NO.

Date 03/05/2018

Μάθημα 9ο

Για μεγάλο μέγεθος δείγματος και υπό ορισμένες συνθήκες ισχύει ότι:

$$\hat{\theta} \overset{\text{προσ.}}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

(θ μονοδιάστατη παράμετρος)

↑ εκτιμητής μέγιστης πιθανοφάνειας

εάν πολυδιάστατο θα υπήρχε πίνακας

όπου $\underline{\underline{nI(\theta)}} = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(\underset{\sim}{x}, \theta)\right)$

με $f(\underset{\sim}{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$

Ξέρουμε: (1) $\dfrac{\hat{\theta} - \theta}{\sqrt{\frac{1}{nI(\theta)}}} \overset{\text{προσ.}}{\sim} N(0,1)$ προσεγγίζω

(2) $P\left(-z_{a/2} \leq \dfrac{\hat{\theta} - \theta}{\sqrt{\frac{1}{nI(\theta)}}} \leq z_{a/2}\right) \simeq 1-a$ απομονώνω το θ

Ξεκινώντας από αυτό το αποτέλεσμα προσπαθώ ν' αντιστρέψω.

Μερικές φορές αντί για $nI(\theta)$ βάζω $nI(\hat{\theta})$

→ (Το αποτέλεσμα δεν αλλάζει καθώς απλά η σύγκλιση είναι ίσως πιο αργή)

NO.

Date

Άσκηση

$X_1, X_2, \ldots, X_n$, ικανοποιητικά μεγάλο τυχαίο δείγμα από την εκθετική: $f(x,\theta) = \theta \cdot e^{-\theta x}$, $x > 0$

Θέλουμε να προσδιορίσουμε προσεγγιστικό Δ.Ε. για το θ.

Λύση

Πρέπει να βρω εκτίμηση μέγιστης πιθανοφάνειας.

$$\prod f(x_i,\theta) = \theta^n e^{-\theta \sum x_i}$$

$$\log \prod f(x_i,\theta) = n\log\theta - \theta \sum x_i$$

$$\frac{\partial}{\partial\theta} \log \prod f(x_i,\theta) = \frac{n}{\theta} - \sum x_i$$

$$\frac{\partial^2}{\partial\theta^2} \log \prod f(x_i,\theta) = -\frac{n}{\theta^2} < 0$$

$$\frac{\partial}{\partial\theta} \log \prod f(x_i,\theta) = 0 \Rightarrow \frac{n}{\hat\theta} = \sum x_i \Rightarrow \boxed{\hat\theta = \frac{n}{\sum x_i} = \frac{1}{\bar x}}$$

Τώρα, θέλω να βρω το $nI(\theta)$

Είναι: $$nI(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(\underline{x},\theta)\right) =$$

$$= -E\left(-\frac{n}{\theta^2}\right) = \frac{n}{\theta^2}$$

Επομένως: $$\frac{\hat\theta - \theta}{\sqrt{\frac{1}{nI(\theta)}}} = \frac{\hat\theta - \theta}{\sqrt{\theta^2/n}} \overset{\text{προσ.}}{\sim} N(0,1)$$

Επόμενο βήμα → παίρνουμε τη σχέση (2):

NO.

Date

$$P\left(-z_{\alpha/2} \leq \frac{(\hat{\theta}-\theta)\sqrt{n}}{\theta} \leq z_{\alpha/2}\right) \simeq 1-\alpha$$

$$P\left(-z_{\alpha/2}\frac{1}{\sqrt{n}} \leq \frac{\hat{\theta}}{\theta} - 1 \leq z_{\alpha/2}\frac{1}{\sqrt{n}}\right) \simeq 1-\alpha$$

$$P\left(-z_{\alpha/2}\frac{1}{\sqrt{n}} + 1 \leq \frac{\hat{\theta}}{\theta} \leq z_{\alpha/2}\frac{1}{\sqrt{n}} + 1\right) \simeq 1-\alpha$$

$$P\left(\frac{\hat{\theta}}{z_{\alpha/2}\frac{1}{\sqrt{n}}+1} \leq \theta \leq \frac{\hat{\theta}}{1-z_{\alpha/2}\frac{1}{\sqrt{n}}}\right) \simeq 1-\alpha$$

Βρήκαμε το προσεγγιστικό διάστημα εμπιστοσύνης (ελαχίστου μήκους) με $\hat{\theta} = \frac{1}{\bar{x}}$

Άσκηση 2.14 βιβλίου

$X_1, X_2, \dots, X_n$ από έναν πληθυσμό που ακολουθεί εκθετική : Εκθ $(1/\lambda)$

Η άσκηση είναι ακριβώς ίδια με την προηγούμενη.

