“Πιθανότητες και Αρχές Στατιστικής” (9η Διάλεξη)

Σωτήρης Νικολετσέας, καθηγητής

Τμήμα Μηχανικών Η/Υ & Πληροφορικής,
Πανεπιστήμιο Πατρών

Ακαδημαϊκό Έτος 2023 - 2024

Σημερινό μάθημα: Εκτιμήτριες συναρτήσεις, σημειακή εκτίμηση παραμέτρων και γραμμική παλινδρόμηση

- Στατιστική συμπερασματολογία (ή "εκτιμητική"): εξαγωγή συμπερασμάτων για το σύνολο ενός πληθυσμού με βάση ένα (σχετικά μικρό) τυχαίο υποσύνολό του (δείγμα)
  - Γίνεται μελέτη των δειγματικών δεδομένων και επαγωγή τους στον (γεννήτορα) πληθυσμό (στατιστική επαγωγή), με μια ορισμένη πιθανότητα εμπιστοσύνης στην επιτυχία αυτής της επαγωγικής διαδικασίας.
- Αντιθέτως, η θεωρία Πιθανοτήτων είναι εγγενώς απαγωγική, δηλαδή ουσιαστικά γνώση για τον πληθυσμό (όλον) μεταφέρεται (απάγεται) στο τυχαίο δείγμα (μέρος).
- Παράδειγμα:
  - πιθανοθεωρία: ξέρω ότι ο χρόνος ζωής λαμπτήρων ακολουθεί εκθετική κατανομή με παράμετρο λ, και υπολογίζω την πιθανότητα συγκεκριμένων χρόνων ζωής σε κάποιους λαμπτήρες.
  - στατιστική επαγωγή: ξέρω ότι οι λαμπτήρες έχουν εκθετικά κατανεμημένο χρόνο ζωής, χωρίς να ξέρω την παράμετρο λ, και συμπεραίνω την λ από τους πραγματικούς χρόνους ζωής σε ένα τυχαίο δείγμα λαμπτήρων.

Βασικές μέθοδοι στατιστικής συμπερασματολογίας

- Σημειακές εκτιμήσεις: Στόχος είναι η κατασκευή μιας εκτίμησης για μία τιμή (δηλαδή ένα σημείο και όχι ένα διάστημα της άγνωστης παραμέτρου), μαζί με τον υπολογισμό του σφάλματος.
  Παράδειγμα: η εκτίμηση της παραμέτρου λ για τον (εκθετικά κατανεμημένο) χρόνο ζωής λαμπτήρων
- Διαστήματα εμπιστοσύνης: Αντί για μια τιμή, αναζητούμε ένα φάσμα τιμών (διάστημα). Αυτό είναι ιδιαίτερα χρήσιμο όταν το δείγμα είναι μικρό ή/και όταν το σφάλμα εκτίμησης είναι σχετικά μεγάλο.
  Παράδειγμα: οι φυσιολογικές τιμές ιατρικών εξετάσεων, το αποδεκτό εύρος τιμών κατά τη μέτρηση αστικών ρύπων
- Έλεγχοι στατιστικών υποθέσεων: ελέγχεται αν μια πρόταση αναφορικά με κάποιο χαρακτηριστικό ενός πληθυσμού είναι σωστή

Σημειακές Εκτιμήσεις

- Σε αυτή τη διάλεξη θα ασχοληθούμε με σημειακές εκτιμήσεις.

- Η σημειακή εκτίμηση γίνεται μέσω εύρεσης και αξιολόγησης εκτιμητριών συναρτήσεων.

- Μέθοδοι εύρεσης εκτιμητριών συναρτήσεων:
  - Μέθοδος ροπών
  - Μέθοδος ελαχίστων τετραγώνων
  - Μέθοδος μέγιστης πιθανοφάνειας

- Βασικά κριτήρια αξιολόγησης εκτιμητριών συναρτήσεων:
  - αμεροληψία ή εγκυρότητα
  - αποτελεσματικότητα ή αξιοπιστία
  - συνέπεια

Τυχαίο Δείγμα - αντιστοιχία με Θεωρία Πιθανοτήτων

- Πληθυσμός: ένα σύνολο ανθρώπων, συσκευών κλπ ("στοιχεία") των οποίων μελετώνται ορισμένα χαρακτηριστικά π.χ. το μηνιαίο εισόδημα των ανθρώπων, η διάρκεια ζωής των λαμπτήρων

- Σημαντική αντιστοιχία με Θεωρία Πιθανοτήτων:
  - πληθυσμός $\leftrightarrow$ συνάρτηση κατανομής F
  - χαρακτηριστικό $\leftrightarrow$ τυχαία μεταβλητή X

  δηλαδή, θεωρούμε ότι η θεωρητική συμπεριφορά του πληθυσμού ακολουθεί κάποια συνάρτηση κατανομής F, και ότι οι συγκεκριμένες τιμές του χαρακτηριστικού σε ένα τυχαίο δείγμα περιγράφονται από μια τυχαία μεταβλητή X που έχει την κατανομή F.

Πιθανοθεωρητική Επαγωγή και Στατιστική Απαγωγή

- Αν ξέραμε την F (π.χ. την παράμετρο λ για την εκθετική κατανομή του χρόνου ζωής των λαμπτήρων) θα μπορούσαμε να εκτιμήσουμε πλήρως την πιθανοθεωρητική συμπεριφορά ενός τυχαίου δείγματος π.χ. του χρόνου ζωής X ενός λαμπτήρα.
- Στην πράξη όμως, δεν ξέρουμε ακριβώς την F (δηλαδή, ξέρουμε ότι είναι εκθετική αλλά αγνοούμε την παράμετρο της λ) και έχουμε μόνο ορισμένες πραγματικές μετρήσεις (δηλαδή τιμές της μεταβλητής X), π.χ. παρατηρούμε τους χρόνους ζωής $X_1, X_2, \ldots, X_\nu$ από ν λαμπτήρες. Σκοπός μας είναι να εκτιμήσουμε την άγνωστη παράμετρο λ.

Τυχαίο δείγμα

Ορισμός: Αν ένας πληθυσμός έχει συνάρτηση κατανομής F, καλούμε τυχαίο δείγμα του ένα σύνολο ανεξάρτητων τυχαίων μεταβλητών

$$X_1, X_2, \ldots, X_\nu$$

που όλες ακολουθούν την συνάρτηση κατανομής F. Γράφουμε: $X_1, X_2, \ldots, X_\nu \sim F$ και καλούμε τον αριθμό ν μέγεθος του δείγματος.

