Διάλεξη 2η
01/03/2019

ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ.

Φύση

Πληθυσμός:

↓

χαρακτηριστικό γνώρισμα ή ιδιότητα μελών πληθ.

↓

Τυχαία μεταβλητή (τ.μ.)

↓

Κατανομή της τ.μ.
(→π.χ. Εκθ(θ) ή $N(\mu,\sigma^2)$)

(*) Εδώ εμφανίζεται ένα πρόβλημα: πάντα στην πράξη οι παράμετροι της κατανομής είναι άγνωστες.
Στο πρόβλημα απαντά η στατιστική συμπερ.

Ορισμός: Αντικείμενο της στατιστικής συμπερασματολογίας είναι να εφοδιάζει με μεθοδολογίες για την προσέγγιση (εκτίμηση) των αγνώστων παραμέτρων της κατανομής.

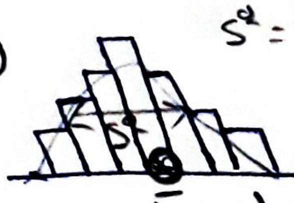
- (I) Εκτίμηση σε σημείο
- (II) Εκτίμηση σε διάστημα
- (III) Έλεγχοι στατιστικών υποθέσεων - Στατιστικά τεστ.

Στατιστικός: Μελέτη μέρους του πληθυσμού (δείγμα)
Το μέρος που θα διαλέξω θα πρέπει να είναι όσο πιο αντιπροσωπευτικό γίνεται. Το κομμάτι αυτό του πληθυσμού είναι το <u>τυχαίο δείγμα</u> (είναι <=> με τον πληθυσμό)

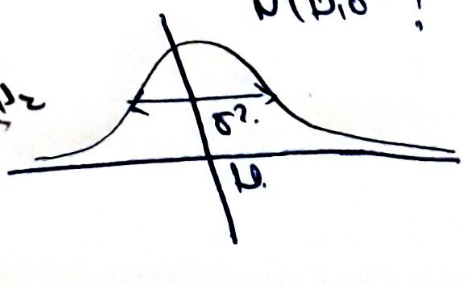
Η τ.μ. $X \Leftrightarrow x_1, x_2, \dots, x_n$.
Το ιστόγραμμα (*) μου δείχνει πώς κατανέμονται οι τυχαίοι αριθμοί n $N(\mu,\sigma^2)$!

(*)

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$$



το ένα προσεγγίζει το άλλο



$$\bar{x} = \frac{1}{n}\sum x_i$$

(I) Εκτίμηση σε σημείο ή σημειοεκτιμητική:

Βασικό στοιχείο: Η ύπαρξη ενός τυχαίου δείγματος (τ.δ.)

Ορισμός: Το τ.δ. μεγέθους n είναι η ανεξάρτητες και ισόνομες παρατηρήσεις, οι $X_1, X_2, \dots, X_n$ της τ.μ. X την οποία μελετάμε.

Διευκρινήσεις:

• Ανεξάρτητα:
- Το X_i δεν επηρεάζει οποιοδήποτε άλλο από τα $X_1, \dots, X_n$.
- Από μαθηματικής πλευράς $f_{X_1, X_2, \dots, X_n} = \prod_{i=1}^{n} f_{X_i}$

• Ισόνομα: Οι $X_1, X_2, \dots, X_n$ ισόνομες, σημαίνει ότι όλες έχουν την ίδια κατανομή $f(x,\theta)$, θ=παράμετρος.

Τα $X_1, \dots, X_n$ έχουν διπλό ρόλο.

(i) είναι ανεξάρτητες και ισόνομες τ.μ.

(ii) έχουν μια αριθμητική τιμή που αντιστοιχεί στην X στο i-μέλος του δείγματος.
παρατήρηση της τ.μ.

o ——— o ——— o ——— o

▶ Έστω τ.δ. $X_1, \dots, X_n$ από πληθυσμό που περιγράφεται από μια κατανομή $f(x,\theta)$ ← εξαρτάται από μια παράμετρο θ.

Ορισμός: Παραμετρικός χώρος (H) είναι το σύνολο στο οποίο παίρνει τιμές η άγνωστη παράμετρος θ. Το $H \subseteq \mathbb{R}$ ή $H \subseteq \mathbb{R}^k$.

π.χ.

α) Έστω τ.δ. $X_1, X_2, \dots, X_n$ από $f(x,\theta)$ να είναι εκθετική με παραμ. θ.

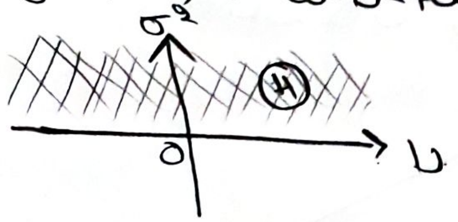
$f(x,\theta) = \theta e^{-\theta x}$, $x>0, \theta>0$

Άρα $\Theta = (0,\infty)$

β) Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό με κατανομή $N(\mu,\sigma^2)$

i) Έστω μ, σ^2 άγνωστα.

