac{2 \sum x_i}{q_2} \leq \theta \leq \frac{2 \sum x_i}{q_1}\right) = 1 - \alpha$$

Δεν μπορεί να βρω ελ. π. κ. άρα βρίσκω ίδιων αρίθ.

$$P(Q \leq q_1) = \alpha/2 \Rightarrow q_1 = \chi_{2n, \alpha/2}^2$$

$$P(Q \geq q_2) = \alpha/2 \Rightarrow q_2 = \chi_{2n, 1-\alpha/2}^2$$

2ος τρόπος

$$F_X(x) = \int_0^x \frac{1}{\theta} e^{-y/\theta} dy = -e^{-y/\theta} \Big|_0^x = 1 - e^{-x/\theta}, \quad x \geq 0$$

$$-2 \sum \log[1 - 1 + e^{-x_i/\theta}] \sim \chi_{2n}^2$$

$$2 \sum \frac{x_i}{\theta} \sim \chi_{2n}^2$$

Άρα καταλήγω στο ίδιο. Άρα εδώ θα μπορούσα να πω ότι να χρησιμοποιήσω και τον 2ο τρόπο.

$$(\pi x) \quad x_1, \dots, x_n \text{ ε.δ. } f(x|\theta) = e^{-(x-\theta)}, x \geq \theta$$

$$\text{Λύση: } \prod_{i=1}^n f(x_i|\theta) = e^{-\sum x_i} e^{n\theta}, \left. \begin{array}{l} x_1 > \theta \\ x_2 > \theta \\ \vdots \\ x_n > \theta \end{array} \right\} \min x_i > \theta$$

$$\text{Άρα } \prod_{i=1}^n f(x_i|\theta) = e^{-\sum x_i} e^{n\theta} I_{(\theta, \infty)}(\theta)$$

$$T = \min x_i = x_{(1)} \text{ επαρκεί.}$$

Τώρα πρέπει να βρω την κατανομή του επαρκούς.

$$F_{x_{(1)}}(y) = P(x_{(1)} \leq y) = \dots = 1 - (1 - F_X(y))^n$$

$$f_{x_{(1)}}(y) = n(1 - F_X(y))^{n-1} f_X(y)$$

$$\text{Α.σ.κ: } F_X(y) = P(X \leq y) = \int_{\theta}^y e^{-(x-\theta)} dx = 1 - e^{-(y-\theta)}, y \geq \theta$$

$$\text{Άρα } f_{x_{(1)}}(y) = n e^{-(n-1)(y-\theta)} e^{-(y-\theta)} = n e^{-n(y-\theta)}, y \geq \theta$$

πρέπει να το γράψω

$$\text{Αλλάζω μεταβλητή: } w = x_{(1)} - \theta \Rightarrow x_{(1)} = w + \theta \Rightarrow dx_{(1)} = dw$$

$$f_w(w) = f_{x_{(1)}}(w+\theta) \cdot |1| = n e^{-nw}, w \geq 0$$

$$\underbrace{x_{(1)} - \theta}_{\text{εξ ατ. } \theta} = w \sim \text{Exp}(n) \equiv \text{Gamma}(1, \frac{1}{n})$$

$$2n(x_{(1)} - \theta) \sim \text{Gamma}(1, 2) = \chi_2^2$$

ανεξάρτητοι του θ. Άρα που κάνει αρκεί να παλιώ με 2 για να καταλήγω σε γνωστό (χ_2^2)

$$Q = 2n(x_{(1)} - \theta)$$

$$\text{Παίρνω: } P(q_1 \leq Q \leq q_2) = 1 - \alpha \text{ και βρίσκω ελπίχ. πηκ.}$$

Δοκιμάσω αν γίνεται με τον άλλο τρόπο.