$\boxed{n \text{ πολύ μεγάλο} \Rightarrow \text{πρέπει να βρω προσεγγιστικό Δ.Ε.}}$

- $f(x,\lambda) = \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}, \quad x \geq 0$

- $\prod f(x_i,\lambda) = \frac{1}{\lambda^n} e^{-1/\lambda \, \Sigma x_i}$

- $\log \prod f(x_i,\lambda) = -n\log\lambda - \frac{1}{\lambda}\Sigma x_i$

NO.

Date

- $\frac{\partial}{\partial \lambda} \log \prod f(x_i,\lambda) = \frac{-n}{\lambda} + \frac{1}{\lambda^2} \sum X_i$

- $\frac{\partial^2}{\partial \lambda^2} \log \prod f(x_i,\lambda) = \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum X_i$

- $\frac{\partial}{\partial \lambda} \log \prod f(x_i,\lambda) = 0 \Rightarrow \boxed{\hat{\lambda} = \frac{\sum X_i}{n} = \bar{X}}$ ⊛

- $\left. \frac{\partial^2}{\partial \lambda^2} \log \prod f(x_i,\lambda) \right|_{\hat{\lambda}=\bar{X}} = \frac{n}{\bar{X}^2} - \frac{2}{\bar{X}^3} n\bar{X} = \frac{-n}{\bar{X}^2} < 0$

Το ⊛ αντιστοιχεί σε σημείο μεγίστου
Άρα εκτιμητής μέγιστης πιθανοφάνειας (και όχι ελάχιστης)

- $n I(\lambda) = -E\left(\frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum X_i\right) = -\left(\frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum E X_i\right) =$
$= -\left(\frac{n}{\lambda^2} - \frac{2}{\lambda^3} n\lambda\right) = \frac{n}{\lambda^2}$

... Συνεχίζουμε όπως πριν...

$\frac{\sum X_i}{n}$ → έχουμε βρει προσεγγιστικό το ΚΟΘ.

Διότι στη συγκεκριμένη άσκηση θα μπορούσαμε να πάρουμε:

$$\frac{\frac{\sum X_i}{n} - \mu}{\sqrt{\sigma^2/n}} \underset{\sim}{\text{προς}} N(0,1)$$

NO.

Date

Άσκηση 2.7

$a = 5\%$, $\theta = 3\mu + 7$,

$X_1 = 8$, $X_2 = 14$, $X_3 = 11$ από την $N(\mu, \sigma^2)$, σ^2 άγνωστο

Όταν $X_1, X_2, \ldots, X_n$ τ.δ. $N(\mu, \sigma^2)$, σ^2 άγνωστη τότε το $100(1-a)\%$ Δ.ε. ελαχίστου μήκους για το μ είναι το:

$$\Big(\underbrace{\bar{X} - t_{n-1, a/2} \frac{S}{\sqrt{n}}}_{= L}, \underbrace{\bar{X} + t_{n-1, a/2} \frac{S}{\sqrt{n}}}_{= U}\Big)$$

$$3\mu + 7 \in (3L + 7, 3U + 7)$$

Δ.ε. για το $3\mu + 7 = \theta$

Πρέπει να υπολογίσω τα $\bar{X}$, S και τελείωσα

(και το $t_{n-1, a/2}$)

Άσκηση 2.8

5 τυχαία ανεξάρτητα δείγματα από $N(\mu_i, \sigma^2)$, $i = 1, .., 5$

όπου μ_i, σ^2 άγνωστα.

Δίνεται:

n_i	6	4	3	7	8
S_i^2	40	30	20	42	50

Να κατασκευάσω 98% ($\Rightarrow a = 2\%$) Δ.ε για το σ^2

το κοινό

Πρέπει να χρησιμοποιήσουμε όλη την πληροφορία για την κοινή παράμετρο σ^2 (5 ΑΣ και τα 5 δείγματα)

Ισχύει: (*) $\frac{(n_1-1)S_1^2}{\sigma^2} \sim X^2_{n_1-1}$

Θα το γράψω και για τα υπόλοιπα 4.

