Μαθημα 3ο

6/3/19

Το τεστ του Kolmogorov - Smirnov

χ^2 τεστ
παραδ. 3-4
(παραδ. 1 5, 2 6
συμπεριφοράς ...)

Εστω $X_1, \ldots, X_n$ τ.δ. απο πληθυσμό με α.σ.κ. $F(x)$.
Ενδιαφέρει ο ελεγχος $H_0: F(x) = F_0(x)$
$H_1: F(x) \neq F_0(x)$

Ο ελεγχος γινεται με το στατιστικό $D_n = \sup_x |F_n(x) - F_0(x)|$
οπου $F_n(x) = (\text{αριθμ. τεαρατ. } x_i \leq x)/n =$
$= \frac{1}{n} \sum_1^n I(x_i \leq x)$ η ε.α.σ.κ

Για επιπεδο σημαντικοτητας a, η H_0 απορρ. οταν $D_n \geq D_{a,n}$ οπου $P(D_n \geq D_{a,n}) = a$ κ οι τιμές $D_{a,n}$ απο πινακα

$\Rightarrow D_n = \max_x |F_n(x) - F_0(x)|$

παράδειγμα 1 (7.3) (στο φυλλάδιο 3ο)

Εχει μπει παράδειγμα ίδιο με τα εξής:
17, 25, 5, 13 ...
$U(0,30)$

$$F(x) = \begin{cases} 0, & x<0 \\ \frac{x}{30}, & 0<x<30 \\ 1, & x \geq 30 \end{cases}$$

x_i : 5 13 17 22

$F_u(x)$: 1/10 2/10

$F_0(x)$: 5/30 13/30 17/30 ...

↳ είναι πιθανότητα.

παράδειγμα 2

33.3, 31.4, 33.4, 33.5, 37.0, 36.2, 34.9, 33.7, 34.4

31.0 άρα $n=10$.

H_0: $F(x) = F_0(x) \equiv N(\mu=32, \sigma^2=3.24)$

H_a: $F(x) \neq F_0(x)$

K-S ;

λύση

x_i :	31.0	31.4	33.3	33.4	33.5	33.7	34.4	34.9	36.2	37.0
$F_{10}(x_i)$:	1/10	2/10	3/10	4/10	5/10	6/10	7/10	8/10	9/10	1
$F_0(x_i)$:	0.2912	0.3707	0.7642	0.7823	0.7967	0.8264	0.9082	0.9463	0.9801	0.9973
$D_{10} = \lvert F_{10}(x_i) - F_0(x_i)\rvert$:	0.2912		0.4642							0.0027

$$F_0(31.0) = P(X \leq 31.0 \mid X \sim N(\mu=32, \sigma^2=3.24)) =$$

$$= P\left(\frac{X-\mu}{\sigma} (= Z) \leq \frac{31-32}{\sqrt{3.24}} \mid Z \sim N(0,1)\right) =$$

$$= P(Z \leq -0.55) = 0.5 - 0.2088 = 0.2912$$

$$D_n = 0.4642 > 0.4642 \; (= D_{10, 0.05})$$

απορρ. H_0.

Προσημικό τεστ (Sign test) (ανάλογο του t-τεστ στη παραμ. στατιστ.)

(θέλουμε συνεχή κ συμμετρικό πληθυσμό)

Έστω $x_1, .., x_n$ τ.δ. από πληθυσμό συνεχή κ συμμετρικό σ.δ. $P(x<\mu) = P(x>\mu) = \frac{1}{2}$, $P(x=\mu)=0$ και

Αν μόνο συνεχής τότε η $P(x<\tilde{\mu}) = P(x>\tilde{\mu}) = \frac{1}{2}$ διάμεσο

Για να ελέγξουμε υποθέσεις της μορφής: $H_0: \mu=\mu_0$ (ή $\tilde{\mu}$)

$H_a: \mu \neq \mu_0$
ή
$H_a: \mu<\mu_0$
ή
$H_a: \mu>\mu_0$

αντικαθιστάμε κάθε μια από τις παρατηρήσεις (μετρήσεις) με ένα + ή − ανάλογα αν είναι μεγαλύτερη ή μικρότερη από το μ_0. Αν είναι ίση με μ_0 την διαγράφουμε και μειώνουμε το n.
Αν η H_0 συνεχής, τα + και − θα πρέπει να είναι εξίσου κατανεμημένα.
Έτσι για X = πλήθος των σημείων + στο δείγμα,
$X \sim Bin(n, p=\frac{1}{2})$ όταν H_0 αληθής και ο αρχικός έλεγχος ισοδύναμα μεταφέρεται: $H_0: p=\frac{1}{2}$ v.
i) $H_a: p\neq\frac{1}{2}$ ή ii) $H_a: p<\frac{1}{2}$ ή iii) $H_a: p>\frac{1}{2}$

Οι κρίσιμες περιοχές είναι (αντίστοιχα):

$X \leq k'_{a/2}$ ή $X \geq k_{a/2}$ (i)

$X \leq k'_a$ (ii)

$X \geq k_a$ (iii)

όπου k'_a κ k_a ο μεγαλύτερος κ μικρότερος αντίστοιχα ακέραιος για τους οποίους:

$$\sum_{y=0}^{k_a}\binom{n}{y}\left(\frac{1}{2}\right)^n \leq a \quad \text{και} \quad \sum_{y=k_a}^{n}\binom{n}{y}\left(\frac{1}{2}\right)^n \leq a, \text{ όπου } X\sim B(n,\tfrac{1}{2})$$

$X \sim Bin(n, \frac{1}{2}) \Rightarrow X \overset{\text{προσ.}}{\sim} N\left(\frac{n}{2}, \frac{n}{4}\right)$ για σχετικά μεγάλα n.

$$Z = \frac{X - \frac{n}{2} - \frac{1}{2}}{\sqrt{n}/2}, \quad X > n/2$$

διόρθωση συνέχειας

$$= \frac{X - \frac{n}{2} + \frac{1}{2}}{\sqrt{n}/2}, \quad X < \frac{n}{2}$$

$[\overset{\text{προσ.}}{\sim} N(0,1)]$ και κρ. περιοχή $|Z| \geq Z_{\frac{a}{2}}$ (i)

$Z \leq -Z_a$ (ii)

$Z \geq Z_a$ (iii).

Παράδειγμα 1 (παρ. 2 (3.1 σημειώσεις Μπατσίδη))

n=19: 163 (+), 165 (+), ~~160~~, 189 (+), 161 (+), 171 (+), 158 (−), 169 (+), 162 (+), 163 (+),
139 (−), 172 (+), 165 (+), 148 (−), 166 (+), 172 (+), 163 (+), 187 (+), 170 (+)

$H_0: \mu = 160$ ~~v~~ $\mu \leq 160$ $H_a: \mu > 160$ $(a = 0.05)$

$n = 18$, $X = 15$ ↳ "+"

κρ. περιοχή από το (iii) $Z \geq Z_{0.05} = 1.645$

$$Z = \frac{X - \frac{n}{2} - \frac{1}{2}}{\sqrt{n}/2} = \frac{15 - \frac{18}{2} - \frac{1}{2}}{\sqrt{18}/2} = 2.59$$

Άρα απορ. η H_0