Μη Παραμετρική Στατιστική

Μαθημα 1ο

20/2/19

· Το κεφάλαιο 7 απο το βιβλίο της Εισαγωγής της στατιστικής.

Κλίμακες Μέτρησης:

1) Ονοματική ή Ονομαστική
2) Διατάξιμη
} αφορούν ποιοτικές μεταβλητές

3) Διαστηματική
4) Αναλογική (λόγου)
} ποσοτικές μεταβλητές

1) Δίνουν τιμές μονο σε κατηγορίες
2) Εαν υπάρχει δυνατοτητα σύγκρισης π.χ. λοχαγός, στρατηγός. (να υπάρχει δηλ. δυνατοτητα διαταξης)
3) Αν μπορεί να μας δώσει τιμή π.χ. θερμοκρασία τοτε μιλάμε για μεταβλητές διαστηματικές.
4) μεταβλητές που παιρνουν την τιμή 0 οπως βάρος, έσοδα

Πολυωνυμική Κατανομή

Γενίκευση Διωνυμικής
n ανεξάρτητες δοκιμές, καθε δοκιμή: $E_1, \ldots, E_k$ στοιχεία (απλά) ενδεχόμενα (αποτελέσματα)

πιθανότητα πραγματοποίησης των αποτελεσμάτων:

$p_i = P(E_i)$, $\sum_{i=1}^{k} p_i = 1$

Έστω X_i = ο αριθμός των εμφανίσεων του E_i, στις n δοκιμές με $\sum_{i=1} x_i = n$

$$(X_1, \dots, X_{k-1}) \sim P_{X_1,\dots,X_{k-1}}(x_1,\dots,x_k) = P(X_1 = x_1, \dots, X_{k-1} = x_{k-1}) =$$

$$= \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \ , \ \sum_{i=1}^{k} p_i = 1 ; \ \sum_{i=1}^{k} x_i = n .$$

ή

$$(X_1, \dots, X_{k-1}) \sim M(n, p_1, \dots, p_{k-1})$$

Προκύπτει ότι $X_1 \sim B(n, p_1)$

Εκτίμηση με την Μέθοδο της Μέγιστης Πιθανοφάνειας

Έστω $X_1, \dots, X_n$ τ.δ. από πληθυσμό με σ.π. $P(x, \theta)$ ή σ.π.π. $f(x, \theta)$. Η συνάρτηση $L(\theta | x) = \prod_{i=1}^{n} P(x_i, \theta) \stackrel{\eta}{=}$ ή $\prod_{i=1}^{n} f(x_i, \theta)$ ονομάζεται συνάρτηση πιθανοφάνειας

Ο εκτιμητής ή εκτιμήτρια $\hat{\theta} = \theta(x_1, \dots, x_n)$ του θ λέγεται Ε.Μ.Π αν $L(\hat{\theta} | x) = \max_{\theta \in \Theta} L(\theta | x)$

όπου Θ ο παραμετρικός χώρος του θ.

Παρατηρήσεις:

i) αντί της $L(\theta|x)$ χρησιμοποιείται η $\ln L(\theta|x)$

ii) —

Παράδειγμα 1

Έστω $X_1, \dots, X_n$ τ.δ. Εκθ$(1/\theta)$

ΕΜΠ του θ;

Λύση

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{\Sigma x_i}{\theta}}$$

$$\ln L(\theta|x) = -n\ln\theta - \frac{\Sigma X_i}{\theta}$$

$$\frac{\partial \ln L(\theta|x)}{\partial\theta} = \frac{-n}{\theta} + \frac{\Sigma X_i}{\theta^2}$$

$$\frac{\partial \ln L(\theta|x)}{\partial\theta} = 0 \Rightarrow \hat{\theta} = \frac{\Sigma X_i}{n} = \bar{X}.$$

$$\left.\frac{\partial^2 \ln L(\theta|x)}{\partial\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n^3}{(\Sigma X_i)^2} < 0$$

Παράδειγμα 2

$X_1, \dots, X_n$ τ.δ. $P(\theta)$. ΕΜΠ του θ;

Λύση

$$L(\theta|x) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\Sigma x_i}}{\Pi(x_i!)}$$

$$\ln L(\theta|x) = -n\theta + \Sigma X_i \ln\theta - \sum_{i=1}^{n} \ln x_i!$$

$$\frac{\partial \ln L(\theta|x)}{\partial\theta} = -n + \frac{\Sigma X_i}{\theta}$$

$$\frac{\partial \ln L(\theta|x)}{\partial\theta} = 0 \Rightarrow \hat{\theta} = \frac{\Sigma X_i}{n} = \bar{X}$$

$$\left.\frac{\partial^2 \ln L(\theta|x)}{\partial\theta^2}\right|_{\theta=\hat{\theta}} = -\frac{n^2}{\Sigma X_i} < 0$$