$\frac{(n_2-1)S_2^2}{\sigma^2} \sim X^2_{n_2-1}$

$\vdots$

$\frac{(n_5-1)S_5^2}{\sigma^2} \sim X^2_{n_5-1}$

Ανεξάρτητα μεταξύ τους αφού και δείγματα ανεξάρτητα.
Άρα τα συνδυάζω:

$$\frac{\overbrace{(n_1-1)S_1^2 + \dots + (n_5-1)S_5^2}^{\text{"}A\text{"}}}{\sigma^2} \sim X^2_{\Sigma n_i - 5} \quad (**)$$

Σαν συμπεριφορά για το (**) η απόδειξη είναι ίδια με το (*). Άρα:

$$P\left(q_1 \leq \frac{A}{\sigma^2} \leq q_2\right) = 1-a \Rightarrow$$

$$P\left(\frac{A}{q_2} \leq \sigma^2 \leq \frac{A}{q_1}\right) = 1-a.$$

$$l = A\left(\frac{1}{q_1} - \frac{1}{q_2}\right)$$

Θα συνεχίσω και θα δείξω ότι δε μπορώ να προσδιορίσω Δ.Ε. ελαχίστου μήκους.

NO.

Date

Ασκ. 2.13, 2.14 εκτός.

Δ.Ε. ίσων ουρών:

$$P(\underset{\frac{A}{\sigma^2}}{\overset{\|}{Q}} \leq q_1) = \alpha/2 \Rightarrow q_1 = X^2_{\Sigma n_i - 5, 1-\alpha/2}$$

$$P(Q \geq q_2) = \alpha/2 \Rightarrow q_2 = X^2_{\Sigma n_i - 5, \alpha/2}$$

Άσκηση 2.12

Έστω X μια παρατήρηση από ένα πληθυσμό με σ.π.π. $f(x,\theta) = (1+\theta)x^\theta$, $0 \leq x \leq 1$. Να κατασκευαστεί Δ.Ε.

Όταν είναι συνεχής α.σ.κ. τότε υπάρχει πάντα αντιστρεπτή ποσότητα.

$$F(x,\theta) = \int_0^x (1+\theta) y^\theta dy = x^{\theta+1}, \quad 0 \leq x \leq 1$$

Είχαμε δείξει $\Rightarrow F(X,\theta) \sim U(0,1)$

$\searrow -2\log F(X,\theta) \sim Εκθ(1/2)$

$$\Rightarrow -2\log X^{\theta+1} = -2(\theta+1)\log X \sim Εκθ(1/2)$$

$$P(q_1 \leq -2(\theta+1)\log X \leq q_2) = 1-\alpha$$

$$P\left(-\frac{q_1}{2\log X} \leq \theta+1 \leq \frac{q_2}{-2\log X}\right) = 1-\alpha$$

$X \in [0,1] \Rightarrow \log X < 0$

άρα $-2\log X > 0$

NO.

Date

$$P\left(\frac{q_1}{-2\log X}-1 \le \theta \le \frac{q_2}{-2\log X}-1\right)=1-a$$

$$\ell=\frac{q_2}{-2\log X}-1-\frac{q_1}{-2\log X}+1=\frac{q_2-q_1}{-2\log X}.$$

$$f_Q(q)=\frac{1}{2}e^{-q/2}, \quad q>0 \quad \longleftarrow \text{είναι φθίνουσα}$$

$$(Q=-2(\theta+1)\log X) \qquad \Downarrow$$

! παραγωγίζω ως προς q_1.

$$\frac{d\ell}{dq_1}=\frac{1}{-2\log X}\left(\frac{dq_2}{dq_1}-1\right)$$

Όπως έχουμε δείξει ότι $\frac{dq_2}{dq_1}=\frac{f_Q(q_1)}{f_Q(q_2)}$

<u>Άρα:</u> $$\frac{d\ell}{dq_1}=\frac{1}{-2\log X}\left(\frac{e^{-q_1/2}-e^{-q_2/2}}{e^{-q_2/2}}\right)>0$$

$$\boxed{q_1=0}$$

Για να βρω το q_2: $\int_0^{q_2}\frac{1}{2}e^{-q/2}\,dq=1-a \Rightarrow$

$$\Rightarrow 1-e^{-q_2/2}=1-a \Rightarrow -q_2/2=\ln a \Rightarrow \boxed{q_2=-2\ln a}$$

$\ln a<0$ αφού $a\in(0,1)$

Άρα βρήκαμε Δ.Ε. ελαχίστου μήκους.

o.

ate

Άσκηση 2.4

$X_1, X_2, \dots, X_n$ τ.δ. $U(0,\theta)$

$f(x,\theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta$

$(0, k(X_1+X_2))$ ΔΕ για την θ με πιθ. $1-a$

όπου $k>1$.

Να προσδιοριστεί η σταθερά k.

