Παρασκευή 27/03/20

Μάθημα 2ο

Εκτίμηση σε Σημείο:

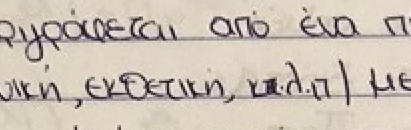
Περιγραφή Προβλήματος - Συμβολισμοί - Ορολογία

(i) Χαρακτηριστικό γνώρισμα μελών πληθυσμού: τ.μ. X ή τ.δ. $\underset{\sim}{X} = (X_1, \dots, X_p)'$. Περιοριζόμαστε στην τ.μ. X

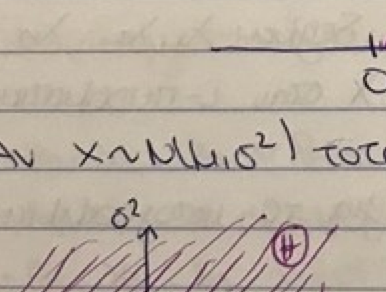
(ii) Το X περιγράφεται από ένα πιθανοθεωρητικό μοντέλο (π.χ. κανονική, εκθετική, κ.λ.π) με κατανομή $f(x,\theta)$ που εξαρτάται από μια παράμετρο θ.

Η θ είναι μονοδιάστατη ή πολυδιάστατη είναι άγνωστη και ανήκει στον παραμετρικό χώρο Θ

π.χ) Αν $X \sim Εκθ(\theta)$ τότε $\Theta = \{\theta \in \mathbb{R};\ \theta > 0\}$.



Αν $X \sim N(\mu, \sigma^2)$ τότε $\Theta = \{(\mu, \sigma^2) \in \mathbb{R}^2 : \mu \in \mathbb{R},\ \sigma^2 > 0\}$



(iii) Η εκτίμηση της $\theta \in \Theta$ βασίζεται σε παρατηρήσεις $X_1, X_2, \dots, X_n$ που συλλέγονται μετρώντας το χαρακτηριστικό X σε n- τυχαία επιλεγμένες πειραματικές μονάδες του πληθυσμού.

Αν X_i παριστά το χαρακτηριστικό X για ~~[illegible]~~ $i = 1, \dots, n$. τότε $X_1, X_2, \dots, X_n$ είναι n- τυχαίες μεταβλητές και $x_1, x_2, \ldots, x_n$ είναι οι παρατηρηθείσες τιμές των τ.μ.

Οι τ.μ. $X_1, \ldots, X_n$ συνιστούν ένα <u>τυχαίο δείγμα</u> (τ.δ.) μεγέθους n αν είναι ανεξάρτητες και ισόνομες, δηλαδή είναι ανεξάρτητες και προέρχονται από την ίδια κατανομή $f(x, \theta)$.

Το $\boxed{\text{πλήθος}}$ των δεδομένων $\boxed{n}$ λέγεται $\boxed{\text{μέγεθος του δείγματος}}$

(iv) Ένας σημειοεκτιμητής της $\theta \in \Theta$ βασίζεται στο τ.δ. $X_1, \ldots, X_n$ και είναι μια συνάρτηση του $X_1, \ldots, X_n$.

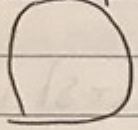
<u>Στατιστική Συνάρτηση (σ.δ.):</u>
Κάθε συνάρτηση $T = T(X_1, X_2, \ldots, X_n)$ του τ.δ. $X_1, X_2, \ldots, X_n$.

<u>Εκτιμητής της $\theta \in \Theta$ ή της $g(\theta)$:</u>
Κάθε στατιστική συνάρτηση που προτείνεται για την εκτίμηση σε σημείο της $\theta \in \Theta$ ή της $g(\theta)$, $\theta \in \Theta$

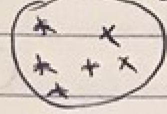
<u>Εκτίμηση της $\theta \in \Theta$ ή της $g(\theta)$:</u>
Η τιμή $T(x_1, x_2, \ldots, x_n)$ του εκτιμητή $T(X_1, \ldots, X_n)$ για την παρατηρηθείσα τιμή $x_1, \ldots, x_n$ του τ.δ. $X_1, \ldots, X_n$.