Παρατήρηση: Μετά τη δειγματοληψία, οι τ.μ. $X_1, X_2, \ldots, X_\nu$ παίρνουν συγκεκριμένες τιμές (πραγματικές μετρήσεις) $x_1, x_2, \ldots, x_\nu$. Οι στατιστικές μέθοδοι βασίζονται σε αυτές ακριβώς τις μετρήσεις x_i.

Εκτιμήτριες Συναρτήσεις

Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_\nu$ από μία συνάρτηση κατανομής $F(x;\theta)$ όπου η παράμετρος θ είναι άγνωστη, και την οποία προσπαθούμε να "εκτιμήσουμε" με βάση το δείγμα. Για το σκοπό αυτό θα χρησιμοποιήσουμε στατιστικές συναρτήσεις (συναρτήσεις ορισμένες στο δείγμα) που καλούνται εκτιμήτριες.

Ορισμός: Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_\nu$ από τη συνάρτηση κτανομής $F = F(x;\theta)$, όπου θ άγνωστη παράμετρος της F.
Οποιαδήποτε συνάρτηση $T(X)$:

$$T = T(X_1, X_2, \ldots, X_\nu)$$

ορισμένη στο δείγμα χωρίς να εξαρτάται από την άγνωστη παράμετρο θ καλείται εκτιμήτρια συνάρτηση της παραμέτρου.

Παρατήρηση: Ακριβώς επειδή η T σκοπεύει στην εκτίμηση της άγνωστης παραμέτρου θ, δεν μπορεί να εξαρτάται από την θ (ή τυχόν άλλες άγνωστες παραμέτρους). Πολλές φορές η $T(X)$ συμβολίζεται με $\hat{\theta}$ και οι αριθμητικές τιμές που παίρνει καλούνται εκτιμήσεις.

Παράδειγμα Εκτιμητριών Συναρτήσεων

Παράδειγμα: Έστω μηχανή που παράγει εξαρτήματα και υποθέτουμε ότι κάθε παραγόμενο εξάρτημα είναι λειτουργικό με πιθανότητα p και ελαττωματικό με πιθανότητα $1-p$ (που χαρακτηρίζει την ποιότητα της μηχανής), στοχαστικά ανεξάρτητα για τα διάφορα εξαρτήματα. Θέλουμε να εκτιμήσουμε την άγνωστη παράμετρο $\theta = p$ επιτυχούς δημιουργίας εξαρτημάτων παρατηρώντας ένα τυχαίο δείγμα $X_1, X_2, \ldots, X_\nu$ από ν εξαρτήματα. Προφανώς οι X_i ακολουθούν την κατανομή Bernoulli:

$$X_1, X_2, \ldots, X_\nu \sim b(\theta), \qquad 0 < \theta < 1$$

όπου $X_i = \begin{cases} 1, & \text{το } i\text{-οστό εξάρτημα είναι λειτουργικό (πιθανότητα } \theta) \\ 0, & \text{το } i\text{-οστό εξάρτημα είναι ελαττωματικό (πιθανότητα } 1-\theta) \end{cases}$

- Μία προφανής εκτιμήτρια συνάρτηση είναι ο μέσος όρος των λειτουργικών προιόντων στο δείγμα:
$$T_1 = \bar{X} = \frac{X_1 + X_2 + \cdots + X_\nu}{\nu}$$
- Άλλες εκτιμήτριες είναι οι $T_2 = X_\nu$, $T_3 = X_1 - X_3 - X_4$, $T_4 = X_3 \cdot X_5 \cdot X_7$, ή ακόμα και η $T_5 = \frac{1}{3}$ (η οποία είναι ουσιαστικά ανεξάρτητη του δείγματος).
- Ωστόσο οι στατιστικές συναρτήσεις $T = 2\theta$, $T = \theta \cdot X_3 + \frac{\bar{X}}{\theta}$ δεν είναι εκτιμήτριες γιατί εξαρτώνται από την άγνωστη παράμετρο θ.

Παράδειγμα Εκτιμήτριας Συνάρτησης

Παράδειγμα: Έστω ότι οι βαθμολογίες σε μια γραπτή εξέταση ακολουθούν την κανονική κατανομή $N(\mu, \sigma^2)$, με μέση τιμή μ (παράμετρος θ_1) και διασπορά σ^2 (παράμετρος θ_2) που μας είναι άγνωστες και προσπαθούμε να τις συμπεράνουμε δειγματοληπτώντας τυχαία ένα μικρό σύνολο από ν γραπτά των οποίων η βαθμολογία βρίσκεται να είναι $X_1, X_2, \ldots, X_\nu$.

- Εκτιμήτριες για τις άγνωστες παραμέτρους θ_1 και θ_2 είναι οι:

  $$T_6 = \bar{X} = \tfrac{X_1+X_2+\cdots+X_\nu}{\nu} \qquad \text{για την } \theta_1 = \mu$$

  και $$T_7 = S^2 = \frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2 \qquad \text{για την } \theta_2 = \sigma^2$$

  (δηλαδή ο δειγματικός μέσος όρος και η δειγματική διασπορά).

  Προφανώς η στατιστική συνάρτηση

  $$T = \sum_{i=1}^{\nu}(X_i - \theta_1)$$

  δεν είναι εκτιμήτρια, γιατί εξαρτάται από το άγνωστο θ_1.

Αμεροληψία Εκτιμητριών

- Όπως είδαμε, για μια άγνωστη παράμετρο μπορούν να κατασκευαστούν πολλές εκτιμήτριες συναρτήσεις.
- Αναζητώντας μια κατά το δυνατόν "καλή" εκτιμήτρια, ένα πρώτο κριτήριο είναι η αμεροληψία.

Ορισμός: Μια εκτιμήτρια T καλείται αμερόληπτη για την άγνωστη παράμετρο θ, άν

$$E(T) = \theta$$

για όλες τις δυνατές τιμές της θ.

Παρατήρηση: Η T, αφού ορίζεται με βάση τις τ.μ. X_i, είναι και αυτή τυχαία μεταβλητή. Η ιδιότητα της αμεροληψίας απαιτεί από μία "καλή" εκτιμήτρια να στοχεύει την άγνωστη παράμετρο θ, με την έννοια ότι οι τιμές που παίρνει (γύρω από την θ) έχουν μέση τιμή θ.