Λύση

$P(0 < \theta < k(X_1+X_2)) = 1-a$

$P\left(X_1+X_2 > \frac{\theta}{k}\right) = 1-a$

άρα πρέπει να βρω την κατανομή του $U = X_1 + X_2$

(Αν πάω με ροπογεννήτρια δε θα μπορέσω να τη βρω)

$\begin{cases} u = X_1 + X_2 \\ v = X_1 \end{cases} \Rightarrow \begin{cases} X_1 = v \\ X_2 = u - v \end{cases}$

$J = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1 \neq 0$

η από κοινού

$f_{u,v}(u,v) = f_{X_1,X_2}(v, u-v) \cdot |J| =$

$= f_{X_1}(v)\, f_{X_2}(u-v) \cdot 1 = \frac{1}{\theta} \cdot \frac{1}{\theta}, \quad (u,v) \in T$

X_1, X_2 από τυχ. δ. $\Rightarrow$ ανεξάρτητες.

$f(x,\theta) = \frac{1}{\theta} \quad 0 \le x \le \theta$

$$T = \begin{cases} 0 \leq v \leq \theta \\ 0 \leq u - v \leq \theta \\ 0 \leq u \leq 2\theta \end{cases}$$

$(0 \leq X_1 \leq \theta,\ 0 \leq X_2 \leq \theta \Rightarrow 0 \leq X_1 + X_2 \leq 2\theta \Rightarrow 0 < u \leq 2\theta)$

$$f_{u,v}(u,v) = \frac{1}{\theta^2}, \quad (u,v) \in T.$$

$$f_u(u) = \int f_{u,v}(u,v)\, dv$$

Θέλω να απαλλαχτώ από την v

<u>Ισχύει:</u> $\begin{cases} 0 \leq v \leq \theta \\ u-\theta \leq v \leq u \end{cases} \rightarrow$

$$\Rightarrow \max\{0, u-\theta\} \leq v \leq \min\{u, \theta\}$$

$\max\{0, u-\theta\} = 0$ όταν $u \leq \theta$, $= u - \theta$ όταν $u > \theta$

$\min\{u, \theta\} = u$ όταν $u \leq \theta$, $= \theta$ όταν $u > \theta$

$$f_u(u) = \begin{cases} \displaystyle\int_0^u \frac{1}{\theta^2}\, dv & 0 \leq u \leq \theta \\ \displaystyle\int_{u-\theta}^{\theta} \frac{1}{\theta^2}\, dv & \theta < u \leq 2\theta \end{cases}$$

NO.

Date

$$f_U(u) = \begin{cases} \dfrac{u}{\theta^2}, & 0 \le u \le \theta \quad (*) \\[2ex] \dfrac{2\theta-u}{\theta^2}, & \theta \le u \le 2\theta \end{cases}$$

$$P\left(X_1+X_2 \ge \frac{\theta}{k}\right) = 1-a \xrightarrow[\frac{\theta}{k}<\theta]{k>1}$$

$$\Rightarrow 1 - P\left(\underbrace{X_1+X_2}_{"U"} \le \frac{\theta}{k}\right) = 1-a \Rightarrow$$

$$\Rightarrow P\left(U \le \frac{\theta}{k}\right) = a$$

$(\theta/k < \theta)$ είμαστε στο $(*)$

$$\Rightarrow \int_0^{\theta/k} \frac{u}{\theta^2}\,du = a \Rightarrow \boxed{k = \frac{1}{\sqrt{2a}}}$$

Έλεγχος Στατιστικών Υποθέσεων

Μία έννοια που παίζει βασικό ρόλο είναι η μηδενική υπόθεση H_0 και είναι η στατιστική υπόθεση που τίθεται προκειμένου να απορριφθεί.

Εναλλακτική υπόθεση είναι αυτή που τίθεται σε αντίθεση από αυτή που θέλουμε να υποθέσουμε.

Βήματα στατιστικού ελέγχου

1ο) Θα πρέπει να προσδιορίσω μια στατιστική συνάρτηση των $X_1, X_2, \dots, X_n$, και των ποσοτήτων που μας ενδιαφέρουν, και η οποία υπό την μηδενική υπόθεση έχει γνωστή κατανομή.

2ο) Να καθορίσω την περιοχή απόρριψης της μηδενικής υπόθεσης.

C = περιοχή απόρριψης = κρίσιμη περιοχή

		ΑΛΗΘΕΙΑ	
		H_0 αληθής	H_a αληθής
Απόφαση στατιστικού	Απορ. H_0	Σφάλμα τύπου I	—
	Αποδ. H_0	—	Σφάλμα τύπου II

Συμβολίζουμε:

$$\begin{cases} a = P(\text{Σφάλμα τύπου I}) = P(\text{απορ. } H_0 \,/\, H_0 \text{ αληθής}) \\ \beta = P(\text{Σφάλμα τύπου II}) = P(\text{αποδ. } H_0 \,/\, H_a \text{ αληθής}) \end{cases}$$

Ταυτόχρονη ελαχιστοποίηση των a, β ? ΑΔΥΝΑΤΗ
(√ Το αναμενόμενο / ιδανικό)

NO.