Σχηματικά Εκτίμηση σε σημείο:

Πληθυσμός. $\xrightarrow{\text{τ.δ.}}$ Τυχαίο δείγμα



Τυχαίες επιλεγμένες παρατηρήσεις
Τυχαία επιλεγμένα μέλη του πληθυσμού



Τα μέλη του έχουν ένα κοινό χαρακτηριστικό, μια τ.μ. X

↓

Η τ.μ. X ακολουθεί μια κατανομή με $f(x,\theta)$, που εξαρτάται από μια άγνωστη παράμετρο $\theta \in \Theta \subseteq \mathbb{R}^m$

↓

Η εκτίμηση της θ αποτελεί το ουσιαστικό βήμα για την εξειδίκευση του μοντέλου $f(x,\theta)$ και τον υπολογισμό των χαρακτηριστικών του.
Οι μεθοδολογίες για την εκτίμηση της θ και τον έλεγχο υποθέσεων, δίνουν περιεχόμενο στο αντικείμενο της Στατιστικής Συμπερασματολογίας

Το X παρατηρείται και καταγράφεται από n μέλη του πληθ. X_i παριστά το χαρακτ. X στο i-μέλος του δείγματος, x_i η παρατηρούμενη τιμή της X_i, $i=1,\dots n$.

↓

Το τ.δ. αποτελεί το μοναδικό όπλο του στατιστικού για την εκτίμηση της θ ή της $f(x,\theta)$ αν αυτή δεν είναι γνωστή.

↓

Επειδή το τ.δ. $X_1 \dots X_n$ είναι n-διάστατο προχωράμε σε συμπύκνωσή του μέσω μιας κατάλληλης συνάρτησής του: $T=T(X_1,\dots,X_n)$

Στατιστικό: κάθε συνάρτηση του τ.δ.

Εκτιμητής: μια στατιστική συνάρτηση που χρησιμοποιείται για την εκτίμηση της θ ή της $g(\theta)$.

Η τιμή του εκτιμητή T στον $x_1,\dots,x_n$ του τ.δ. λέγεται εκτίμηση της θ ή της $g(\theta)$

Ποιότητα - Κριτήρια επιλογής εκτιμητών

Έστω $T(\underset{\sim}{X}) = T(X_1, \dots, X_n)$ εκτιμητής της $g(\theta)$

Η ποιότητα του εξαρτάται από το πόσο κοντά είναι στην $g(\theta)$ δηλαδή από την ποσότητα $|T(\underset{\sim}{X}) - g(\theta)|$.

Η $|T(\underset{\sim}{X}) - g(\theta)|$ είναι διάσπαση και δεν μπορεί να χρησιμοποιηθεί γιατί είναι η ίδια τ.μ. Για αυτό μπορεί να χρησιμοποιηθεί το Μέσο Απόλυτο Σφάλμα: $E_\theta(|T(\underset{\sim}{X}) - g(\theta)|)$ ή το

$\boxed{\text{Μέσο Τετραγωνικό Σφάλμα: } E_\theta([T(\underset{\sim}{X}) - g(\theta)]^2).}$

Συνάρτηση Ζημιας:

$L(d, \theta)$: εκφράζει τη ζημιά αν εκτιμήσουμε την $g(\theta)$ με την d.

Πρέπει να ικανοποιεί:

i) $L(d, \theta) \geq 0 \quad \forall \theta, d$.

ii) $L(g(\theta), \theta) = 0 \quad \forall \theta \in \Theta$

Περιοριζόμαστε στο μέσο τετραγωνικό σφάλμα (ΜΤΣ) του εκτιμητή $T(\underset{\sim}{X})$ της $g(\theta)$ που ορίζεται

$$ΜΤΣ(T, \theta) = E_\theta(T(\underset{\sim}{X}) - g(\theta))^2.$$

Πρόταση:

Ισχύει $\boxed{ΜΤΣ(T, g(\theta)) = Var(T) + [E(T) - g(\theta)]^2}, \theta \in \Theta$

Απόδειξη.

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

όμως $g(\theta) \in \mathbb{R}$ άρα.

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

Παράδειγμα:

Έστω τ.δ. $X_1 \ldots X_n$ με κατανομή $Εκθ(1/\theta)$ και σ.π.π.

$f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x>0$, $\theta>0$. Έστω η ε.ο.

$T = T(\underset{\sim}{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ για την εκτίμηση της θ. Να υπολογισθεί το $ΜΤΣ(\bar{X},\theta)$.