Αμεροληψία Εκτιμητριών - Παραδείγματα (I)

- Στο παράδειγμα με τα ελαττωματικά προϊόντα έχουμε:

$$E(T_1) = E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_\nu}{\nu}\right) = \frac{1}{\nu}\sum_{i=1}^{\nu} E(X_i) =$$
$$= \tfrac{1}{\nu} \cdot \nu \cdot \theta = \theta$$

$$E(T_2) = E(X_\nu) = \theta$$

$$E(T_3) = E(X_1 - X_3 - X_4) = E(X_1) - E(X_3) - E(X_4) =$$
$$= \theta - \theta - \theta = -\theta$$

$$E(T_4) = E(X_3 X_5 X_7) = E(X_3)E(X_5)E(X_7) = \theta \cdot \theta \cdot \theta = \theta^3$$

$$E(T_5) = \tfrac{1}{3}$$

Άρα αμερόληπτες είναι μόνο οι T_1 και T_2. (Ωστόσο, η T_4 είναι αμερόληπτη για την παράμετρο θ^3).

Αμεροληψία Εκτιμητριών - Παραδείγματα (II)

- Στο παράδειγμα με τις βαθμολογίες είναι:

$$E(T_6) = E(\bar{X}) = E\left(\frac{X_1 + \cdots + X_\nu}{\nu}\right) = \frac{1}{\nu}\sum_{i=1}^{\nu} E(X_i) =$$
$$= \tfrac{1}{\nu} \cdot \nu \cdot \mu = \mu = \theta_1$$

$$E(T_7) = E(S^2) = E\left[\frac{1}{\nu - 1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2\right] =$$
$$= \frac{1}{\nu - 1}\sum_{i=1}^{\nu} E(X_i - \bar{X})^2 = \sigma^2 = \theta_2$$

Άρα η T_6 είναι αμερόληπτη για την θ_1 (όχι όμως για την θ_2) και η T_7 είναι αμερόληπτη για την θ_2 (όχι όμως για την θ_1).

Αποτελεσματικότητα ή Αξιοπιστία Εκτιμητριών Συναρτήσεων

Είδαμε λοιπόν ότι για μια άγνωστη παράμετρο μπορεί να υπάρχουν πολλές αμερόληπτες εκτιμήτριες. Μεταξύ τους, είναι διαισθητικά σωστό να επιλέξουμε ως καλύτερη εκείνη με τη μικρότερη διασπορά, άρα με υψηλότερη συγκέντρωση, γύρω από τη μέση τιμή που (λόγω της αμεροληψίας) είναι η άγνωστη παράμετρος.

Ορισμός: Μεταξύ δύο αμερόληπτων εκτιμητριών T_1 και T_2 για μία άγνωστη παράμετρο θ (οπότε $E(T_1) = E(T_2) = \theta$) η T_1 θα θεωρείται καλύτερη (αποτελεσματικότερη) από την T_2, ως προς το κριτήριο της ελάχιστης διασποράς, άν για κάθε θ:

$$\mathrm{Var}(T_1) \leq \mathrm{Var}(T_2)$$

Παράδειγμα: Στο παράδειγμα με τα ελαττωματικά προϊόντα έχουμε:

$$\mathrm{Var}(T_1)=\mathrm{Var}(\bar{X})=\mathrm{Var}\left(\frac{X_1 + \cdots + X_\nu}{\nu}\right) = \frac{1}{\nu^2}\sum_{i=1}^{\nu}\mathrm{Var}(X_i) =$$

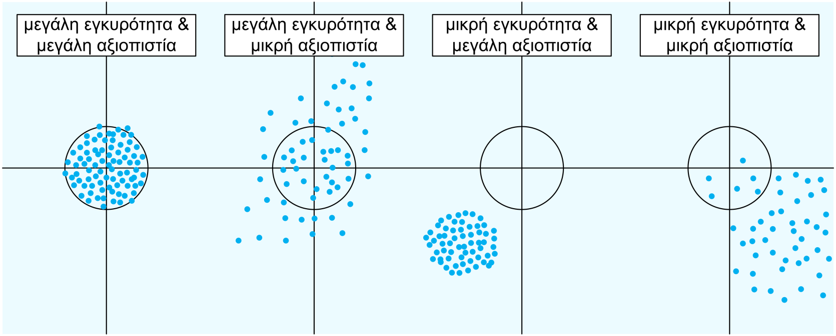
$$= \tfrac{1}{\nu^2} \cdot \nu \cdot \theta(1-\theta) = \tfrac{\theta(1-\theta)}{\nu}$$

ενώ $\mathrm{Var}(T_2)=\mathrm{Var}(X_\nu)=\theta(1-\theta)$, άρα καλύτερη είναι η T_1.

Εγκυρότητα και Αξιοπιστία

- Μια εκτιμήτρια δεν αρκεί να έχει μικρή διασπορά για να θεωρείται καλή, πρέπει επιπλέον να είναι αμερόληπτη, δηλαδή οι τιμές που παίρνει να είναι "γύρω" από την άγνωστη παράμετρο και να έχουν μέση τιμή την παράμετρο θ.
  Π.χ. στο παράδειγμα με τα ελαττωματικά προϊόντα, η σταθερή εκτιμήτρια ($T_5 = \frac{1}{3}$) έχει τη μικρότερη δυνατή διασπορά (0). Ωστόσο, δεν είναι αμερόληπτη γιατί "στοχεύει" πάντα το $\frac{1}{3}$ και όχι το άγνωστο θ. Δηλαδή, η αμεροληψία εγγυάται την "εγκυρότητα" της εκτιμήτριας, υπό την έννοια της αποφυγής θεμελιωδών (συστηματικών) λαθών λόγω της μη στόχευσης της παραμέτρου.
- Από την άλλη μεριά, οι αποκλίσεις της εκτιμώμενης τιμής από την πραγματική, μπορεί να οφείλονται σε τυχαίες επιδράσεις (μικρή ακρίβεια στην εκτίμηση) λόγω μεγάλης διασποράς των τιμών της εκτιμήτριας συνάρτησης σε σειρά επαναλήψεων της μέτρησης. Τα τυχαία σφάλματα σε πολλαπλές επαναλήψεις περιγράφονται από την ιδιότητα της αξιοπιστίας.

Εγκυρότητα και Αξιοπιστία Εκτιμητριών



Κριτήριο μέσου τετραγωνικού σφάλματος

Ποσότητα που επιτυγχάνει τον ταυτόχρονο υπολογισμό τόσο της εγκυρότητας όσο και της αξιοπιστίας.

Ορισμός: Καλούμε "λάθος" της εκτιμήτριας, για συγκεκριμένο δείγμα x, την ποσότητα: $$e(x) = \hat{\theta}(x) - \theta$$
(προφανώς η ποσότητα αυτή εξαρτάται όχι μόνο από την εκτιμήτρια αλλά και το συγκεκριμένο δείγμα).