$$F_X(x) = 1 - e^{-(x-\theta)} \quad x \geq \theta$$

$$-2 \sum \log[1 - (1 - e^{-(x_i - \theta)})]$$

$$2 \sum (x_i - \theta) \sim \chi^2_{2n}$$

$$2 \sum x_i - 2n\theta \sim \chi^2_{2n}$$

(ΠX) $X_1, \dots, X_n$ τ.δ. $f(x|\theta) = \frac{1}{\theta}, 0 < x < \theta$

Λύση: $T = X_{(n)}$ επαρκής

$$F_{X_{(n)}}(y) = P(\max X_i \leq y) = F_X^n(y)$$

$$f_{X_{(n)}}(y) = n f_X^{n-1}(y) f_X(y)$$

Αρκ: $F_X(y) = P(X \leq y) = \int_0^y \frac{1}{\theta} dy = \frac{y}{\theta}, 0 < y < \theta$

$$f_T(t) = n \frac{t^{n-1}}{\theta^{n-1}} \frac{1}{\theta}, 0 < t < \theta$$

Θέσω: $W = \frac{T}{\theta} \Rightarrow T = \theta W$
 $dt = dW \cdot \theta$

$$f_W(w) = f_T(\theta w) |\theta| = n \frac{\theta^{n-1} w^{n-1}}{\theta^{n-1}} \frac{1}{\theta} \theta = n w^{n-1}, 0 < w < 1$$

Ανέβαφα ότι: $W = \frac{T}{\theta}$ έχει $f_W(w) = n w^{n-1}, 0 < w < 1$

$$P(q_1 \leq \frac{T}{\theta} < q_2) = 1 - \alpha \Rightarrow P(\frac{T}{q_2} < \theta < \frac{T}{q_1}) = 1 - \alpha$$

Πρόβλεψη δ.ε. ελαχ. πιθανότητας

Σημ: Αν έστω ο τον 2ος τρόπο: $f_X(x) = \frac{x}{\theta}, 0 < x < \theta$

$$\text{Έχω: } -2 \sum \log \frac{x_i}{\theta} \sim \chi^2_{2n}$$

$$-2 \sum \log x_i + 2n \log \theta \sim \chi^2_{2n}$$

Αρα $P(q_1 \leq -2 \sum \log x_i + 2n \log \theta \leq q_2) = 1 - \alpha$

$$P(\frac{q_1 + 2 \sum \log x_i}{2n} \leq \log \theta \leq \frac{q_2 + 2 \sum \log x_i}{2n}) = 1 - \alpha$$

από εδώ
 $\Delta \varepsilon \nu$
 μπορεί να
 βρεθεί μ.

Σημείωση : $X_1, \dots, X_n$ τ.δ. $f(x|\theta_1, \theta_2) = \frac{\theta_2}{\theta_1} x^{\theta_2-1} e^{-\frac{x}{\theta_1}}$, $x \geq 0$
 θ_2 γνωστό, $\theta_1, \theta_2 > 0$

Λόγος : $\prod_{i=1}^n \frac{\theta_2}{\theta_1} x_i^{(\theta_2-1)} e^{-\frac{x_i}{\theta_1}} = \left(\frac{\theta_2}{\theta_1}\right)^n \prod_{i=1}^n x_i^{(\theta_2-1)} e^{-\frac{\sum x_i}{\theta_1}}$

$T = \sum x_i^{\theta_2}$ επαρκές

(Αν ήταν θ_1, θ_2 άγνωστο τότε επαρκές : $(X_1, \dots, X_n)$)

Για να βρω την κατανομή του T .

$y = x^{\theta_2} \Rightarrow x = y^{1/\theta_2}$

$$f_Y(y) = f_X(y^{1/\theta_2}) \left| \frac{1}{\theta_2} y^{\frac{1}{\theta_2}-1} \right| = \frac{\theta_2}{\theta_1} y^{(\theta_2-1)/\theta_2} e^{-y/\theta_1} \frac{1}{\theta_2} y^{\frac{1}{\theta_2}-1} =$$

$$= \frac{1}{\theta_1} e^{-y/\theta_1}, y \geq 0 \quad \text{Exp}\left(\frac{1}{\theta_1}\right)$$

Έλεγχος Στατιστικών Υποθέσεων

1^ο βήμα : Διατύπωση της στατιστικής υπόθεσης

2^ο βήμα : Ανάπτυξη μεθοδολογίας που συμπίπτει στο τ.δ για τον έλεγχο της υπόθεσης

3^ο βήμα : Έφαρμογή συμπερασμάτων

Προβλημα : Μηδενική Υπόθεση (η υπόθεση που κρίνεται με την ελπίδα να απορριφθεί) (H_0)

Εναλλακτική Υπόθεση (H_a)

