12/5/2017

Στατιστική Συμπερασματολογία

Κατασκευή Αντιστρεπτών Ποσοτήτων

ΕΡΩΤΗΜΑ: Πως βρίσκουμε αντιστρεπτές ποσότητες για κατανομές εκτός κανονικής;

ΠΡΟΤΑΣΗ ① (α) Έστω τ.μ. X με συνεχή κατανομή F_θ, $\theta \in \Theta$. Τότε η τ.μ. $Y = F_\theta(X) \sim U(0,1)$ $\forall \theta \in \Theta$

(β) Η τ.μ. $Z = -2\log F_\theta(X) \sim \chi^2_2$

(γ) Η τ.μ. $W = -2\log(1-F_\theta(X)) \sim \chi^2_2$

ΠΡΟΤΑΣΗ ② Έστω τυχ. δείγμα $X_1, \dots, X_n$ από συνεχή κατανομή F_θ, $\theta \in \Theta$. Τότε οι τ.μ.

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

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

Οι Q, Q^* αντιστρεπτές ποσότητες

ΑΠΟΔΕΙΞΗ

(α) $\longleftarrow$ Απόδειξη στο 1ο μάθημα

(β) Έστω $Y = F_\theta(X) \sim U(0,1) \rightsquigarrow f_Y(y) = \begin{cases} 1, & 0<y<1 \\ 0, & \text{αλλού} \end{cases}$

Θεωρώ το μετασχηματισμό

$\zeta = h(y) = -2\log y$, Αφού $y \in (0,1)$ το $\zeta > 0$

Ο h είναι αντιστρέψιμος

$y = h^{-1}(\zeta) = e^{-\zeta/2}$

$\frac{d h^{-1}(\zeta)}{d\zeta} = -\frac{1}{2} e^{-\zeta/2} \neq 0$ και συνεχής $\zeta > 0$

Άρα $f_Z(\zeta) = f_Y(h^{-1}(\zeta)) \left| \frac{d h^{-1}(\zeta)}{d\zeta} \right| = 1 \left| -\frac{1}{2} e^{-\zeta/2} \right| = \frac{1}{2} e^{-\zeta/2}$, $\zeta > 0$

Άρα η τ.μ. $Z = -2\log F_\theta(X) \sim \text{Εκθ}(\tfrac{1}{2}) \overset{(*)}{\equiv} G(1,2) \overset{(**)}{\equiv} \chi^2_2$

Γνωρίζουμε $\text{Εκθ}(\lambda) \equiv G(1, 1/\lambda)$ $(*)$

$G(\tfrac{n}{2}, 2) \equiv \chi^2_n$ $(**)$

Απόδειξη Πρόταση 2 (α)

$$Q^* = -2\sum_{i=1}^{n} \log F_\theta(x_i) = \sum_{i=1}^{n}(-2\log F_\theta(x_i)) \overset{\text{ΠΡΟΤΑΣΗ 1Β}}{\sim} \sum_{i=1}^{n} \chi^2_2 \overset{X_i \text{ ανεξ}}{\equiv} \chi^2_{\sum_{i=1}^{n} 2} \equiv \chi^2_{2n} \quad \square$$

ΠΑΡΑΔΕΙΓΜΑ 1 Έστω τ.δ. $X_1, \ldots, X_n$ από πληθυσμό με κατανομή $f_\theta(x) = \theta x^{\theta-1}$, $\theta > 0$, $0<x<1$ (Βήτα$(\theta, 1)$). Να κατασκευαστεί δ.ε. ίσων ουρών για θ

Πρέπει να βρω αντιστρεπτή ποσότητα

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

Απαιτείται η $F_\theta(x)$

$$F_\theta(x) \overset{\text{ορ}}{=} \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} \int_{-\infty}^{x} f_X(t)\,dt, & -\infty < x < 0 \\ \int_{-\infty}^{x} f_X(t)\,dt, & 0<x<1 \\ \int_{-\infty}^{x} f_X(t)\,dt, & x>1 \end{cases} = \begin{cases} 0, & -\infty < x < 0 \\ \int_0^x \theta t^{\theta-1}\,dt, & 0<x<1 \\ \int_0^1 \ldots\,dt, & x>1 \end{cases} \Rightarrow F_\theta(x) = \begin{cases} 0, & -\infty<x<0 \\ x^\theta, & 0<x<1 \\ 1, & x>1 \end{cases}$$

Επομένως

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

Αφού Q αντιστρεπτό $\exists\, q_1, q_2$ με $0<q_1, q_2$ τέτοια ώστε

$$1-a = P(q_1 < Q < q_2) = P(q_1 < -2\theta\sum \log X_i < q_2) \Rightarrow$$

$$\Rightarrow 1-a = P\left(\frac{q_1}{-2\sum\log X_i} < \theta < \frac{q_2}{-2\sum \log X_i}\right)$$

Άρα το

$$\left(\frac{q_1}{-2\sum\log X_i}, \frac{q_2}{-2\sum\log X_i}\right)$$ είναι ένα $100(1-a)\%$ δ.ε για θ.