Ορισμός: Καλούμε μέσο τετραγωνικό σφάλμα (mean square error) τη μέση τιμή του τετραγώνου των λαθών μιας εκτιμήτριας $\hat{\theta}$:
$$MSE(\hat{\theta}) = E[(\hat{\theta}(x) - \theta)^2]$$
Ορισμός: Για συγκεκριμένο δείγμα x, καλούμε δειγματοληπτική απόκλιση της $\hat{\theta}$, την ποσότητα:
$$d(x) = \hat{\theta}(x) - E(\hat{\theta}(x)) = \hat{\theta}(x) - E(\hat{\theta})$$
Ορισμός: Καλούμε διασπορά της εκτιμήτριας $\hat{\theta}$, τη μέση τιμή των τετραγώνων των δειγματοληπτικών αποκλίσεων:
$$\mathrm{Var}(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2]$$
Για αμερόληπτες εκτιμήτριες είναι $E(\hat{\theta}) = \theta$, άρα $\mathrm{Var}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$

Ορισμός: Καλούμε μεροληψία (bias) της εκτιμήτριας $\hat{\theta}$ την ποσότητα
$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Παρατηρήσεις

Είναι:

$$MSE(\hat{\theta}) = \text{Var}(\hat{\theta}) + [B(\hat{\theta})]^2$$

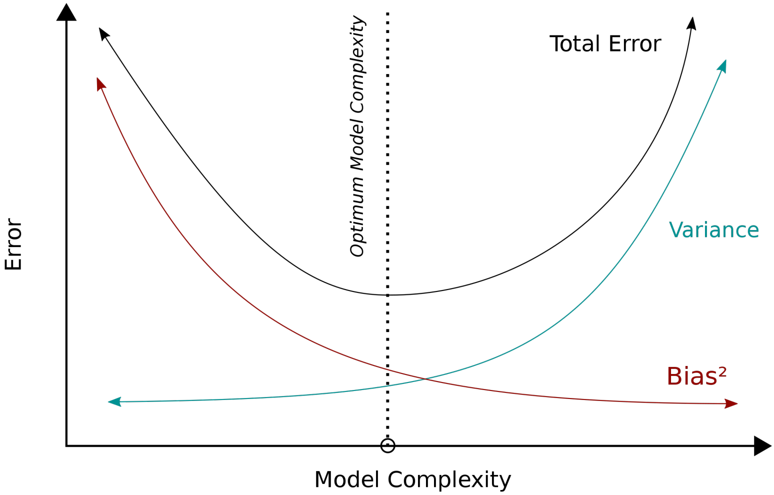
δηλαδή το μέσο τετραγωνικό σφάλμα μιας εκτιμήτριας είναι ίσο με τη διασπορά της συν το τετράγωνο της μεροληψίας της.

<u>Ορισμός</u>: Μεταξύ δυο εκτιμητριών $\hat{\theta}_1$ και $\hat{\theta}_2$ για μια άγνωστη παράμετρο θ, η $\hat{\theta}_1$ θεωρείται καλύτερη της $\hat{\theta}_2$ ως προς το κριτήριο του ελάχιστου τετραγωνικού σφάλματος, αν

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$$

<u>Παρατήρηση</u>: Αν οι $\hat{\theta}_1$ και $\hat{\theta}_2$ είναι αμερόληπτες, τότε $B(\hat{\theta}_1) = B(\hat{\theta}_2) = 0$ και το κριτήριο ουσιαστικά συνίσταται στην επιλογή της εκτιμήτριας με μικρότερη διασπορά (δηλαδή της πιο αποτελεσματικής).

The bias-variance trade-off



The bias-variance trade-off

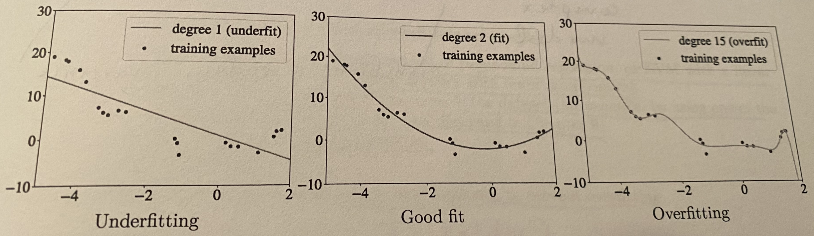


Figure 2: Examples of underfitting (linear model), good fit (quadratic model), and overfitting (polynomial of degree 15).

Δύο βασικές εκτιμήτριες: δειγματικός μέσος και δειγματική διασπορά

- Δειγματικός μέσος:

$$\bar{X} = \bar{X}_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} X_i$$

(πρόκειται επομένως για τον αριθμητικό μέσο των τ.μ. του δείγματος)

- Δειγματική διασπορά:

$$S^2 = S_\nu^2 = \frac{1}{\nu - 1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2$$

Σημαντικό θεώρημα

- <u>Θεώρημα</u>: Έστω $X_1, X_2, \ldots, X_\nu$ τυχαίο δείγμα από την κατανομή F με μέση τιμή μ και διασπορά σ^2. Είναι:
$$E(\bar{X}) = \mu, \quad \mathrm{Var}(\bar{X}) = \tfrac{\sigma^2}{\nu}, \quad E(S^2) = \sigma^2.$$

  Επίσης, αν $\mu_4 = E[(X_i - \mu)^4] < \infty$ τότε
$$\mathrm{Var}(S^2) = \tfrac{1}{\nu}\left(\mu_4 - \tfrac{\nu-3}{\nu-1}\sigma^4\right)$$

Σημαντικό θεώρημα - Απόδειξη

Είναι:

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

Επειδή οι X_i είναι ανεξάρτητες, η διασπορά έχει γραμμικότητα και:

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{\nu}\sum_{i=1}^{\nu} X_i\right) = \frac{1}{\nu^2}\sum_{i=1}^{\nu}\mathrm{Var}(X_i) = \frac{1}{\nu^2}\nu\sigma^2 = \frac{\sigma^2}{\nu}$$

