Ασκήσεις Εβδομάδας 1

Εμπειρικές εκτιμήτριες, μέση τιμή και διασπορά

Συμβολισμοί. Οι $X_1, \dots, X_n$ αποτελούν τυχαίο δείγμα, ενώ $x_1, \dots, x_n$ είναι οι παρατηρηθείσες τιμές. Για σταθερό μέγεθος δείγματος γράφουμε

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad M_k' = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad M_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k.$$

Οι αντίστοιχες πληθυσμιακές ροπές είναι $\mu = \mathbb{E}(X_1)$, $\mu_k' = \mathbb{E}(X_1^k)$ και $\mu_k = \mathbb{E}[(X_1 - \mu)^k]$, όταν υπάρχουν οι απαιτούμενες απόλυτες ροπές. Τα πεζά $\overline{x}, m_k', m_k$ δηλώνουν τιμές στα δεδομένα. Ο δείκτης του μεγέθους διατηρείται μόνο όταν συγκρίνουμε διαφορετικά δείγματα.

$$\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le t\}}, \qquad F^{-}(u) = \inf\{t \in \mathbb{R} : F(t) \ge u\}, \quad 0 < u < 1.$$

Τα $\mathbb{P}, \mathbb{E}, \mathbb{V}, \mathrm{Cov}$ δηλώνουν πιθανότητα, μέση τιμή, διασπορά και συνδιακύμανση. Με $X_{(1)} \le \cdots \le X_{(n)}$ συμβολίζουμε τις διατεταγμένες παρατηρήσεις, με τις επαναλήψεις τους.

Για συγκεκριμένη πυκνότητα ή συνάρτηση πιθανότητας f χρησιμοποιούμε τη σύμβαση $S_f = \{t : f(t) > 0\}$. Στις ασκήσεις εκτίμησης πιθανότητας επιτρέπονται εκτιμήσεις στο $[0, 1]$. Τα αριθμητικά δεδομένα του φύλλου είναι υποθετικά.

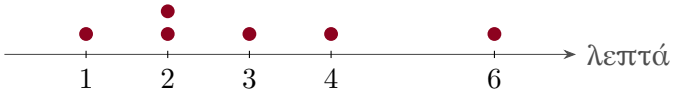
Αναφορές. G. Casella και R. L. Berger, *Στατιστική Συμπερασματολογία*, ελληνική μετάφραση της 2ης έκδοσης, Λήμμα 5.2.5 και Ασκήσεις 5.3, 5.15, 5.21. Σ. Τρέβεζας, *Μη Παραμετρική Στατιστική*, Κάλλιπος (2025), §§3.1–3.2 και 3.12.

Άσκηση 1. Έξι χρόνοι αναμονής

Βασική

Σε μια στάση καταγράφηκαν οι χρόνοι αναμονής έξι επιβατών, σε λεπτά:

$$x = (1, 2, 2, 3, 4, 6).$$



Σχήμα 1. Κάθε σημείο αντιστοιχεί σε μία παρατήρηση. Οι ίσες τιμές δεν παραλείπονται.

Στην άσκηση παραλείπουμε τον δείκτη του μεγέθους του δείγματος.

(α) Φτιάξτε πίνακα συχνοτήτων. Πόσο βάρος δίνει η εμπειρική κατανομή σε κάθε διαφορετική τιμή;

(β) Γράψτε και σχεδιάστε την εμπειρική συνάρτηση κατανομής $\widehat{F}(t)$ για κάθε $t \in \mathbb{R}$. Ποιο είναι το μέγεθος του άλματος στο 2;

(γ) Εκτιμήστε εμπειρικά τις πιθανότητες αναμονής ακριβώς 2 λεπτών, περισσότερο από 3 λεπτά, και περισσότερο από 2 αλλά το πολύ 4 λεπτά. Γράψτε τα αντίστοιχα ενδεχόμενα.

(δ) Υπολογίστε τον μέσο $\overline{x}$, τη δεύτερη ροπή m_2', την εμπειρική διασπορά m_2, το εύρος r και τη συνήθη διάμεσο d.

(ε) Ποιο είναι το στήριγμα της εμπειρικής συνάρτησης πιθανότητας; Είναι το διάστημα $[1, 6]$;

(στ) Υποθέτουμε το μοντέλο $\mathrm{Unif}[0, \theta]$, όπου $\theta > 0$ είναι το σταθερό άγνωστο μεσοδιάστημα των λεωφορείων. Βρείτε τις εκτιμήσεις από τους κανόνες $\widehat{\theta} = X_{(n)}$ και $\overline{\theta} = 2\overline{X}$. Δώστε άλλο δείγμα έξι θετικών χρόνων στο οποίο οι δύο τιμές διαφέρουν.