Παρατήρηση: Το δ.ε. έχει ότι είναι συναρτήσεις της επαρκούς $\prod_{i=1}^{n} X_i$

δε ίσων ουρών

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

Για το ίσων ουρών δ.ε. τα q_1, q_2 θα πρέπει:

$$P(Q>q_2) = \frac{a}{2} = P(Q<q_1)$$

$$P(Q>q_2) = \frac{a}{2} \Rightarrow P(\chi^2_{2n} > q_2) = a/2 \Rightarrow q_2 = \chi^2_{2n, a/2}.$$

$$P(Q<q_1) = a/2 = 1-P(Q>q_1) = a/2$$

$$\Rightarrow P(Q>q_1) = 1-a/2 \quad q_1 = \chi^2_{2n, 1-a/2}$$

Παράδειγμα: 2

Έστω $X_1,\dots,X_n$ τ.δ. από $U(0,\theta)$

Να κατασκευαστεί δ.ε. ίσων ουρών για θ.

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & 0<x<\theta \\ 0, & \text{αλλού} \end{cases} \qquad F_\theta(x) = \begin{cases} 0, & x<0 \\ \frac{x}{\theta}, & 0<x<\theta \\ 1, & x>\theta \end{cases}$$

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

$Q = -2\sum \log X_i + 2n\log\theta \sim \chi^2_{2n}$

άρα Q αντιστρεπτή.

Αφού η Q αντιστρεπτή, $\exists\, q_1, q_2 : 0<q_1<q_2$

$1-a = P(q_1<Q<q_2) = P(q_1 < -2\sum \log X_i + 2n\log\theta < q_2)$

$= P\left(e^{q_1/2n}\left(\prod_{i=1}^{n} X_i\right)^{1/n} < \theta < e^{q_2/2n}\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right)$

Άρα ένα $100(1-a)\%$ δε είναι

$$\left(e^{\frac{q_1}{2n}}\left(\prod_{i=1}^{n} X_i\right)^{1/n},\ e^{\frac{q_2}{2n}}\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right)$$

και το ίσων ουρών

$$\left(e^{\frac{\chi^2_{2n,1-\frac{a}{2}}}{2n}}\left(\prod_{i=1}^{n} X_i\right)^{1/n},\ e^{\frac{\chi^2_{2n,a/2}}{2n}}\left(\prod_{i=1}^{n} X_i\right)^{1/n}\right)$$

Παρατήρηση: Ικανοποιημένοι; ΟΧΙ $\longrightarrow$ δεν εξαρτάται από το επαρκές $X_{(n)}$ που συμβαίνει ε- είναι ο ΕΜΠ και ο ΑΟΕΔ

Παραδειγμα 3 :

Εστω τ.δ. $X_1, \dots, X_n$ από $U(0,\theta)$

Εστω $X_{(n)} = \max\{X_1, \dots, X_n\}$

(α) Νδο η $Q = \frac{X_{(n)}}{\theta}$ είναι αντιστρεπτή

(β) Να βρεθούν δ.ε ίσων ουρών για θ

(γ) -ιι- -ιι- δ.ε ελαχίστου μήκους για θ.

Αρα $f(x) = \begin{cases} 1/\theta, & 0<x<\theta \\ 0, & \text{αλλιώ} \end{cases}$ $\quad F(x) = \begin{cases} 0 & x<0 \\ \frac{x}{\theta}, & 0<x<\theta \\ 1, & x>\theta \end{cases}$

Θυμίζω: $f_{X_{(n)}}(x) = n\left[F(x)\right]^{n-1} f(x)$

(α) Εστω $Y = X_{(n)} \rightsquigarrow$ η κατανομή της Y είναι:

$$f_Y(y) = n\left[\frac{y}{\theta}\right]^{n-1} \frac{1}{\theta} \Rightarrow$$

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

Θεωρώ το μετασχηματισμό

$q = h(y) = \frac{y}{\theta}, \quad q \in (0,1)$

$y = h^{-1}(q) = \theta q$

$\frac{dh^{-1}(q)}{dq} = \theta \neq 0$ συνεχής

Αρα $f_Q(q) = f_Y(h^{-1}(q)) \left|\frac{dh^{-1}(q)}{dq}\right|$

$\Rightarrow f_Q = n\frac{(\theta q)^{n-1}}{\theta^n}\theta$

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

Συνοπτικά

$f_Q(q) = \begin{cases} nq^{n-1}, & 0<q<1 \\ 0, & \text{αλλιώ} \end{cases}$

η f_Q είναι ανεξάρτητη της θ

Αρα Q αντιστρεπτή

(α) $Q = \frac{X_{(n)}}{\theta}$ αντιστρεπτή αφού

$$f_Q(q) = \begin{cases} n q^{n-1}, & 0<q<1 \\ 0, & \text{αλλού} \end{cases}$$

(ε) Το δ.ε. ίσων ουρών είναι $\left(\frac{X_{(n)}}{\sqrt[n]{1-\frac{a}{2}}}, \frac{X_{(n)}}{\sqrt[n]{a/2}} \right)$ — Αποτέλεσμα

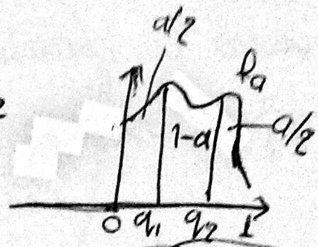
(β) Αφού $Q = \frac{X_{(n)}}{\theta}$ αντιστρεπτή $\exists q_1, q_2$ με $0<q_1<q_2<1$ και τέτοια ώστε:

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

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

Άρα το $\left(\frac{X_{(n)}}{q_2}, \frac{X_{(n)}}{q_1} \right)$ είναι ένα $100(1-a)\%$ δ.ε. για θ.

Το ίσων ουρών που ζητάμε, προκύπτει για q_1, q_2



$$P(Q>q_2) = \frac{a}{2} = P(Q<q_1)$$

$$\frac{a}{2} = P(Q>q_2) = \int_{q_2}^{1} f_Q(q)\,dq = \int_{q_2}^{1} n q^{n-1}\,dq = q^n \Big|_{q_2}^{1} = 1-q_2^n \Rightarrow q_2 = \sqrt[n]{1-\frac{a}{2}}$$

$$\frac{a}{2} = P(Q<q_1) = \int_0^{q_1} f_Q(q)\,dq = \int_0^{q_1} n q^{n-1}\,dq = q^n \Big|_0^{q_1} = q_1^n \Rightarrow q_1 = \sqrt[n]{a/2}$$

(γ) Το μήκος του $\left(\frac{X_{(n)}}{q_2}, \frac{X_{(n)}}{q_1} \right)$ είναι $g = X_{(n)}\left(\frac{1}{q_1} - \frac{1}{q_2}\right)$

Άρα ζητείται ελαχ. του g ως προς q_1, q_2 ή ισοδύναμα του

$g^* = \frac{1}{q_1} - \frac{1}{q_2}$ υπό την συνθήκη $\int_{q_1}^{q_2} f_Q(q)\,dq = 1-a$

$$\frac{dg^*}{dq_1} = \frac{d}{dq_1}\left(\frac{1}{q_1}\right) - \frac{d}{dq_1}\left(\frac{1}{q_2}\right)$$
$$= -\frac{1}{q_1^2} - \frac{d}{dq_2}\left(\frac{1}{q_2}\right)\frac{dq_2}{dq_1}$$

$$\frac{dg^*}{dq_1} = -\frac{1}{q_1^2} + \frac{1}{q_2^2}\frac{dq_2}{dq_1}$$

Από περιορισμό: $1-a = \int_{q_1}^{q_2} n q^{n-1}\,dq \Rightarrow 1-a = q_2^n - q_1^n \quad (1)$

Παραγωγίζοντας την (1)

$\frac{d}{dq_1}(1-a) = \frac{d}{dq_1} q_2^n - \frac{d}{dq_1} q_1^n$

$\Rightarrow \frac{d}{dq_2}(q_2^n)\frac{dq_2}{dq_1} - nq_1^{n-1} = 0$

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

$\Rightarrow \frac{dq_2}{dq_1} = \frac{q_1^{n-1}}{q_2^{n-1}}$ (2)

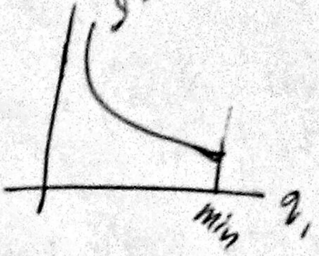
Άρα από (2) η

$\frac{dg^*}{dq_1} = -\frac{1}{q_1^2} + \frac{1}{q_2^2}\frac{q_1^{n-1}}{q_2^{n-1}}$

$\frac{dg^*}{dq_1} = 0 \Rightarrow q_1^{n+1} = q_2^{n+1} \Rightarrow q_1 = q_2$ αδύνατο γιατί $q_1 < q_2$.

Άρα έτσι δεν μπορώ να βρω ελάχιστο.