Για την S^2 είναι:

$$\begin{aligned}
S^2 &= \frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i - \bar{X})^2 = \frac{1}{\nu-1}\sum_{i=1}^{\nu}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \\
&= \frac{1}{\nu-1}\sum_{i=1}^{\nu} X_i^2 - \frac{2}{\nu-1}\sum_{i=1}^{\nu} X_i\bar{X} + \frac{1}{\nu-1}\sum_{i=1}^{\nu}\bar{X}^2 = \\
&= \frac{1}{\nu-1}\sum_{i=1}^{\nu} X_i^2 - \frac{2}{\nu-1}\bar{X}\sum_{i=1}^{\nu} X_i + \frac{1}{\nu-1}\nu\bar{X}^2 = \\
&= \frac{1}{\nu-1}\sum_{i=1}^{\nu} X_i^2 - \frac{2}{\nu-1}\bar{X}\nu\bar{X} + \frac{1}{\nu-1}\nu\bar{X}^2 = \\
&= \frac{1}{\nu-1}\left(\sum_{i=1}^{\nu} X_i^2 - \nu\bar{X}^2\right)
\end{aligned}$$

$$\begin{aligned}
\Rightarrow E(S^2) &= \frac{1}{\nu-1}\left(\sum_{i=1}^{\nu} E(X_i^2) - \nu E(\bar{X}^2)\right) = \\
&= \frac{1}{\nu-1}\left(\sum_{i=1}^{\nu}(\mu^2 + \sigma^2) - \nu(\mathrm{Var}(\bar{X}) + E^2(\bar{X}))\right) = \\
&= \frac{1}{\nu-1}\left(\nu(\mu^2 + \sigma^2) - \frac{\nu\sigma^2}{\nu} - \nu\mu^2\right) = \frac{1}{\nu-1}(\nu-1)\sigma^2 = \sigma^2 \qquad \square
\end{aligned}$$

Στοχαστική σύγκλιση τυχαίων μεταβλητών

- Ορισμός: Λέμε ότι η ακολουθία τ.μ. T_ν ($\nu = 1, 2, \dots$) συγκλίνει στοχαστικά (ή αλλιώς συγκλίνει κατά πιθανότητα) στο θ ανν
$$\lim_{\nu \to \infty} \Pr\{|T_\nu - \theta| \geq \epsilon\} = 0$$
για οποιοδήποτε (οσοδήποτε μικρό) $\epsilon > 0$. Γράφουμε $T_\nu \xrightarrow{p} \theta$ καθώς $\nu \to \infty$
- Με άλλα λόγια, καθώς το ν μεγαλώνει, η T_ν πλησιάζει οσοδήποτε θέλουμε την τιμή θ με πολύ μεγάλη πιθανότητα που τείνει στο 1.

Συνέπεια εκτιμητριών συναρτήσεων

- Ορισμός: Έστω τυχαίο δείγμα $X_1, X_2, \ldots, X_\nu$ από την κατανομή $F(x;\theta)$ όπου θ είναι άγνωστη παράμετρος και T_ν ακολουθία εκτιμητριών συναρτήσεων:
$$T_\nu = T_\nu(X_1, X_2, \ldots, X_\nu), \;\; \nu = 1, 2, \ldots$$
Η T_ν καλείται συνεπής για την παράμετρο θ αν
$$T_\nu \xrightarrow{p} \theta \text{ καθώς } \nu \to \infty$$
- Με άλλα λόγια, αν χρησιμοποιούμε συνεπή εκτιμήτρια, τότε $\forall \epsilon > 0$ (οσοδήποτε μικρό) είναι
$$\lim_{\nu \to \infty} \Pr\{|T_\nu - \theta| \geq \epsilon\} = 0$$
και επομένως αν πάρουμε ένα “καταλλήλως μεγάλο” δείγμα πλησιάζουμε “πάρα πολύ” την πραγματική τιμή της θ.

Σχέση συνέπειας με αποτελεσματικότητα και αμεροληψία

Θεώρημα: Αν T_ν αμερόληπτες εκτιμήτριες και $\mathrm{Var}(T_\nu) \to 0$ καθώς $\nu \to \infty$, τότε η T_ν είναι συνεπής.

Απόδειξη: Αφού η T_ν είναι αμερόληπτη, είναι $E(T_\nu) = \theta$, και επομένως (από την ανισότητα Chebyshev):

$$\Pr\{|T_\nu - \theta| \geq \epsilon\} \leq \frac{\mathrm{Var}(T_\nu)}{\epsilon^2} \ \forall \epsilon > 0$$

Άρα, για $\epsilon > 0$ σταθερό, είναι

$$0 \leq \Pr\{|T_\nu - \theta| \geq \epsilon\} \leq \frac{\mathrm{Var}(T_\nu)}{\epsilon^2} \to 0 \text{ καθώς } \nu \to \infty$$

Άρα $T_\nu \xrightarrow{p} \theta$. □

Ασυμπτωτική αμεροληψία

Πρόκειται για ασθενέστερη συνθήκη σε σχέση με την αμεροληψία, αφού αντί για $E(T_\nu) = \theta$ αρκεί

$$E(T_\nu) \to \theta \text{ καθώς } \nu \to \infty$$

Θεώρημα: Αν T_ν ασυμπτωτικά αμερόληπτη εκτιμήτρια και $\mathrm{Var}(T_\nu) \to 0$ (καθώς $\nu \to \infty$), τότε η T_ν είναι συνεπής.

Απόδειξη: Έστω $E(T_\nu) = \delta_\nu \to \theta$. Τότε

$$\Pr\{|T_\nu - \theta| \geq \epsilon\} \leq \frac{E[(T_\nu-\theta)^2]}{\epsilon^2} \tag{1}$$

Αλλά, $E[(T_\nu - \theta)^2] = E[((T_\nu - \delta_\nu) + (\delta_\nu - \theta))^2] =$

$$= \mathrm{Var}(T_\nu) + E[(\delta_\nu - \theta)^2] + 2E[(T_\nu - \delta_\nu)(\delta_\nu - \theta)] =$$
$$= \mathrm{Var}(T_\nu) + (\delta_\nu - \theta)^2 + 2(\delta_\nu - \theta)E[(T_\nu - \delta_\nu)] =$$
$$= \mathrm{Var}(T_\nu) + (\delta_\nu - \theta)^2 \to 0$$

αφού $\mathrm{Var}(T_\nu) \to 0$ και $\delta_\nu \to \theta$. Άρα

$$E[(T_\nu - \theta)^2] \to 0 \text{ καθώς } \nu \to \infty$$

και η συνέπεια προκύπτει από την (1).

Ασυμπτωτική αμεροληψία

Παρατήρηση 1: Μια αμερόληπτη εκτιμήτρια δεν είναι πάντα συνεπής (πρέπει επιπλέον να έχει διασπορά που τείνει στο μηδέν καθώς το μέγεθος του δείγματος αυξάνει). Αντιθέτως, μια συνεπής εκτιμήτρια είναι οπωσδήποτε ασυμπτωτικά αμερόληπτη.

