Κεφάλαιο 2: Εκτιμητική

Γ. Ψαρράκος

Πανεπιστήμιο Πειραιώς

Τμήμα Οικονομικής Επιστήμης

Β΄ Εξάμηνο

28 Φεβρουαρίου 2013

Σκοπός: Εκτίμηση της παραμέτρου ή των παραμέτρων μιας κατανομής που εκφράζει κάποιο χαρακτηριστικό (πχ ύψος, βάρος κτλ) ενός πλυθυσμού.

Για παράδειγμα, έστω ότι κάποιο χαρακτηριστικό X σε ένα πληθυσμό ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$ με παραμέτρους τη μέση τιμή μ (άγνωστο) και διασπορά σ^2 (άγνωστο). Θέλουμε μέσω ενός τυχαίου δείγματος να εκτιμήσουμε τις δύο αυτές παραμέτρους, συμβ. $\hat{\mu}$ και $\hat{\sigma^2}$,

$$\hat{\mu} = \text{συναρτήσει του τ.δ.}, \quad \hat{\sigma^2} = \text{συναρτήσει του τ.δ.}$$

Ορισμός

Τυχαίο δείγμα μεγέθους n καλείται ένα σύνολο ανεξάρτητων και ισόνομων τ.μ. $X_1, X_2, \ldots, X_n$.

Ορισμός

Στατιστική ή δειγματική συνάρτηση λέγεται κάθε συνάρτηση $T(X_1, X_2, \ldots, X_n)$ που εξαρτάται μόνο από το τ.δ. $X_1, X_2, \ldots, X_n$ (δεν εξαρτάται από τις παραμέτρους που θέλουμε να εκτιμήσουμε)

Γνωστές στατιστικές συναρτήσεις είναι

- Δειγματικός μέσος

$$(\hat{\mu} =)\ \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- Δειγματικές ροπές τάξεως r

$$m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r = \frac{X_1^r + X_2^r + \ldots + X_n^r}{n}$$

- Δειγματική διασπορά

$$(\hat{\sigma^2} =)\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Εφ' όσον τα στατιστικά $\overline{X}$, S^2 κτλ είναι τυχαίες μεταβλητές μπορούμε να υπολογίσουμε στατιστικά μέτρα (όπως μέσες τιμές, ροπές ανώτερης τάξεως, διασπορές)

Θεώρημα

Έστω $X_1, X_2, \ldots, X_n$ ένα τυχαίο δείγμα από ένα πληθυσμό με μέση τιμή μ και διασπορά σ^2. Τότε

$$E(\overline{X}) = \mu,$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

και

$$E(S^2) = \sigma^2.$$

Βασικά κριτήρια αξιολόγησης εκτιμητών.

Δύο βασικά κριτήρια αξιολόγησης εκτιμητών είναι

- Αμερόληπτος εκτιμητής.
- Εκτιμητής ελαχίστης διασποράς.

1) Αμεροληψία: Ένας εκτιμητής $\hat{\theta}$ μιας παραμέτρου θ λέγεται αμερόληπτος αν

$$E(\hat{\theta}) = \theta, \quad \text{όπου} \quad \hat{\theta} = T(X_1, X_2, \ldots, X_n).$$

Γενικά η διαφορά $b(\theta) = E(\hat{\theta}) - \theta$ καλείται μεροληψία. Προφανώς η μεροληψία ενός αμερόληπτου εκτιμητή είναι 0. Όπως είδαμε ο δειγματικός μέσος και η δειγματική διασπορά είναι δύο αμερόληπτοι εκτιμητές δηλ. αν

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

και

$$\hat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

τότε

$$E(\hat{\mu}) = E(\overline{X}) = \mu \quad \text{και} \quad E(\hat{\sigma^2}) = E(S^2) = \sigma^2.$$

Ελάχιστη Διασπορά: Μεταξύ των αμερόληπτων εκτιμητών επιλέγουμε εκείνον που έχει τη μικρότερη διασπορά. Επομένως, αν βρούμε δύο αμερόληπτους εκτιμητές για την παράμετρο θ

$$T_1 = T_1(X_1, X_2, \ldots, X_n) \text{ και } T_2 = T_2(X_1, X_2, \ldots, X_n)$$

με

$$Var(T_1) > Var(T_2),$$

τότε ο εκτιμητής T_2 είναι καλύτερος από τον T_1.
Συνεπως η βέλτιστη επιλογή εκτιμητή είναι η εύρεση αμερόληπτων εκτιμητών ελαχίστης διασποράς (α.ε.ε.δ.). Σε ορισμένες περιπτώσεις που δε μπορούμε να βρούμε α.ε.ε.δ., αναζητούμε ασυμπτωτικούς α.ε.ε.δ. ($n \to \infty$).