Ο Στατιστικός βγάζει ένα καύονα απόφασης για την αλήθεια ή μη της μηδενικής υπόθεσης. Αυτό τον οδηγεί σε δύο περιπτώσεις αποφάσεων :

$\alpha = P(\text{βλάπτα τύπου I}) = P(\text{απορ. } H_0 \mid H_0 \text{ αληθινή})$ (επίπεδο σημαντικότητας)

$\beta = P(\text{βλάπτα τύπου II}) = P(\text{αποδ. } H_0 \mid H_a \text{ αληθινή})$

$\gamma = 1 - \beta = P(\text{απορ. } H_0 \mid H_a \text{ αληθινή})$ (ισχύς)

Στην κλ. στατ. εμμέρ. σταθεροποιούμε το α και προσπαθούμε να βρούμε τον καύονα απόφασης που ελαχιστοποιεί το β ή μεγιστοποιεί το γ .

Από αυτό θα θερμοποιήσουμε το α είναι να είναι
 γνωστή η $P(\text{ερ. είναι } I)$ ή θα γνωρίσω εκ των προτέρων
 την $P(\text{ερ. τύπου } I)$

Επιμέλεια ελήφθη να απορ. την H_0 γιατί τότε ξέρω ότι το βράσημα είναι α .

Αντί: όταν καθορίζει παρέρ (πέρικέρ την κατακοπή) της ε.ρ.
 (βωθεμ)

Κατασκευή στατιστικών τεστ για τον έλεγχο αντήρ προς αντήρ

1^η μέθοδος: $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$ (Κλ. βρ. βωπ.) ήμπα Neyman-Pearson

Το πλέον ισχυρό τεστ προδωρίφεται ως εξής:

$$\frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} \leq k \quad (\text{πέρικέρ απόρριψης})$$

όπου το k προδωρίφεται τ.ω. $P(\text{απορ. } H_0 | H_0 \text{ αληθιν.}) = \alpha$

2^η μέθοδος: (Θεωρία Bayes) $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$

d_0 = αποδοχή της H_0

d_1 = αποδοχή της H_1

$$L(\theta_0, d_0) = 0 \quad L(\theta_1, d_0) = k_0 > 0$$

$$L(\theta_0, d_1) = k_1 > 0 \quad L(\theta_1, d_1) = 0$$

Minimax τεστ: Είναι εκείνο για το οποίο: $R(\theta_0, d_1) = R(\theta_1, d_0) \xrightarrow[\text{απόρριψης}]{\text{λέρω}} R(\theta, d)$

$$\Rightarrow k_1 \cdot P(\text{απόρριψης } d_1 | \theta = \theta_0) = k_0 \cdot P(\text{απόρριψης } d_0 | \theta = \theta_1) \Rightarrow$$

$$\Rightarrow k_1 \cdot P(\text{απορ. } H_0 | H_0 \text{ αληθιν.}) = k_0 \cdot P(\text{απορ. } H_0 | H_1 \text{ αληθιν.})$$

$$\Rightarrow k_1 \cdot P\left(\frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} \leq k \mid H_0 \text{ αληθιν.}\right) = k_0 \cdot P\left(\frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} > k \mid H_1 \text{ αληθιν.}\right)$$

Αν αυτό που αληθιν. είναι το πέρ θ_0 προδωρίφω το k .

3^η μέθοδος: (Bayes rule) : Η περιοχή απόρριψης είναι (δηλ. θα απορρ. όταν:)

$$\frac{\pi(\theta_0) \prod_{i=1}^n f(x_i | \theta_0)}{\pi(\theta_1) \prod_{i=1}^n f(x_i | \theta_1)} = \frac{\pi(\theta_0 | x)}{\pi(\theta_1 | x)} < \frac{L(\theta_1, d_0)}{L(\theta_0, d_1)} \Rightarrow$$

$$\Rightarrow \frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} < \frac{L(\theta_1, d_0) \pi(\theta_1)}{L(\theta_0, d_1) \pi(\theta_0)}$$

Συνοψίζοντας: Περιοχή απόρριψης: $\frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} \leq k$

1^η μέθοδος: κ τ.ω. $P(\text{αναρ. } H_0 | \theta = \theta_0) = \alpha$

2^η μέθοδος: $k_1 \cdot P(\text{αναρ. } H_0 | H_0 \text{ αλ.}) = k_0 \cdot P(\text{Αποδ. } H_0 | H_0 \text{ αλ.})$