Τότε $\Theta = \{(\mu,\sigma^2) : -\infty<\mu<+\infty \, , \sigma^2>0\} \subseteq \mathbb{R}^2$.

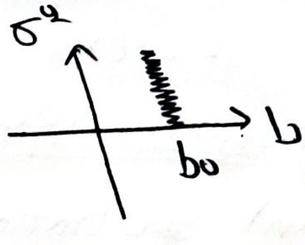


ii) Έστω $\sigma^2 = \sigma_0^2$ γνωστή

$\Theta = \{\mu \in \mathbb{R}\} = \mathbb{R}$

iii) Έστω $\mu = \mu_0$ γνωστό.

$\Theta = \{\sigma^2 : \sigma^2>0\} = (0,\infty)$



Στατιστική Συνάρτηση: Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό με κατανομή $f(x,\theta)$, $\theta\in\Theta$. Στατιστική συνάρτηση ονομάζεται οποιαδήποτε συνάρτηση του τ.δ. $X_1,\dots,X_n$.

π.χ.

$T = T(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$.

$T = T(X_1,\dots,X_n) = \frac{1}{n-1}\sum_{i=1}^{n} (X_i-\bar{X})^2 = S^2$.

Ορισμός: Εκτιμητής είναι η στατιστική συνάρτηση την οποία χρησιμοποιούμε για την εκτίμηση (προσέγγιση) της άγνωστης παραμέτρου $\theta \in \Theta$.

Κριτήρια επιλογής εκτιμητών - Μέσο τετραγωνικό σφάλμα:

Έστω τ.δ. $X_1, \dots, X_n$ από $f(x;\theta)$, $\theta \in \Theta$ και έστω ότι ενδιαφερόμαστε για την κατασκευή ενός εκτιμητή της $g(\theta)$, $\theta \in \Theta$.
Ένα κριτήριο επιλογής εκτιμητή θα είναι: $|T(X_1, \dots, X_n) - g(\theta)|$

Όμως το κριτήριο αυτό έχει αδυναμίες:

1η αδυναμία: Απόλυτη τιμή.

2η αδυναμία: Εξαρτάται από το θ, το οποίο επίσης μου είναι άγνωστο.

Για να αντιμετωπίσω τις αδυναμίες αυτές:
Θεωρώ ως κριτήριο: $\boxed{E(T(X_1, \dots, X_n) - g(\theta))^2}$ το οποίο ονομάζεται μέσο τετραγωνικό σφάλμα (ΜΤΣ)

Ορισμός: Αν $T(X_1, \dots, X_n)$ εκτιμητής της $g(\theta)$. Το μέσο τετραγωνικό σφάλμα του εκτιμητή T για την εκτίμηση της $g(\theta)$ συμβολίζεται με $ΜΤΣ(T, g(\theta))$ και ορίζεται να είναι:

$$\boxed{ΜΤΣ(T, g(\theta)) = E[T(X_1, \dots, X_n) - g(\theta)]^2}$$

Πρόταση: $ΜΤΣ(T, g(\theta)) = Var(T) + (E(T) - g(\theta))^2$

Αποδ.

$Var(\omega) = E(\omega^2) - (E(\omega))^2$. Άρα $E(\omega^2) = Var(\omega) + (E\omega)^2$

$ΜΤΣ(T, g(\theta)) \stackrel{ορ}{=} E(T - g(\theta))^2 = Var(T - g(\theta)) + [E(T - g(\theta))]^2$

$= Var(T) + [E(T) - E(g(\theta))]^2 = Var(T) + (E(T) - g(\theta))^2$

Π.χ. Έστω τ.δ. $X_1, \dots, X_n$ από $Εκθ(\frac{1}{\theta})$, $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}$, $x > 0$, $\theta > 0$. Έστω ο εκτιμητής $T = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$ της θ. Να βρεθεί το ΜΤΣ του $\bar{X}$ για την εκτίμηση της θ.

Λύση

$\boxed{ΜΤΣ(\bar{X}, \theta) = Var(\bar{X}) + (E(\bar{X}) - \theta)^2}$ $E(\bar{X}) = E(\frac{1}{n} \sum X_i) = \frac{1}{n} \sum E(X_i)$

$\xrightarrow[\text{Άρα } X_i \sim Εκθ(\frac{1}{\theta}) \Rightarrow E(X_i) = \frac{1}{1/\theta} = \theta.]{X_i \text{ - ισόνομες}} \frac{1}{n} \sum \theta = \frac{1}{n} \cdot n\theta = \theta.$

$Var(\bar{X}) = Var(\frac{1}{n} \sum X_i) = \frac{1}{n^2} Var(\sum X_i) \stackrel{(*)}{=} \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i)$

$= \frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{(1/\theta)^2} = \frac{1}{n^2} \sum \theta^2$

$= \frac{1}{n^2} \cdot n\theta^2 = \frac{\theta^2}{n}$

(*) αν οι $\omega_1, \dots, \omega_n$ ανεξάρτ.
τότε $Var(\sum_{i=1}^{n} a_i \omega_i) = \sum_{i=1}^{n} a_i^2 Var(\omega_i)$

Άρα $ΜΤΣ = \theta^2/n - 0 = \frac{\theta^2}{n}$

και αν $n \to \infty$ τότε $\underline{\underline{ΜΤΣ = 0}}$ (επαρκής εκτ.)