Από $ΜΤΣ(\bar{X},\theta) = Var(\bar{X}) + [E(\bar{X})-\theta]^2$.

Υπολογίζω της $E(\bar{X})$, $Var(\bar{X})$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1/\theta} = \frac{1}{n}\sum_{i=1}^{n}\theta = \theta.$$

Άρα $E(\bar{X}) = \theta$

Αν ~~Ε~~ οι τ.μ. $W_1 \ldots W_n$ είναι ανεξάρτητες τότε:

$$Var\left(\sum_{i=1}^{n} a_i W_i\right) = \sum_{i=1}^{n} a_i^2 Var(W_i)$$ με a_i σταθερές

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\left(\frac{1}{\theta}\right)^2} =$$

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

Άρα $ΜΤΣ(\bar{X},\theta) = \frac{\theta^2}{n} + [\theta-\theta]^2 = \frac{\theta^2}{n}.$

Παρατήρηση

Αν $n \to \infty$ τότε $ΜΤΣ(\bar{X},\theta) \to 0$

Το ΜΤΣ εισήχθη ως ένα ποσοτικό κριτήριο για να εκφράσει την "εγγύτητα" του εκτιμητή $T = T(X_1,\ldots,X_n)$ με την παράμετρο θ ή $g(\theta)$ την οποία σχεδιάζεται για να προσεγγίσει.

Το ΜΤΣ χρησιμοποιείται για τη σύγκριση δύο εκτιμητών της ίδιας παραμέτρου.

Αποδεκτός Εκτιμητής:

Για δυο εκτιμητές $T_1 = T_1(\underset{\sim}{X})$ και $T_2 = T_2(\underset{\sim}{X})$ της θ, ο εκτιμητής T_1 ορίζεται να είναι καλύτερος του T_2 με κριτήριο το ΜΤΣ αν και μόνο αν

i) $MTS(T_1, \theta) \leq MTS(T_2, \theta)$, $\forall \theta \in \Theta$ και

ii) $MTS(T_1, \theta) < MTS(T_2, \theta)$, για τουλάχιστον ένα $\theta \in \Theta$

Ο T_2 ονομάζεται μη αποδεκτός εκτιμητής.
Ενας εκτιμητής είναι αποδεκτός αν-ν δεν υπάρχει καλύτερος από αυτόν.
Προτιμότερος εκτιμητής ειναι αυτός με το μικρότερο ΜΤΣ.

Ορισμός αμερόληπτου εκτιμητή:

Ένας εκτιμητής $T = T(\underset{\sim}{X})$ λέγεται αμερόληπτος εκτιμητής της $g(\theta)$, $\theta \in \Theta$ αν-ν $E[T(\underset{\sim}{X})] = g(\theta)$, για όλα τα $\theta \in \Theta$

Ενας εκτιμητής που δεν είναι αμερόληπτος, λέγεται μεροληπτικός

Ορισμός μεροληψίας:

Αν $T = T(\underset{\sim}{X})$ ένας εκτιμητής της $g(\theta)$, η μεροληψία του εκτιμητή T για την εκτίμηση της $g(\theta)$ συμβολίζεται με $b(T, \theta)$ και ορίζεται

$$\boxed{b(T, \theta) = E[T(\underset{\sim}{X})] - g(\theta), \ \theta \in \Theta}$$

Παρατήρηση:

Δυο εκτιμητές T_1 και T_2 που είναι αμερόληπτοι εκτιμητές της $g(\theta)$ τοτε αυτοί είναι "συγκεντρωμένοι" γύρω απο την ίδια τιμή $g(\theta)$.

Αν περιοριστούμε σε αμερόληπτους εκτιμητές της $g(\theta)$, αναιρείται η επίδραση της άγνωστης θ κατα την ~~προσέγγιση~~ σύγκριση τους.

Πρόταση:

Εστω $T = T(\underline{X})$ αμερόληπτος εκτιμητής της $g(\theta)$. Τότε το ΜΤΣ $(T, g(\theta))$ γίνεται:

$$\boxed{ΜΤΣ(T, g(\theta)) = Var(T(\underline{X})), \theta \in \Theta}$$

Απόδειξη:

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

αν T αμερόληπτος της $g(\theta) \Rightarrow E(T) = g(\theta)$.

