Κατασκευή Αντιστρεπτών Ποσοτήτων

ΠΡΟΤΑΣΗ

Έστω τ.δ $X_1, \dots, X_n$ από συνεχή κατανομή με α.σ.κ $F_\theta(x)$.

Τότε (α) $Q = -2 \sum_{i=1}^{n} \log F_\theta(x_i) \sim \chi^2_{2n}$

(β) $Q^* = -2 \sum_{i=1}^{n} \log(1 - F_\theta(x_i)) \sim \chi^2_{2n}$

Παράδειγμα

1) Έστω τ.δ $X_1, \dots, X_n$ από κατανομή $f(x,\theta) = \theta x^{\theta-1}$, $0<x<1$, $\theta>0$ (Beta$(\theta,1)$). Να κατασκευαστεί δ.ε ίσων ουρών για την παράμετρο θ.

Λύση

Εύρεση α.σ.κ F_θ

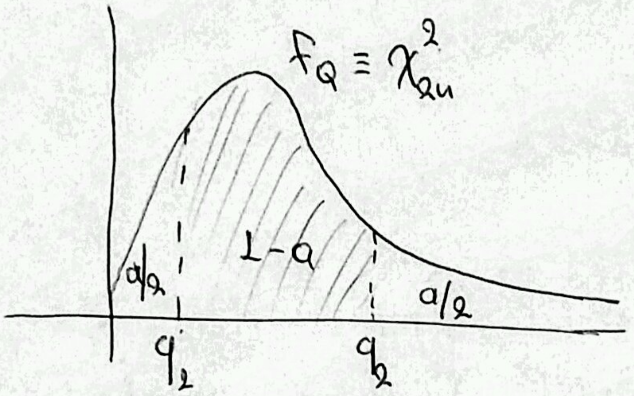
$$\forall x \in \mathbb{R},\ F_\theta(x) \overset{\text{ορ}}{=} P(X \le x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0 & , x \le 0 \\ \int_0^x \theta t^{\theta-1} dt = x^\theta & , 0<x<1 \\ 1 & , x \ge 1 \end{cases}$$

Θεωρώ την $Q = -2 \sum_{i=1}^{n} \log F_\theta(x_i) = -2 \sum_{i=1}^{n} \log x_i^\theta$

Επίσης $Q = -2 \sum_{i=1}^{n} \log x_i^\theta = -2\theta \sum_{i=1}^{n} \log x_i \sim \chi^2_{2n}$

Η Q είναι συνάρτηση των X_i, συνάρτηση της θ και έχει κατανομή ανεξάρτητη της θ $\Rightarrow$ Q αντιστρεπτή

Αφού Q αντιστρεπτή $\exists\, q_1, q_2 \in (0,\infty)$ με $q_1 < q_2$ τέτοια ώστε:



$$1-a = P(q_1 < Q < q_2) =$$

$$= P(q_1 < -2\theta\Sigma\log x_i < q_2) =$$

$$= P\left(-\frac{q_1}{2\Sigma\log x_i} < \theta < -\frac{q_2}{2\Sigma\log x_i}\right)$$

Ένα δ.ε για το θ με β.ε. $100(1-a)\%$ είναι

$$\left(-\frac{q_1}{2\Sigma\log x_i},\ -\frac{q_2}{2\Sigma\log x_i}\right)$$

αφού $0 < x_i < 1 \Rightarrow \log x_i < 0$ γι' αυτό τα q_1, q_2 δεν είναι αντίθετα

Για δ.ε ίσων ουρών επιλέγω τα q_1, q_2 έτσι ώστε:

$P(Q \geq q_2) = \frac{a}{2} \Rightarrow$ και $P(Q \leq q_1) = \frac{a}{2} \Rightarrow$

$P(\chi^2_{2n} \geq q_2) = \frac{a}{2} \Rightarrow$

$1 - P(Q \geq q_1) = \frac{a}{2} \Rightarrow$

$q_2 = \chi^2_{2n,\frac{a}{2}}$

$P(Q \geq q_1) = 1 - \frac{a}{2} \Rightarrow$

$P(\chi^2_{2n} \geq q_1) = 1 - \frac{a}{2} \Rightarrow$

$q_1 = \chi^2_{2n,1-\frac{a}{2}}$

Άρα το δ.ε ίσων ουρών είναι:

$$\left(-\frac{\chi^2_{2n,1-a/2}}{2\Sigma\log x_i},\ -\frac{\chi^2_{2n,a/2}}{2\Sigma\log x_i}\right)$$

Παρατήρηση

1. Μπορούσε να ζητά δ.ε ελαχίστου μήκους. Τότε τα q_1, q_2 θα προσδιορίζονται από αριθμητική επίλυση συστήματος (βλ. π.χ κανονικής)

2. Το επαρκές είναι το $\prod_{i=1}^{n} X_i$. Παρατηρούμε ότι το δ.ε ίσων ουρών εξαρτάται από το επαρκές. Διαισθητικά: Είναι αποδεκτό δ.ε

2) Έστω τ.δ $X_1, \dots, X_n$ από κατανομή $U(0,\theta)$, $0<x<\theta$, $\theta>0$
Να κατασκευαστεί δ.ε ίσων ουρών για την θ.