Ορισμός: Έστω 2 εκτιμητές T_1 και T_2 για την εκτίμηση της $g(\theta)$, $\theta \in \Theta$. Ο εκτιμητής T_1 λέγεται καλύτερος από τον T_2 με κριτήριο το ΜΤΣ αν $ΜΤΣ(T_1, g(\theta)) \leq ΜΤΣ(T_2, g(\theta))$, $\forall \theta \in \Theta$.

Παρατήρηση: ⓐ Ο T_2 λέγεται μη αποδεκτός
ⓑ Ένας εκτιμητής λέγεται αποδεκτός αν έχει το μικρότερο ΜΤΣ.

Αμεροληψία εκτιμητή:

$$ΜΤΣ(T, g(\theta)) = \underbrace{Var(T)}_{\geq 0} + \underbrace{(E(T) - g(\theta))^2}_{\geq 0}$$

Ορισμός: (αμερόληπτος εκτιμητής)
Ο εκτιμητής $T = T(X_1, \dots, X_n)$ λέγεται αμερόληπτος εκτιμητής της $g(\theta)$ αν η $E(T) = g(\theta)$ αν η $E(T) = g(\theta)$

Ορισμός: Αν T ένας εκτιμητής του $g(\theta)$ η μεροληψία του T ορίζεται να είναι: $b(T, g(\theta)) = E(T) - g(\theta)$

Παρατήρηση: ⓐ Αν T αμερόληπτος της $g(\theta)$, τότε $b(T, g(\theta)) = 0$.
ⓑ Αν ο εκτιμητής T είναι αμερόληπτος της $g(\theta)$ τότε $ΜΤΣ(T, g(\theta)) = Var(T)$

Διαισθητικά: Αν περιοριστούμε στην κλάση των αμερόληπτων εκτιμητών της $g(\theta)$, τότε η $Var(T)$ είναι το κριτήριο ποιότητας του εκτιμητή.

Ορισμός: Ονομάζουμε αμερόληπτο ομοιόμορφα ελάχιστης διακύμανσης (ΑΟΕΔ) εκτιμητή, έναν εκτιμητή $T = T(X_1, \dots, X_n)$, αν ο T είναι αμερόληπτος της $g(\theta)$, αν: (i) ο T είναι αμερόληπτος της $g(\theta)$ και (ii) έχει την μικρότερη διακύμανση στην κλάση όλων των αμερόληπτων εκτιμητών της $g(\theta)$.

Πρόταση: Αν ο ΑΟΕΔ εκτιμητής της $g(\theta)$ υπάρχει, είναι ΜΟΝΑΔΙΚΟΣ.

Παραδείγματα:

(1.) Έστω τ.δ. $X_1, \dots, X_n$ από κατανομή $f(x,\theta)$, $\theta \in \Theta$ και μ η μέση τιμή της κατανομής αυτής. Ν.δ.ο. ο $\bar{X} = \frac{1}{n}\sum X_i$
- i) αμερόληπτος της μ
- ii) $Var(\bar{X}) = \frac{\sigma^2}{n}$, $\sigma^2 = Var(X)$

Λύση

(i) Αρκεί ν.δ.ο. $E(\bar{X}) = \mu$.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i) \underset{\text{ισόνομες}}{\overset{X_i}{=}} \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n \cdot \mu = \mu.$$

(ii) $$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}Var\left(\sum X_i\right) \underset{\text{ανεξ.}}{\overset{X_i}{=}} \frac{1}{n^2}\sum Var(X_i)$$

$$\underset{\text{ισον.}}{\overset{X_i}{=}} \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

(2) Έστω τ.δ. $X_1, \dots, X_n$ από κατανομή $f(x,\theta)$, $\theta \in \Theta$ με διακύμανση σ^2. Ν.δ.ο. η δειγματική διακύμανση $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ είναι αμερόληπτος εκτιμητής της σ^2.

Λύση

Αρκει νδο $E(S^2)=\sigma^2$.

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2\bar{x}\cdot x_i + \bar{x}^2)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n\bar{x}^2\right)$$

$$\xrightarrow[\sum x_i = n\bar{x}]{\bar{x}=\frac{1}{n}\sum x_i} \frac{1}{n-1}\left(\sum x_i^2 - 2\bar{x}\, n\bar{x} + n\bar{x}^2\right)$$

$$\Rightarrow S^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

Αρα $E(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}E(x_i^2) - nE(\bar{x}^2)\right)$

$$\Rightarrow E(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(Var(x_i)+E(x_i)^2\right) - n\left(Var(\bar{x}) + (E\bar{x})^2\right)\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$$

$$= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right)$$

$$= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2.$$