Date

Θέλαμε να ξεκαθαρίσουμε αυτό το ανακάτεμα

⇓

Θα προκαθορίσω το a (δηλ. θα πω ότι θέλω να δουλέψω για a=5%, 1%, 10%)

a = επίπεδο σημαντικότητας του ελέγχου

Άρα δεν με απασχολεί η ελαχιστοποίηση του a γιατί το θεωρώ γνωστό

Στη συνέχεια, θα προσπαθήσω να βρω τη διαδικασία που ελαχιστοποιεί το β (σκοπός μας)

ή εναλλακτικά μεγιστοποιεί το 1-β.
Συμβολίζω γ=1-β και το ονομάζω ισχύς του τεστ.

και $1-\beta = 1 - P(\text{αποδ. } H_0 / H_a \text{ αληθής}) =$
$= P(\text{απορ. } H_0 / H_a \text{ αληθής}) = \gamma$
συμπλήρωμα

Επίσης, p-τιμή (Sig) είναι η μικρότερη τιμή του επιπέδου σημαντικότητας για την οποία απορρίπτεται η μηδενική υπόθεση.

Ο κανόνας λέει: αν p-τιμή ≥ a τότε η μηδενική υπόθεση H_0 δεν απορρίπτεται.

αν p-τιμή < a τότε η H_0 απορρίπτεται

NO

Date

Έχουμε 2 κατηγορίες υποθέσεων

απλές υποθέσεις — σύνθετες υποθέσεις

<u>Απλή υπόθεση</u>: είναι εκείνη η οποία μας δίνει πλήρη γνώση για το τι συμβαίνει όταν ισχύει

<u>π.χ.</u> $H_0 : \theta = 7$ (απλή) (<u>απλή προς απλή</u>)
$H_a : \theta = 9$ (απλή)

<u>π.χ</u> $H_0 : \theta = 7$ (απλή) (<u>απλή προς σύνθετη</u>)
$H_a : \theta > 7$ (σύνθετη)
μας δίνει πληροφορία μόνο για την φορά

Άλλος διαχωρισμός

αριστερόπλευρος έλεγχος — δεξιόπλευρος έλεγχος — δίπλευρος έλεγχος.

(One-Tailed)
(Two-Tailed)

Μια σχέση που θα πρέπει να ικανοποιείται είναι η:
P (απορ H_0 / H_0 αληθής) = α

Επίσης θα πρέπει να βρω την βέλτιστη διαδικασία

$\Downarrow$

Διαδικασία με την <u>μεγαλύτερη ισχύ</u>

Μεταξύ 2 τεστ του ίδιου επιπέδου σημαντικότητας προτιμότερο είναι εκείνο το οποίο έχει τη μεγαλύτερη ισχύ

Η πολυωνυμική μετράει πόσες φορές εμφανίζεται ένα ενδεχόμενο $X_1, X_2 \dots$

Άσκηση 1.17

$X_1, X_2, \dots, X_k$ τ.δ. από έναν πληθυσμό με πολυωνυμική:

$$p(X_1, X_2, \dots, X_k) = \frac{n!}{X_1!\,X_2! \dots X_k!} \cdot p_1^{X_1} p_2^{X_2} \dots p_k^{X_k} \quad (1)$$

όπου $\boxed{\sum_{i=1}^{k} p_i = 1}$ $\boxed{\sum_{i=1}^{k} X_i = n}$ $\quad X_k = n - \sum_{i=1}^{k-1} X_i$

$$(1) = \frac{n!\, p_1^{X_1} p_2^{X_2} p_3^{X_3} \dots p_{k-1}^{X_{k-1}} (1-p_1-p_2-\dots-p_{k-1})^{n-X_1-\dots-X_{k-1}}}{X_1!\,X_2! \dots X_{k-1}!\,(n-X_1-X_2-\dots-X_{k-1})!}$$

A

Παίρνω εκτίμηση μέγιστης πιθανοφάνειας:

$\log p(X_1, \dots, X_{k-1}) = \dots$

και $\frac{\partial}{\partial p_i} \log p = \log A + X_1 \log p_1 + X_2 \log p_2 + \dots$

Άρα βρίσκω $\boxed{\hat{p}_i = \frac{X_i}{n}}$