3^η μέθοδος: $k = \frac{L(\theta_1, d_0) \pi(\theta_1)}{L(\theta_0, d_1) \pi(\theta_0)}$

(ΠΧ) $X_1, \dots, X_n$ τ.δ. από $N(\theta, 100)$, $n=100$ Να ελέγξετε την:
Υπόθεση: $H_0: \theta = 75 \quad \forall \quad H_a: \theta = 78$

i) Ισχυρότατο test 5%

ii) Minimax $L(\theta_0, d_1) = 3$
 $L(\theta_1, d_0) = 1$

iii) Bayes rule $L(\theta_0, d_1) = L(\theta_1, d_0)$
 $\pi(\theta_0) = \pi(\theta_1)$

Σε όλες τις περιπτώσεις να βρεθεί το α και το γ.

$$\text{Λύση: } \frac{\prod_{i=1}^n f(x_i | \theta_0)}{\prod_{i=1}^n f(x_i | \theta_1)} \leq k \Rightarrow \frac{e^{-\frac{1}{2\sigma^2} \sum (x_i - \theta_0)^2}}{e^{-\frac{1}{2\sigma^2} \sum (x_i - \theta_1)^2}} \leq k \Rightarrow$$

$$\Rightarrow e^{-\frac{1}{2\sigma^2} \{ \sum (x_i - \theta_0)^2 - \sum (x_i - \theta_1)^2 \}} \leq k \Rightarrow$$

$$\Rightarrow e^{-\frac{1}{2\sigma^2} \{ -2\theta_0 \sum x_i + n\theta_0^2 + 2\theta_1 \sum x_i - n\theta_1^2 \}} \leq k \quad \textcircled{1} \Rightarrow$$

$$\Rightarrow -\frac{1}{2\sigma^2} \{ 2\sum x_i (\theta_1 - \theta_0) + n(\theta_0^2 - \theta_1^2) \} \leq \ln k \Rightarrow$$

$$\Rightarrow 2\sum x_i (\theta_1 - \theta_0) + n(\theta_0^2 - \theta_1^2) \geq (-2\sigma^2) \ln k \Rightarrow$$

$$\Rightarrow \sum x_i (\theta_1 - \theta_0) + n(\theta_0^2 - \theta_1^2) \geq -\sigma^2 \ln k - \frac{n}{2} (\theta_0^2 - \theta_1^2) \Rightarrow$$

$$\Rightarrow \sum x_i \geq \frac{-\sigma^2 \ln k - (\frac{n}{2} (\theta_0^2 - \theta_1^2))}{\theta_1 - \theta_0} \quad \text{περιοχή απόρριψης}$$

Η 1^η μέθοδος λέει ότι προσδιορίζω το k π.ω. $P(\text{απορ. } H_0 | H_0 \text{ αλ.}) = \alpha$
 Άρα έχω: $\alpha = P(\sum x_i \geq k_1 | x_i \sim N(\theta_0, \sigma^2)) = P(\sum x_i \geq k_1 | x_i \sim N(n\theta_0, n\sigma^2))$
 $\Rightarrow P\left(\frac{\sum x_i - n\theta_0}{\sqrt{n\sigma^2}} \geq \frac{k_1 - n\theta_0}{\sqrt{n\sigma^2}} \mid H_0 \text{ αληθής}\right) = \alpha$

$$\text{όπου: } \frac{k_1 - n\theta_0}{\sqrt{n\sigma^2}} = Z_\alpha \Rightarrow k_1 = \sqrt{n\sigma^2} Z_\alpha + n\theta_0$$

Το επίπεδο σημαντικότητας που ζητάει να βρω είναι το α που βρήκα παραπάνω.

Για την ισχύ:

$$\begin{aligned} \gamma &= P(\text{απορ. } H_0 | H_1 \text{ αλ.}) \Rightarrow \gamma = P(\sum x_i \geq k_1 | \sum x_i \sim N(n\theta_1, n\sigma^2)) = \\ &= \dots = 1 - \Phi\left(\frac{k_1 - n\theta_1}{\sqrt{n\sigma^2}}\right) \end{aligned}$$