Παρατήρηση 2: Η συνέπεια (όπως και η ασυμπτωτική αμεροληψία) είναι ασυμπτωτικές ιδιότητες, δηλαδή αφορούν σε πολύ μεγάλα, μεταβαλλόμενα μεγέθη δειγμάτων. Αντιθέτως, η αμεροληψία και η αποτελεσματικότητα αναφέρονται σε δείγματα σχετικά μικρού, σταθερού μεγέθους.

Σύνοψη

- ασυμπτωτική αμεροληψία: $E(T_\nu) \to \theta$
- αμεροληψία: $E(T_\nu) = \theta$
- συνέπεια: $T_\nu \xrightarrow{p} \theta$

Δειγματικός μέσος και δειγματική διασπορά

Θεώρημα: Έστω $X_1, X_2, \ldots, X_\nu$ τυχαίο δείγμα από κατανομή F με μέση τιμή μ και διασπορά $\sigma^2 < \infty$. Η ακολουθία των δειγματικών μέσων

$$\bar{X} = \bar{X}_\nu = \frac{1}{\nu} \sum_{i=1}^{\nu} X_i$$

είναι συνεπής για την μέση τιμή μ της κατανομής, ενώ η ακολουθία των δειγματικών διασπορών

$$S^2 = S_\nu^2 = \frac{1}{\nu - 1} \sum_{i=1}^{\nu} (X_i - \bar{X})^2$$

είναι συνεπής για το σ^2.

Απόδειξη: Δείξαμε πριν ότι $E(\bar{X}) = \mu$ και $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{\nu} \to 0$ καθώς $\nu \to \infty$ και επομένως η $\bar{X}$ είναι συνεπής για τη μ (επίσης είναι αμερόληπτη). Επίσης, δείξαμε ότι $E(S^2) = \sigma^2$ και $\mathrm{Var}(S^2) = \frac{1}{\nu} \left(\mu_4 - \frac{\nu-3}{\nu-1} \right) \sigma^4 \to 0$ καθώς $\nu \to \infty$, και άρα η S^2 είναι συνεπής (και αμερόληπτη) για την σ^2.

Δειγματικός μέσος και δειγματική διασπορά

Παρατήρηση: Επομένως, μπορούμε πάντα να βρίσκουμε συνεπείς (και αμερόληπτες) εκτιμήτριες για την πληθυσμιακή μέση τιμή και την πληθυσμιακή διασπορά για ΟΠΟΙΑΔΗΠΟΤΕ κατανομή, υπό την έννοια ότι τέτοιες εκτιμήτριες είναι, για οποιαδήποτε κατανομή, ο δειγματικός μέσος όρος $\bar{X}$ (για τη μ) και η δειγματική διασπορά S^2 (για τη σ^2).

Μέθοδοι εύρεσης εκτιμητριών

- Μέθοδος των ροπών
- Μέθοδος μέγιστης πιθανοφάνειας
- Μέθοδος ελάχιστων τετραγώνων

Πρόκειται για <u>γενικές μεθοδολογίες</u> (επιπλέον των ad hoc μεθόδων και των θεωρητικών αναλύσεων) που επιτρέπουν την υιοθέτηση <u>κάποιας αρχής βελτιστοποίησης</u> κατά την κατασκευή εκτιμητριών <u>με επιθυμητές ιδιότητες</u>.

Μέθοδος των ροπών

- Είναι η παλαιότερη μέθοδος
- Βασική ιδέα:
  - εξίσωση μερικών (έστω K) από τις πρώτες ροπές της πληθυσμιακής κατανομής περί το μηδέν με τις αντίστοιχες ροπές του τυχαίου δείγματος.
  - Παίρνουμε τόσες εξισώσεις (ροπές) όσος είναι ο αριθμός των αγνώστων παραμέτρων που θέλουμε να εκτιμήσουμε
  - Λύνουμε το $K \times K$ σύστημα και βρίσκουμε εκτιμήσεις των παραμέτρων
- Θυμίζουμε ότι, οι πληθυσμιακές ροπές τάξης r είναι $m_r = E(X^r)$ ενώ η δειγματική ροπή τάξης r είναι $\nu_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
- Παρατηρήσεις:
  - Οι προκύπτουσες εκτιμήτριες είναι συχνά μεροληπτικές
  - Επίσης, είναι δύσκολο να διατυπωθεί συμπέρασμα σχετικά με την αποτελεσματικότητα
  - Ωστόσο η μέθοδος είναι πολύ απλή και κάτω από ορισμένες συνθήκες οι εκτιμήτριες που παράγει είναι συνεπείς.

Παράδειγμα

Έστω $X_1, X_2, \ldots, X_\nu$ τυχαίο δείγμα από ομοιόμορφο πληθυσμό (κατανομή $U(a,b)$) με παραμέτρους $a = 0$ και $b = \theta$ (άγνωστο) α) Να βρεθεί εκτιμήτρια της παραμέτρου b με τη μέθοδο των ροπών και β) να ελεγχθεί η συνέπειά της.

<u>Λύση</u>: α) Η πρώτη πληθυσμιακή ροπή περί την αρχή είναι:

$$m_1 = E(X^1) = \frac{a+b}{2} = \frac{\theta}{2}$$

Η πρώτη δειγματική ροπή περί το μηδέν είναι:

$\nu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

Άρα από το (1×1) σύστημα $m_1 = \nu_1$ παίρνουμε $\frac{\theta}{2} = \bar{X}$ και η εκτιμήτρια είναι $\hat{\theta} = 2\bar{X}$.

β) Είναι

$E(\hat{\theta}) = E(2\bar{X}) = E(\frac{2}{n}(X_1 + X_2 + \cdots + X_n) = \frac{2}{n} n E(X_i) = 2\frac{\theta}{2} = \theta$

και

$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(2\bar{X}) = 4\mathrm{Var}(\frac{X_1+X_2+\cdots+X_n}{n}) = \frac{4}{n^2} \cdot n \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n} \to 0$

καθώς $n \to \infty$ και επομένως η $\hat{\theta} = 2\bar{X}$ είναι συνεπής.