Αποτελεσματικότητα οποιανδήποτε εκτιμητών.

Στην προηγούμενη παράγραφο εξετάσαμε την αποτελεσματικότητα εκτιμητών που είναι αμερόληπτες

προτιμάμε α.ε.ε.δ.

Τι επιλογή εκτιμητών κάνουμε όταν διαθέτουμε και μη αμερόληπτους εκτιμητές;

Ορισμός

Έστω ο εκτιμητής $\hat{\theta} = T(X_1, X_2, \ldots, X_n)$ *της παραμέτρου* θ. *Το μέσο τετραγωνικό σφάλμα ορίζεται ως*

$$mse(\hat{\theta}) = E((\hat{\theta} - \theta)^2) \quad (= Var(\hat{\theta}) + [b(\hat{\theta})]^2 \quad \text{Άσκηση})$$

Στην ειδική περίπτωση όπου $E(\hat{\theta}) = \theta$ δηλαδή έχουμε αμεροληψία, τότε

$$mse(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = E((\hat{\theta} - E(\hat{\theta}))^2) = Var(\hat{\theta})$$

Βέλτιστος εκτιμητής είναι αυτός που έχει το μικρότερο μέσο τετραγωνικό σφάλμα.

Μια ακόμα επιθυμητή ιδιότητα είναι η **συνέπεια**.

Ορισμός

Ένας εκτιμητής $\hat{\theta} = \hat{\theta}_n = T(X_1, X_2, \ldots, X_n)$ *καλείται συνεπής όταν:*

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta \ \ (\delta\eta\lambda\alpha\delta\acute{\eta} \lim_{n\to\infty} b(\theta_n) = 0)$$

και

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0.$$

Τα δύο παραπάνω δίνουν ισοδύναμα

$$\lim_{n\to\infty} mse(\hat{\theta}_n) = \lim_{n\to\infty} \left(Var(\hat{\theta_n}) + [b(\hat{\theta}_n)]^2 \right) = 0$$

Για παράδειγμα ο δειματικός μέσος $\overline{X}$ παρατηρήσεων που προέρχονται από μία κατανομή με μέση τιμή μ και διασπορά σ^2 είναι συνεπής εκτιμητής του μ. Πράγματι, έχουμε

$$E(\overline{X}) = \mu, \quad Var(\overline{X}) = \frac{\sigma^2}{n},$$

οπότε

$$\lim_{n\to\infty} E(\overline{X}) = \lim_{n\to\infty} \mu = \mu$$

και

$$\lim_{n\to\infty} Var(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

(Ισοδύναμα

$$\lim_{n\to\infty} mse(\overline{X}) = \lim_{n\to\infty} \left(Var(\overline{X}) + [b(\overline{X})]^2 \right) = \lim_{n\to\infty} \left(\frac{\sigma^2}{n} + 0^2 \right) = 0.$$

)

Τ. Παπαιωάννου και Λούκας Σ.Β. Εισαγωγή στη Στατιστική, 2002.

Χ.Κ. Αγιακλόγλου και Θ.Ε. Μπένος. Εισαγωγή στην Οικονομετρική Ανάλυση, 2003.

Γ. Ηλιόπουλος. Βασικές Μέθοδοι Εκτίμησης Παραμέτρων, 2006.

Χ.Κ. Φράγκος. Στατιστική Επιχειρήσεων, 1998.

Δ.Α. Ιωαννίδης. Στατιστικές Μέθοδοι, 1999.

Μ. Μπούτσικας. Στατιστική ΙΙΙ, Σημειώσεις 2003.