Άσκηση 2. Διάρκεια χρήσης και χρέωση

Βασική–μέση

Οι $X_1, \dots, X_n$ είναι τυχαίο δείγμα μη αρνητικών χρόνων χρήσης μιας υπηρεσίας. Για γνωστή Borel-μετρήσιμη συνάρτηση h, θέτουμε

$$T_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i).$$

(α) Αν $\mathbb{E}|h(X_1)| < \infty$, υπολογίστε την $\mathbb{E}(T_n)$. Αν επιπλέον $\mathbb{E}[h^2(X_1)] < \infty$, βρείτε τη $\mathbb{V}(T_n)$.

(β) Η χρέωση είναι 2 ευρώ σταθερά και 1 ευρώ ανά λεπτό: $h(x) = 2 + x$. Ποιο χαρακτηριστικό εκτιμά η T_n; Βρείτε μέση τιμή και διασπορά της, για $\sigma^2 < \infty$.

(γ) Εφαρμόστε το (α) στην M_k'. Εκφράστε τα αποτελέσματα με τις μ_k', μ_{2k}' και δηλώστε τις αναγκαίες υποθέσεις ύπαρξης.

(δ) Για $\mathbb{E}(X_1^2) < \infty$, βρείτε την $\mathbb{E}(\overline{X}^{\,2})$. Ποια χαρακτηριστικά εκτιμούν εμπειρικά οι M_2' και $\overline{X}^{\,2}$; Είναι πάντοτε ίσα;

Άσκηση 3. Ποιο όριο χρησιμοποιούμε;

Μέση

Έστω τυχαίο δείγμα χρόνων αναμονής με άγνωστη συνεχή συνάρτηση κατανομής F και πεπερασμένη άγνωστη μέση τιμή μ.

(α) Για $p = \mathbb{P}(X_1 > 5)$, θέτουμε $\widehat{p}_n = n^{-1} \sum_i \mathbf{1}_{\{X_i > 5\}}$. Βρείτε την κατανομή του $n\widehat{p}_n$, τη μέση τιμή και τη διασπορά της $\widehat{p}_n$.

(β) Αν αντικαταστήσουμε το 5 με το άγνωστο μ, προκύπτει στατιστική συνάρτηση; Ποια είναι η κατανομή του πλήθους των παρατηρήσεων πάνω από το μ;

(γ) Αν αντικαταστήσουμε το 5 με τον $\overline{X}$, ο κανόνας υπολογίζεται από το δείγμα. Για $n = 2$, δείξτε ότι ακριβώς μία παρατήρηση υπερβαίνει τον δειγματικό μέσο, με πιθανότητα 1. Είναι ανεξάρτητες οι δύο δείκτριες;

(δ) Αν δεν παρατηρηθεί καμία αναμονή πάνω από 5 λεπτά, αποδείξαμε ότι $p = 0$; Βρείτε την πιθανότητα αυτού του αποτελέσματος για $0 < p < 1$.

Άσκηση 4. Προστίθεται ένας ακόμη επιβάτης

Μέση

Από n χρόνους αναμονής κρατάμε τον $\overline{X}_n$ και το άθροισμα

$$Q_n = \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = n M_{2,n}.$$

Καταγράφεται ένας ακόμη χρόνος X_{n+1}.

(α) Δείξτε ότι $\overline{X}_{n+1} = (n\overline{X}_n + X_{n+1})/(n+1)$.

(β) Αποδείξτε ότι

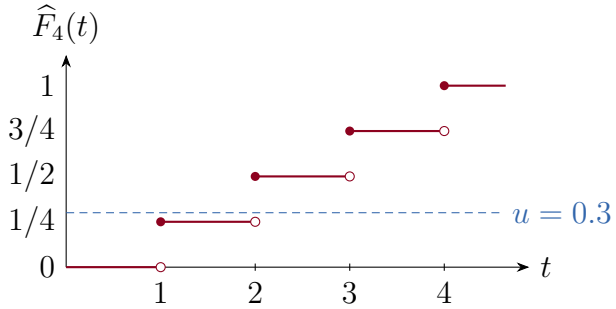
$$Q_{n+1} = Q_n + \frac{n}{n+1}(X_{n+1} - \overline{X}_n)^2.$$

(γ) Στα δεδομένα της Άσκησης 1 προστίθεται χρόνος 3 λεπτών. Βρείτε νέο μέσο και νέα εμπειρική διασπορά χωρίς να υπολογίσετε ξανά όλες τις αποκλίσεις.