Παράδειγμα

Να βρεθούν με τη μέθοδο των ροπών εκτιμήτριες συναρτήσεις της μέσης τιμής και της διασποράς σ^2 ενός κανονικού πληθυσμού με κατανομή $N(\mu, \sigma^2)$

Λύση:
Έχουμε 2 άγνωστες παραμέτρους, άρα παίρνουμε 2 ροπές (εξισώσεις).
Οι πληθυσμιακές ροπές 1ης και 2ης τάξης είναι:

$$m_1 = E(X) = \mu \text{ και } m_2 = E(X^2) = \mathrm{Var}(X) + E^2(X) = \sigma^2 + \mu^2$$

Οι αντίστοιχες δειγματικές ροπές είναι:

$$\nu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \text{ και } \nu_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Άρα, το 2×2 σύστημα είναι:

$$\begin{cases} m_1 = \nu_1 \\ m_2 = \nu_2 \end{cases} \Leftrightarrow \begin{cases} \mu = \bar{X} \ (1) \\ \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \ (2) \end{cases}$$

Παράδειγμα (συνέχεια)

Άρα, η εκτιμήτρια της μ είναι η $\hat{\theta_1} = \bar{X}$ και είναι συνεπής. Η εκτιμήτρια $\hat{\theta_2}$ της σ^2 είναι:

$$(2) \Rightarrow \sigma^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \mu^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) = \\ = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

την οποία και παίρνουμε ως εκτιμήτρια της διασποράς σ^2. Είναι $\hat{\theta_2} = \frac{n-1}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = (1 - \frac{1}{n})S^2$

όπου S^2 η δειγματική διασπορά.

Επομένως, $E(\hat{\theta_2}) = E[(1 - \frac{1}{n})S^2] = (1 - \frac{1}{n})E(S^2) = (1 - \frac{1}{n})\sigma^2$

Άρα $E(\hat{\theta_2}) \neq \sigma^2$ και η $\hat{\theta_2}$ δεν είναι αμερόληπτη (αλλά είναι ασυμπτωτικά αμερόληπτη).

(χρησιμοποιήθηκε σχέση από την διαφάνεια 46 της 8ης διάλεξης).

Μέθοδος μέγιστης πιθανοφάνειας

- Είναι μια από τις πλέον σημαντικές μεθόδους.
- Βασική ιδέα: εύρεση εκτιμητριών τέτοιων ώστε να μεγιστοποιείται η πιθανότητα εμφάνισης των τιμών του συγκεκριμένου τυχαίου δείγματος.
- Έστω τυχαίο δείγμα από κατανομή $f(x;\theta)$. Η (πληθυσμιακή) κατανομή πιθανότητας του δείγματος είναι προφανώς (θεωρούμε το θ σταθερό):
$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$
Αν τώρα θεωρήσουμε σταθερές τις (πραγματοποιημένες) τιμές του δείγματος και την θ ως μεταβλητή, θα έχουμε τη συνάρτηση πιθανοφάνειας (likelihood) του δείγματος:
$$L(x_1, x_2, \ldots, x_n; \theta) = L(\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta)$$
(ουσιαστικά πρόκειται για την ίδια ποσότητα, απλώς αλλάζουν οι σταθερές και η μεταβλητή.)
- Συνήθως, για κάθε άγνωστη παράμετρο θ, υπάρχει μια μοναδική τιμή του θ που μεγιστοποιεί τη συνάρτηση πιθανοφάνειας $L(\theta)$. Η τιμή αυτή ονομάζεται εκτιμήτρια μέγιστης πιθανοφάνειας (MLE) της θ και συμβολίζεται $\hat{\theta}$.

Μέθοδος μέγιστης πιθανοφάνειας

Παρατήρηση:
Η αρχή της μέγιστης πιθανοφάνειας διαισθητικά βασίζεται στο ότι το παρατηρηθέν δείγμα είναι ένα γεγονός $E = \{x_1, x_2, \ldots, x_n\}$ και η συνάρτηση πιθανοφάνειας είναι η πιθανότητα εμφάνισής του.
Αφού όμως το γεγονός συνέβη πρέπει να έχει μεγάλη πιθανότητα, και η μεγιστοποίηση της $L(\theta) = L(x_1, x_2, \ldots, x_n; \theta)$ σημαίνει επιλογή εκείνης της τιμής του θ που μεγιστοποιεί την πιθανότητα εμφάνισης του $E = \{x_1, x_2, \ldots, x_n\}$.

Περιγραφή της MLE μεθόδου

Ουσιαστικά χρησιμοποιούμε κλασσικές μεθόδους εύρεσης μέγιστων τιμών παραγωγίσιμων συναρτήσεων:

1. Κατασκευάζουμε τη συνάρτηση πιθανοφάνειας των δεδομένων του δείγματος με βάση την κατανομή που ακολουθούν:

$$L(\theta) = L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

2. Παίρνουμε το λογάριθμο της $L(\theta)$ (για διευκόλυνση των πράξεων) και παραγωγίζουμε ως προς την άγνωστη θ, βρίσκουμε δηλαδή την

$$\frac{d \ln L(\theta)}{d\theta}$$

3.Βρίσκουμε την τιμή της θ (υποψήφια εκτιμήτρια $\hat{\theta}$) ώστε

$$\frac{d \ln L(\theta)}{d\theta} = 0 \Rightarrow \hat{\theta} = \ldots$$

4. Δείχνουμε ότι η δεύτερη παράγωγος είναι αρνητική ώστε η λύση να αντιστοιχεί σε ολικό μέγιστο:

$$\left. \frac{d^2 \ln L(\theta)}{d^2\theta} \right|_{\theta=\hat{\theta}} < 0$$

Παράδειγμα

Δίνεται τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ από εκθετική κατανομή άγνωστης μέσης τιμής θ. Να βρεθεί εκτιμήτρια της θ με την MLE μέθοδο.

Λύση: Η συνάρτηση πιθανοφάνειας είναι:

$L(\theta) = \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \left(\frac{1}{\theta}\right)^n e^{\frac{-\sum_{i=1}^n x_i}{\theta}} \Rightarrow$

$\Rightarrow \ln L(\theta) = \ln\left(\frac{1}{\theta}\right)^n + \ln e^{\frac{-\sum_{i=1}^n x_i}{\theta}} = \cdots = -n\ln\theta - \frac{\sum_{i=1}^n x_i}{\theta}$

Άρα $\frac{d\ln L(\theta)}{d\theta} = 0 \Leftrightarrow \cdots \Leftrightarrow \theta = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$

και η υποψήφια εκτιμήτρια είναι η $\hat{\theta} = \bar{X}$

Επίσης $\frac{d^2\ln L(\theta)}{d^2\theta} < 0 \Leftrightarrow \cdots \Leftrightarrow n\theta < 2n\bar{X}$ που ισχύει για $\theta = \hat{\theta} = \bar{X}$ και επομένως η MLE είναι πράγματι η $\hat{\theta} = \bar{X}$.