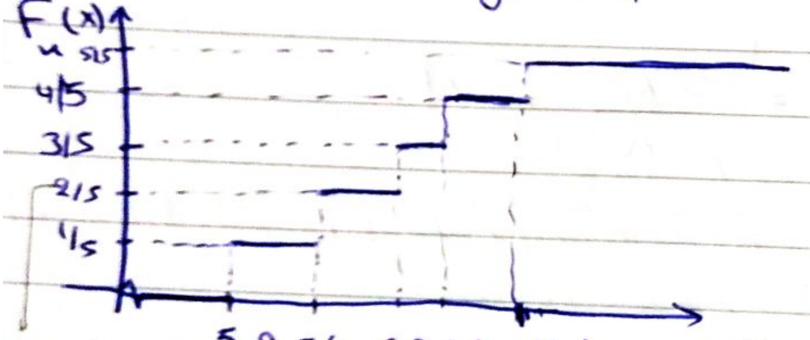
Εμπειρική Αθροιστική Συνάρτηση Κατανομής

Έστω $X_1, \ldots, X_n$ τ.δ. από πληθυσμό με α.σ.κ $F(x)$.
Η ΕΑΣΚ $F_n(x)$ ισούται με το μέρος (ποσοστό) των X_i και που είναι μικρότερα ή ίσα με το x, $-\infty < x < +\infty$. Δηλ. $F_n(x) = \frac{\text{αριθμός } X_i \leq x}{n}$, $x \in \mathbb{R}$

$$= \frac{\sum_{i=1}^{n} I_{(-\infty, x)}(x_i)}{n}, \ x \in \mathbb{R} \quad \text{όπου } I_c(x_i) = \begin{cases} 1, & x_i \in C \\ 0, & x_i \notin C \end{cases}$$

παράδειγμα

6.2, 5.6, 7.1, 6.4, 5.2. Οι τιμές και η εασκ φαίνονται στο γράφημα.



2 τιμές μικρότερες του 5.6 άρα $\frac{2}{5}$

ΕΑΣΚ = βηματική ΑΣΚ.

$P(X = 6.2) = \frac{1}{5}$

Θεώρημα 1

Έστω $X_1, \dots, X_n$ τ.δ. από πληθυσμό με ασκ $F(x)$. Για σταθερό x, η τ.μ. $n \cdot F_n(x) \sim B(n, F(x))$ όπου $F_n(x)$ η εασκ.

απόδειξη (ίσως ερώτ.)

Οι δυνατές τιμές της εασκ ανήκουν στο σύνολο $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$. Για σταθερό x, η τιμή της εασκ είναι ίση με $\frac{k}{n}$, $k = 0, 1, \dots, n$, αν-ν

k το πλήθος από δειγματικές τιμές είναι $\leq x$. Άρα, ο υπολογισμός $P\left(F_n(x) = \frac{k}{n}\right)$ είναι η πιθανότητα να έχουμε k επιτυχίες σε n δοκιμές ενός διωνυμικού πειράματος με πιθανότητα $p = P(X_i \leq x) = F(x)$.

Άρα $P\left(F_n(x) = \frac{k}{n}\right) = \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}, k = 0, \dots, n$

$= P(nF_n(x) = k)$ άρα $n \cdot F_n(x) \sim B(n, F(x))$

Θεώρημα 2

(α) Η εασκ $F_n(x)$ είναι συνεπής εκτιμητής της $F(x)$

(β) Η $F_n(x)$ ακολουθεί ασυμπτωτική κανονική κατανομή με μέση τιμή $F(x)$ και διακύμανση $F(x) \cdot [1 - F(x)] / n$.

Ο συνεπής εκτιμητής είναι αμερόληπτος και έχει διακύμανση $\xrightarrow[n]{} 0$.

αποδειξη:

$E(nF_n(x)) = n \cdot F(x) \rightsquigarrow E(F_n(x)) = F(x)$ άρα αμερολ.

$Var(n F_n(x)) = n F(x)[1 - F(x)] \rightsquigarrow n^2 Var F_n(x) =$

$= n F(x)[1 - F(x)] \to Var F_n(x) = \frac{F(x)[1 - F(x)]}{n} \xrightarrow[n \to \infty]{} 0$.

Άρα $X \sim B(n, p) \rightsquigarrow X \overset{προσ.}{\sim} N(np, np(1-p))$