Άσκηση 5. Ποσοστημόρια σε τέσσερις παρατηρήσεις

Βασική–μέση

Για τους χρόνους $x = (1, 2, 3, 4)$ δίνεται το γράφημα της εμπειρικής συνάρτησης κατανομής.



Σχήμα 2. Η εμπειρική συνάρτηση κατανομής είναι δεξιά συνεχής. Οι γεμάτοι κύκλοι δηλώνουν την τιμή της στα σημεία άλματος.

(α) Βρείτε τις $\widehat{F}_4^{-}(0.5)$ και $\widehat{F}_4^{-}(0.3)$.

(β) Ποιοι αριθμοί είναι διάμεσοι της εμπειρικής κατανομής; Ελέγξτε τις ανισότητες $\widehat{F}_4(q-) \le 1/2 \le \widehat{F}_4(q)$ για $q = 2.5$.

(γ) Για $u = 0.5$ και $u = 0.3$, υπολογίστε την εκτίμηση με παρεμβολή: $r = 5u$, $j = \lfloor r \rfloor$, $w = r - j$ και $(1-w)x_{(j)} + w x_{(j+1)}$.

(δ) Για $u = 0.3$, η τιμή της παρεμβολής ικανοποιεί τις ανισότητες $\widehat{F}_4(q-) \le u \le \widehat{F}_4(q)$; Είναι η τιμή της γενικευμένης αντιστρόφου;

Άσκηση 6. Πόσοι εξυπηρετούνται έγκαιρα;

Μέση–εμβάθυνση

Οι $X_1, \dots, X_n$ είναι ανεξάρτητοι και ισόνομοι χρόνοι εξυπηρέτησης με συνάρτηση κατανομής F. Για γνωστό σταθερό t, η $\widehat{F}_n(t)$ εκτιμά την πιθανότητα ολοκλήρωσης το αργότερο σε χρόνο t.

(α) Βρείτε τις $\mathbb{E}[\widehat{F}_n(t)]$ και $\mathbb{V}[\widehat{F}_n(t)]$ με χρήση δεικτριών.

(β) Για δύο γνωστά όρια $s < t$, δείξτε ότι

$$\operatorname{Cov}\big(\widehat{F}_n(s), \widehat{F}_n(t)\big) = \frac{F(s)(1 - F(t))}{n}.$$

Υπόδειξη: $\mathbf{1}_{\{X_i \le s\}} \mathbf{1}_{\{X_i \le t\}} = \mathbf{1}_{\{X_i \le s\}}$.

(γ) Ποια πιθανότητα εκτιμά η $\widehat{F}_n(t) - \widehat{F}_n(s)$; Βρείτε τη διασπορά της με δείκτριες και με τον τύπο του (β).

(δ) Βρείτε n που εξασφαλίζει $\mathbb{V}[\widehat{F}_n(t)] \le 1/400$ για κάθε F και κάθε σταθερό t.

Άσκηση 7. Το μέγιστο και η διάμεσος του πληθυσμού

Μέση

Έστω τυχαίο δείγμα από απολύτως συνεχή κατανομή F με πυκνότητα f. Αν m είναι διάμεσος του πληθυσμού, τότε $F(m) = 1/2$.

(α) Βρείτε την $\mathbb{P}(X_{(n)} > m)$, πρώτα για $n = 2$ και έπειτα για γενικό n.

(β) Για $n \ge 2$, βρείτε την πιθανότητα το $[X_{(1)}, X_{(n)}]$ να περιέχει το m. Πότε δεν το περιέχει;

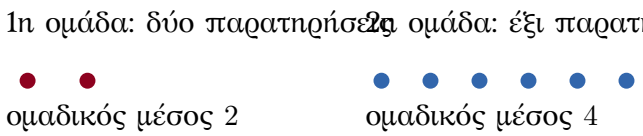
(γ) Επαναλάβετε το (β) για σημείο q_u με $F(q_u) = u$, $0 < u < 1$.

(δ) Μπορεί το μέγιστο να υπερβαίνει τη διάμεσο και ταυτόχρονα να είναι μικρότερο από το άνω άκρο του S_f; Δώστε παράδειγμα.