Μέθοδος ελάχιστων τετραγώνων

- Η μέθοδος αποσκοπεί στην εύρεση εκτιμητριών που βελτιστοποιούν όσο το δυνατόν περισσότερες επιθυμητές ιδιότητες.
- Κριτήριο επιλογής: Η ελαχιστοποίηση του αθροίσματος των τετραγώνων των αποκλίσεων ("σφαλμάτων") μεταξύ των δειγματοληπτικών τιμών και των προς εκτίμηση (θεωρητικών, από την κατανομή) τιμών της τυχαίας μεταβλητής.
- Παράδειγμα: Αν προσπαθούμε να εκτιμήσουμε την μέση τιμή μ ενός πληθυσμού, επιλέγουμε ως εκτιμήτρια την τιμή $\hat{\mu}$ του μ που για το συγκεκριμένο δείγμα ελαχιστοποιεί την συνάρτηση Q:

$$Q = \sum_{i=1}^{n} (X_i - \mu)^2$$

Μέθοδος ελάχιστων τετραγώνων

Η ελαχιστοποίηση της Q γίνεται με παραγώγιση (όπως και στην MLE). Στο συγκεκριμένο παράδειγμα προκύπτει εύκολα ότι η παράγωγος της Q (ως προς μ) είναι

$$2 \cdot \mu \cdot n - 2 \cdot \sum_{i=1}^{n} X_i$$

που μηδενίζεται (και επίσης έχει ελάχιστο) για $\mu = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$.

Άρα τελικώς επιλέγουμε:

$$\hat{\mu} = \bar{X}$$

Παρατήρηση: Η μέθοδος οδηγεί σε εκτιμήτριες που ελαχιστοποιούν την διακύμανση (μεταξύ όλων των αμερόληπτων εκτιμητριών). Αρκετά γενικά, οι εκτιμήτριες αυτές είναι συνεπείς.

Γραμμική Παλινδρόμηση (Linear Regression)

- Συχνά εξετάζουμε τη σχέση μεταξύ 2 ή περισσότερων μεταβλητών.
- π.χ. τη σχέση ύψους και βάρους, το ποσοστό υγρασίας σε ένα χημικό προϊόν σε σχέση με την υγρασία του περιβάλλοντος στο οποίο παράγεται.
- συνήθως υπάρχει μία μοναδική εξαρτημένη μεταβλητή Y η οποία εξαρτάται από τις τιμές μιας (ή περισσότερων) ανεξάρτητων μεταβλητών εισόδου $x_1, x_2, \ldots, x_r$
- ο σκοπός μας είναι, παρατηρώντας ένα μικρό δείγμα, να μπορέσουμε να συμπεράνουμε τη σχέση της Y ως προς τις μεταβλητές x_i.

Γραμμική Παλινδρόμηση (Linear Regression)

- Συχνά η σχέση αυτή είναι γραμμική, δηλαδή υπάρχουν (άγνωστες) σταθερές $\beta_0, \beta_1, \ldots, \beta_r$ ώστε:

  $$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r$$

  Αν μπορέσουμε να συμπεράνουμε τα β_i τότε μπορούμε να προβλέψουμε με ακρίβεια τις τιμές της Y.

- Στην πράξη στη γραμμική αυτή σχέση υπεισέρχεται ένα τυχαίο λάθος ϵ και η σχέση γίνεται

  $$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \epsilon$$

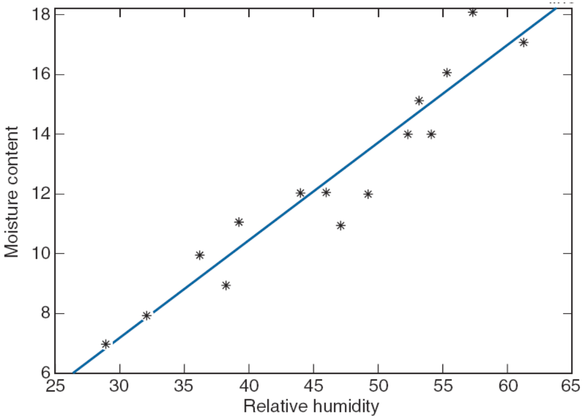
  όπου ϵ τυχαία μεταβλητή με μέση τιμή 0.

  Η σχέση αυτή λέγεται εξίσωση γραμμικής παλινδρόμησης. Οι συντελεστές β_i λέγονται συντελεστές παλινδρόμησης και λέμε ότι η εξαρτημένη μεταβλητή Y παλινδρομεί στις ανεξάρτητες μεταβλητές x_i.

Απλή γραμμική παλινδρόμηση

- Όταν $r = 1$ (δηλαδή η εξαρτημένη μεταβλητή είναι μία) τότε μιλάμε για απλή γραμμική παλινδρόμηση με εξίσωση

$$Y = \alpha + \beta x + \epsilon$$



Εκτίμηση ελαχίστων τετραγώνων των παραμέτρων της γραμμικής παλινδρόμησης

- Έστω ότι παρατηρούμε τις τιμές Y_i ως προς τις εισόδους x_i $(i = 1, 2, \ldots, n)$ και θέλουμε να υπολογίσουμε τα α και β.
- Σκεφτόμαστε ότι, αν A είναι εκτίμηση της α και B είναι εκτίμηση της β, τότε η εκτίμηση για την είσοδο x_i θα είναι $A + Bx_i$ και το τετράγωνο της απόκλισης από την πραγματική παρατηρηθείσα τιμή Y_i θα είναι $(Y_i - A - Bx_i)^2$
- Το άθροισμα των τετραγώνων των αποκλίσεων αυτών είναι

$$SS = \sum_{i=1}^{n} (Y_i - A - Bx_i)^2$$

- Η μέθοδος συνίσταται ακριβώς στην εύρεση των A και B που ελαχιστοποιούν αυτό το άθροισμα.

Απλή γραμμική παλινδρόμηση (II)

Θεώρημα: Οι εκτιμητές ελαχίστων τετραγώνων των β, α για τις τιμές Y_i, x_i $(i = 1, 2, \ldots, n)$ είναι:

$$B = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{x} \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$A = \bar{Y} - B\bar{x}$$

Η ευθεία γραμμή $A + Bx$ καλείται ευθεία παλινδρόμησης.

Παράδειγμα: Στο παράδειγμα με την υγρασία εύκολα βρίσκουμε ότι $\alpha = -2.51$, $\beta = 0.32$ άρα η ευθεία παλινδρόμησης είναι

$$Y = -2.51 + 0.32x$$