Ας κοιτάξουμε μονοτονία g^* ως προς q_1



Παρατηρώ ότι: $\frac{dg^*}{dq_1} < 0$ γιατί $q_1 < q_2$

Άρα η g^* φθίνουσα συνάρτηση του q_1

Άρα θέλω να βρω άνω φράγμα για το q_1

Από την (1)

$q_1^n = q_2^n - 1 + a \xRightarrow{q_2 < 1} q_1^n < 1 - 1 + a = a$

$\Rightarrow q < \sqrt[n]{a}$

Αφού η $g^* \downarrow$ φθίνουσα συνάρτηση του q_1, ελαχιστοποιείται ως προς q_1 για $q_1 = \sqrt[n]{a}$

Για $q_1 = \sqrt[n]{a}$ από την (1) το $q_2 = 1$

Το δε ελάχ. μήκος είναι:

$\left(X(n), \frac{X(n)}{\sqrt[n]{a}} \right)$

ΕΛΕΓΧΟΙ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ
—
ΣΤΑΤΙΣΤΙΚΑ ΤΕΣΤ

Παιδίατρος: Γνωρίζει από τη βιβλιογραφία ότι το μέσο βάρος παιδιών 10 ετών είναι 40kgr.

Από την εμπειρία του μπορεί να του δημιουργείται η εντύπωση ότι το δεδομένο αυτό έχει αλλάξει

Δίλημμα: η υποψία ή εικασία ή άποψη ή ιδέα

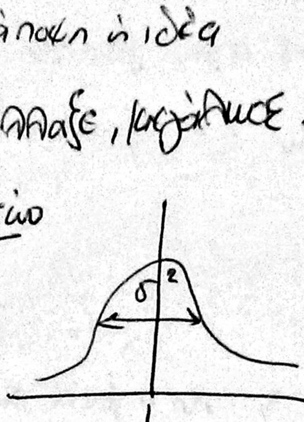
Το βάρος είναι 40kgr ή άλλαξε, μεγάλωσε

Πληθυσμός: Παιδιά ηλικίας 10 ετών

τ.μ.: → X ← βάρος

↑

κατανομή



⟹ $H_0: \mu = 40kg$ μηδενική υπόθεση $H_a: \mu > 40kg$ εναλλακτική υπόθεση

↑

είναι εκείνη η υπόθεση που έχει την εντύπωση δεν πρέπει να ~~...~~ απορριφθεί

Για να επιλέξω ανάμεσα στην H_0 ή την H_a πρέπει να διαθέτω ένα τ.δ.
$X_1, X_2, \dots, X_n$ από $N(\mu, \sigma^2)$

Φορμαλισμός

Έστω τ.δ. $x_1, \dots, x_n$ από πληθυσμό με κατανομή $f(x,\theta)$, $\theta \in \Theta \subseteq \mathbb{R}$

Έστω προς έλεγχο το ζεύγος υποθέσεων

$H_0: \theta = \theta_0$ (θ_0 γνωστό) έναντι της $H_a: \theta = \theta_a$ ($\theta_a \neq \theta_0$)

μηδενική — εναλλακτική

Αν οι υποθέσεις προσδιορίζουν πλήρως την κατανομή του πληθυσμού λέγονται <u>απλές</u>

(π.χ. Αν ο πληθυσμός $f(x) = \lambda e^{-\lambda x}$ και η $H_0: \lambda = 4$

τότε η f προσδιορίζεται πλήρως αφού γίνεται $f(x) = 4e^{-4x}$)

Μια υπόθεση λέγεται <u>σύνθετη</u> αν δεν προσδιορίζεται πλήρως η κατανομή του πληθυσμού

(π.χ. $H_0: \lambda > 4$)

Το σημείο αναφοράς είναι το τ.δ. $x_1 \dots x_n$ ή μια συνάρτηση του

Ονομάζουμε στατιστική συνάρτηση του τεστ (σ.σ.τ) μια συνάρτηση του τυχαίου δείγματος τ.δ και (πιθανώς) της παραμέτρου, η οποία θα χρησιμοποιηθεί για την επιλογή των υποθέσεων

Κρίσιμη περιοχή: λέγεται μια περιοχή στο σύνολο τιμών της σ.σ.τ που οδηγεί στην απόρριψη της H_0 με βάση το εξής κριτήριο:

- Αν η τιμή της στατιστικής συνάρτησης του τεστ, περιλαμβάνεται στην Κ.Π

όπου (Κ.Π: Κρίσιμη Περιοχή) τότε απορρίπτουμε την H_0 ενώ αν η τιμή της ~~στατιστικής~~ σ.σ.τ. δεν περιλαμβάνεται στην Κ.Π. τότε δεν μπορούμε να απορρίψουμε την H_0.