Λύση

Εύρεση F_θ.

Αν $X \sim U(a,b)$
$$F_X(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & 0<x<b \\ 1, & x \ge b \end{cases}$$

Άρα $F_\theta(x) = \dfrac{x}{\theta}$, $0<x<\theta$

$$Q = -2\sum_{i=1}^{n} \log F_\theta(x_i) = -2\sum_{i=1}^{n} \log x_i + 2n\log\theta \sim \chi^2_{2n}$$

Άρα η Q είναι αντιστρεπτή. Επομένως $\exists\, q_1, q_2 > 0$ με $q_1 < q_2$:

$$1-a = P(q_1 < Q < q_2) \Rightarrow 1-a = P\left(q_1 < -2\sum_{i=1}^{n} \log X_i + 2\log\theta < q_2\right) \Rightarrow$$

$$\Rightarrow 1-a = P\left(e^{q_1/2n} \left(\prod X_i\right)^{1/n} < \theta < e^{q_2/2n} \left(\prod X_i\right)^{1/n}\right)$$

Άρα ένα δ.ε είναι το $\left(e^{q_1/2n} (\Pi X_i)^{1/n}, e^{q_2/2n} (\Pi X_i)^{1/n} \right)$

και το δ.ε ίσων ουρών προκύπτει για $q_1 = \chi^2_{2n, 1-\frac{a}{2}}$, $q_2 = \chi^2_{2n, \frac{a}{2}}$

<u>Παρατήρηση</u>

Το επαρκές για την $U(0,\theta)$ είναι $Y_{(n)}$. Παρατηρούμε ότι το δ.ε δεν σχετίζεται με το $X_{(n)}$. <u>Διαισθητικά</u>: Δεν είναι αποδεκτό δ.ε. Δεν συνιστείται.

3) Έστω τ.δ $X_1, \ldots, X_n$ από κατανομή $U(0,\theta)$ $0<x<\theta$, $\theta>0$

α. νδο $Q = X_{(n)}/\theta$ είναι αντιστρεπτή ποσότητα

β. να κατασκευαστεί δ.ε ίσων ουρών για την θ

γ. —"— ελαχίστου μήκους για την θ.

<u>Λύση</u>

α. Έστω $Y = X_{(n)} = \max\{X_1, \ldots, X_n\}$

Τότε (1ο μάθημα) $f_Y(y) = n[F(y)]^{n-1} f(y)$, $0<y<\theta$

όπου $f(y) = \frac{1}{\theta}$ και $F(y) = \frac{y}{\theta}$, $0<y<\theta$

Άρα $f_Y(y) = \frac{n}{\theta^n} y^{n-1}$, $0<y<\theta$

Για την εύρεση της κατανομής του $Q = X_{(n)}/\theta = \frac{Y}{\theta}$ με αλλαγή μεταβλητών

$$f_Q(q) = nq^{n-1}, \quad 0<q<1.$$

Άρα Q αντιστρεπτή.

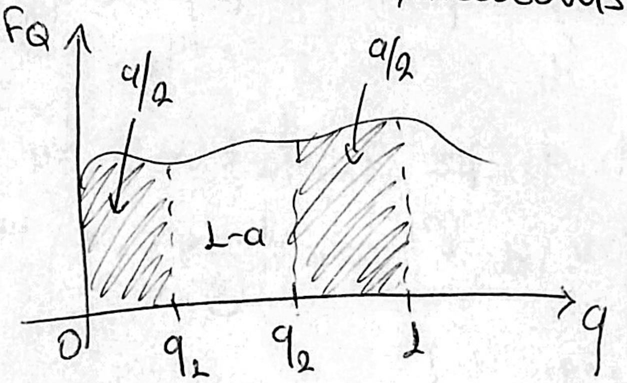
β. Αφού Q αντιστρεπτή $\exists\, q_1, q_2 \in (0,1)$ με $q_1 < q_2$:

$$1-\alpha = P(q_1 < Q < q_2) = P\left(q_1 < \frac{X_{(n)}}{\theta} < q_2\right) = P\left(\frac{X_{(n)}}{q_2} < \theta < \frac{X_{(n)}}{q_1}\right)$$

Από τον ορισμό $100(1-\alpha)\%$ δ.ε έχουμε ότι ένα δ.ε για το θ είναι

$$\left(\frac{X_{(n)}}{q_2}, \frac{X_{(n)}}{q_1}\right)$$

Για την εύρεση διαστήματος εμπιστοσύνης ίσων ουρών τα q_1, q_2.