Αρα $ΜΤΣ(T, g(\theta)) = Var\ T$. ∎

Ο δειγματικός μέσος ($\bar{X}$) και η δειγματική διακύμανση (S^2) εκτιμούν αμερόληπτα τα αντίστοιχα πληθυσμιακά μεγέθη μ και σ^2

Παράδειγμα:

Εστω τ.δ $X_1, \ldots, X_n$ από πληθυσμό με κατανομή με μέση τιμή μ και διακύμανση σ^2. ($E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $i = 1, \ldots n$). Με $\mu \in \mathbb{R}$ και $\sigma^2 > 0$ Εστω $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ο δειγματικός μέσος και $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ η δειγματική διακύμανση. Τότε

α) ο $\bar{X}$ είναι αμερόληπτος εκτιμητής του μ, $E(\bar{X}) = \mu$.

β) $Var(\bar{X}) = \sigma^2/n$.

γ) Ο S^2 είναι αμερόληπτος εκτιμητής της σ^2, $E(S^2) = \sigma^2$

Λύση

α)

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu.$$

β) $$Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

$\underline{\gamma)}$ $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right)$

$$E(S^2) = \frac{1}{n-1}\left\{\sum_{i=1}^{n}E(X_i^2) - nE(\bar{X})^2\right\} = \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left[Var X_i + [EX_i]^2\right] - n\left[Var(\bar{X}) + (E\bar{X})^2\right]\right\}$$

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

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

Αν περιοριστούμε στην κλάση των αμερόληπτων εκτιμητών της $g(\theta)$, ο καλύτερος από αυτούς, με την έννοια του ΜΤΣ, είναι εκείνος που ελαχιστοποιεί την διακύμανση

Ορισμός ΑΟΕΔ εκτιμητή: (Αμερόληπτου Ομοιόμορφα Ελάχιστης Διακύμανσης)

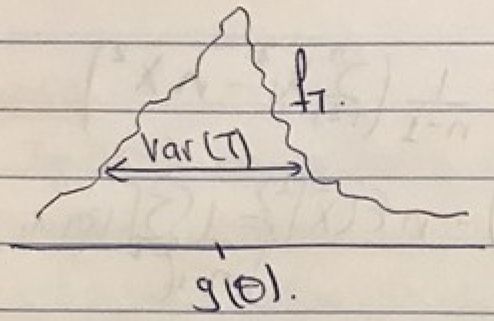
Έστω τ.δ $X_1, \dots, X_n$ από πληθυσμό με κατανομή $f(x;\theta)$, $\theta \in \Theta$ και έστω $T = T(\underline{X}) = T(X_1, \dots, X_n)$ ένας εκτιμητής της $g(\theta)$. Ο T λέγεται ΑΟΕΔ εκτιμητής της $g(\theta)$ αν είναι αμερόληπτος και έχει τη μικρότερη διακύμανση μεταξύ όλων των άλλων αμερόληπτων εκτιμητών της $g(\theta)$, $\theta \in \Theta$.

Αν T_1, ένας άλλος αμερόληπτος εκτιμητής της $g(\theta)$ τότε

$$Var(T) \leq Var(T_1) \quad \forall \theta \in \Theta$$

ΑΟΕΔ εκτιμητής:
(i) είναι αμερόληπτος
(ii) έχει την μικρότερη διακύμανση μεταξύ άλλων

Η ιδιότητα της αμεροληψίας (i) αναγκάζει τον ΑΟΕΔ εκτιμητή T να συγκεντρώνεται γύρω από τη $g(\theta)$ άρα η σ.π. του εκτιμητή είναι συγκεντρωμένη γύρω από την $g(\theta)$. Η ιδιότητα (ii) τον αναγκάζει να είναι πολύ συγκεντρωμένη γύρω από την $g(\theta)$



Οι τιμές του T με την μεγαλύτερη πιθανότητα βρίσκονται κοντά στην $g(\theta)$. Οι τιμές όμως της T είναι εκτιμήσεις της $g(\theta)$ άρα οι εκτιμήσεις βρίσκονται πολύ κοντά στην ποσότητα που εκτιμούν

Θεώρημα Μοναδικότητας ΑΟΕΔ εκτιμητή:

Αν υπάρχει ένας ΑΟΕΔ εκτιμητής $T = T(X_1, \dots X_n)$ της $g(\theta)$, $\theta \in H$ με $Var_\theta(T) < \infty$ τότε ο T είναι μοναδικός

[Η απόδειξη υπάρχει στις σημειώσεις του αλλά δεν θα ζητηθεί!]