Δηλαδή τα q_1, q_2 τέτοια ώστε $\frac{\alpha}{2} = P(Q \le q_1)$ και $\frac{\alpha}{2} = P(Q \ge q_2)$

- $P(Q \le q_1) = \frac{\alpha}{2} \Rightarrow \int_0^{q_1} f_Q(q)\,dq = \int_0^{q_1} n q^{n-1}\,dq = q^n \Big|_0^{q_1} = q_1^n \Rightarrow$

$$\Rightarrow q_1 = \sqrt[n]{\frac{\alpha}{2}}$$

- $P(Q \ge q_2 = \int_{q_2}^{1} f_Q(q)\,dq = \int_{q_2}^{1} n q^{n-1}\,dq = q^n \Big|_{q_2}^{1} = 1 - q_2^n \Rightarrow$

$$\Rightarrow q_2 = \sqrt[n]{1 - \frac{\alpha}{2}}$$

Άρα το δ.ε ίσων ουρών θα είναι $\left(\frac{X_{(n)}}{\sqrt[n]{1-\alpha/2}}, \frac{X_{(n)}}{\sqrt[n]{\frac{\alpha}{2}}}\right)$

γ. Βρήκαμε ότι με βάση την $Q = \frac{X_{(n)}}{\theta}$ ένα δ.ε είναι το

$$\left(\frac{X_{(n)}}{q_2}, \frac{X_{(n)}}{q_1}\right)$$

Για να βρω δ.ε ελαχίστου μήκους αρκεί να βρω τα q_1, q_2 με $q_1, q_2 \in (0,1)$ και $q_1 < q_2$ που ελαχιστοποιούν το μήκος

$$l = \frac{X_{(n)}}{q_1} - \frac{X_{(n)}}{q_2} = X_{(n)}\left(\frac{1}{q_1} - \frac{1}{q_2}\right)$$ ή κάνω ελάχιστο το $l^* = \frac{1}{q_1} - \frac{1}{q_2}$

υπό του περιορισμό ότι $P(q_1 < Q < q_2) = 1-a$ ή ισοδύναμα

$$1-a = \int_{q_1}^{q_2} f_Q(q)\,dq = \int_{q_1}^{q_2} n q^{n-1}\,dq = q^n \Big|_{q_1}^{q_2} = q_2^n - q_1^n$$

Άρα ζητώ q_1, q_2 με $q_1, q_2 \in (0,1)$ και $q_1 < q_2$ τέτοια ώστε να ελαχιστοποιείται η $l^* = \frac{1}{q_1} - \frac{1}{q_2}$ (1) υπό του περιορισμό $q_2^n - q_1^n = 1-a$ (2)

Από την (1) έχω: $\frac{dl^*}{dq_1} = -\frac{1}{q_1^2} - \frac{dq_2}{dq_1}\frac{d(1/q_2)}{dq_2} = -\frac{1}{q_1^2} + \frac{1}{q_2^2}\frac{dq_2}{dq_1}$

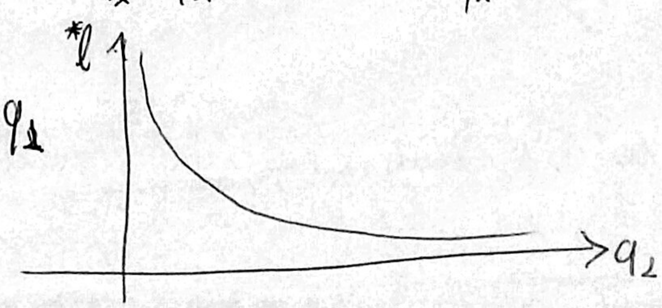
Από την (2) παραγωγίζοντας την ως προς q_1 έχω:

$$\frac{dq_2}{dq_1}\cdot\frac{dq_2^n}{dq_2} - nq_1^{n-1} = 0 \Rightarrow \frac{dq_2}{dq_1}\cdot nq_2^{n-1} = nq_1^{n-1} \Rightarrow$$

$$\Rightarrow \frac{dq_2}{dq_1} = \left(\frac{q_1}{q_2}\right)^{n-1}$$

Άρα $\frac{dl^*}{dq_1} = -\frac{1}{q_1^2} + \frac{1}{q_2^2}\left(\frac{q_1}{q_2}\right)^{n-1} \xRightarrow[q_1 < q_2]{q_1, q_2 \in (0,1)} \frac{dl^*}{dq_1} < 0$

Άρα $l^* \downarrow$ φθίνουσα ως προς q_1



Άρα η ℓ^* θα ελαχιστοποιείται ως προς q_1 στο άνω φράγμα του q_1

Αρκεί να βρω την μεγαλύτερη τιμή του q_2

Από την ② $q_2^n = q_2^n - 1 + a \xrightarrow{q_2 < 1} q_1^n \leq 1 - 1 + a = a \Rightarrow$

$\Rightarrow q_2 \leq \sqrt[n]{a}$

Επομένως ℓ^* φθίνουσα συνάρτηση του q_1 για όλα τα $q_1 \leq \sqrt[n]{a}$

Άρα ℓ^* ελαχιστοποιείται για $q_1 = \sqrt[n]{a}$ και από την ②

$q_2^n = 1 - a + (\sqrt[n]{a})^n = 1 \Rightarrow q_2 = 1$

Τελικά το δ.ε ελαχίστου μήκους είναι $\left(X_{(n)}, \dfrac{X_{(n)}}{\sqrt[n]{a}}\right)$

Παρατήρηση

Και των ίσων ουρών και του ελαχίστου μήκους εξαρτώνται από το επαρκές $X_{(n)}$. Διαισθητικά είναι αποδεκτό και καλύτερο του ελαχίστου μήκους.

III Έλεγχοι Στατιστικών Υποθέσεων - Στατιστικά Τεστ.

Έστω τ.δ $\underset{\sim}{X} = (X_1, \ldots, X_n)$ από πληθυσμό $f(x, \theta)$ $\theta \in \Theta \subseteq \mathbb{R}$

(I) Εκτίμηση σε σημείο : $T(\underset{\sim}{X}) \simeq \theta$ (ή $g(\theta)$)

(II) Εκτίμηση σε δ.ε $(L(\underset{\sim}{X}), U(\underset{\sim}{X})) : P(L(X) < \theta < U(X)) = 1 - a$

Έννοιες – Ορολογία

1. Στατιστικές υποθέσεις

$H_0: \theta = \theta_0$ (θ_0 γνωστό) , $H_a: \begin{cases} \theta > \theta_0 \\ \theta = \theta_0 \\ \theta < \theta_0 \end{cases}$

ή

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

$H_0 \leftarrow$ μηδενική υπόθεση (είναι η υπόθεση που ευελπιστούμε να απορριφθεί)

$H_a \leftarrow$ εναλλακτική υπόθεση

Μια υπόθεση θα λέγεται απλή όταν προσδιορίζει απολύτως την κατανομή $f(x,\theta)$ (π.χ η $H_0: \theta = 5$)

Αλλιώς θα λέγεται σύνθετη υπόθεση (π.χ η $H_0: 0 < \theta < 5$)

Στατιστικό Τεστ: Εννοούμε μια μεθοδολογία/τεχνική που θα μου δώσει την δυνατότητα να αποφασίσω αν θα απορρίψω ή όχι την H_0

2. Για να φτιάξουμε ένα στατιστικό τεστ θα στηριχτούμε στο τ.δ.

Στατιστική Συνάρτηση του Τεστ (Σ.Σ.Τ)

Κάθε συνάρτηση του τ.δ $\underset{\sim}{X} = (X_1, \ldots, X_n)$ και της παραμέτρου θ

ΣΣΤ $\rightsquigarrow T(\underset{\sim}{x}, \theta)$

3. Κρίσιμη Περιοχή (κ.π)

Κρίσιμη περιοχή εννοούμε ένα υποσύνολο του συνόλου τιμών της $T(\underset{\sim}{x}, \theta)$

$$C = \{\underline{x} : T(\underline{x},\vartheta) \in A \subseteq \mathbb{R}\} \left(= \{\underline{x} : T(\underline{x},\vartheta) \gtrless k\} \right)$$

Κρίσιμο σημείο (κ.σ)

4. Κανόνας λήψης απόφασης

Αν η τιμή της ΣΣΤ περιλαμβάνεται στην κ.π => Απορρίπτω Ho

Αλλιώς δεν μπορώ να απορρίψω την Ho