Άσκηση 8. Δύο ομάδες με διαφορετικό πλήθος

Μέση–εμβάθυνση

Δύο ομάδες καταγραφών είναι οι X_1, X_2 και $Y_1, \dots, Y_6$. Όλες οι παρατηρήσεις είναι ανεξάρτητες και ισόνομες, με μέση τιμή μ και διασπορά $0 < \sigma^2 < \infty$.



Σχήμα 3. Ίσο βάρος στις ομάδες δεν σημαίνει ίσο βάρος στις παρατηρήσεις. Τα σημεία δείχνουν μόνο το πλήθος κάθε ομάδας.

(α) Αν οι ομαδικοί μέσοι είναι 2 και 4 λεπτά, βρείτε τον μέσο των οκτώ παρατηρήσεων. Είναι ο απλός μέσος των δύο ομαδικών μέσων;

(β) Για γνωστό βάρος $0 \le \lambda \le 1$, θέτουμε $T_\lambda = \lambda \overline{X}_2 + (1-\lambda)\overline{Y}_6$. Βρείτε μέση τιμή και διασπορά.

(γ) Ποια λ ελαχιστοποιεί τη διασπορά; Πόσο βάρος παίρνει τότε καθεμία από τις οκτώ παρατηρήσεις;

(δ) Για τυχαίο δείγμα μεγέθους n και γνωστά πραγματικά βάρη με $\sum_i w_i = 1$, δείξτε ότι

$$\mathbb{V}\left(\sum_{i=1}^{n} w_i X_i\right) = \sigma^2 \sum_{i=1}^{n} w_i^2 \ge \frac{\sigma^2}{n}.$$

Πότε υπάρχει ισότητα;

Άσκηση 9. Μια σπάνια μεγάλη αναμονή

Μέση–εμβάθυνση

Στην πρώτη υπηρεσία ο χρόνος εξυπηρέτησης είναι πάντοτε 1 λεπτό. Στη δεύτερη είναι 1 λεπτό με πιθανότητα 0.99 και 20 λεπτά με πιθανότητα 0.01.

(α) Βρείτε σε κάθε υπηρεσία το στήριγμα της συνάρτησης πιθανότητας, τη μέση τιμή και το άνω άκρο του στηρίγματος.

(β) Για δέκα ανεξάρτητες καταγραφές, ποια είναι η πιθανότητα να εμφανιστούν μόνο χρόνοι ενός λεπτού; Ποια εμπειρική κατανομή προκύπτει;

(γ) Από δέκα τέτοιους χρόνους μπορούμε να συμπεράνουμε με βεβαιότητα ότι καμία εξυπηρέτηση δεν διαρκεί πάνω από 10 λεπτά;

(δ) Αν στη δεύτερη υπηρεσία οι δυνατές τιμές είναι 1 και $b > 10$ με πιθανότητες $1-\varepsilon$ και ε, επιλέξτε $\varepsilon > 0$ ώστε η πιθανότητα δέκα μονάδων να υπερβαίνει το 0.95, όσο μεγάλο και αν είναι το b.

Άσκηση 10. Πεντάλεπτα χωρίς πελάτες

Προαιρετική

Στο ιδεατό μοντέλο, τα πλήθη πελατών σε n πεντάλεπτα είναι ανεξάρτητα και ισόνομα:

$$X_i \sim \text{Poisson}(\lambda), \qquad \lambda > 0.$$

Μας ενδιαφέρει το $p_0 = \mathbb{P}_\lambda(X_1 = 0) = e^{-\lambda}$.

(α) Γράψτε την εκτιμήτρια A_n από τη σχετική συχνότητα των μηδενικών παρατηρήσεων και την B_n που αντικαθιστά το λ με τον $\overline{X}$ στον τύπο $e^{-\lambda}$.

(β) Για $x = (0, 1, 0, 2, 1, 2)$, υπολογίστε τις δύο εκτιμήσεις. Συμπίπτουν;

(γ) Βρείτε τις $\mathbb{E}_\lambda(A_n)$ και $\mathbb{V}_\lambda(A_n)$.

(δ) Βρείτε τις $\mathbb{E}_\lambda(B_n)$ και $\mathbb{V}_\lambda(B_n)$ με την ταυτότητα

$$\mathbb{E}(e^{-tK}) = \exp\big(a(e^{-t} - 1)\big), \qquad K \sim \text{Poisson}(a), \quad t \geq 0.$$

Χρησιμοποιήστε $K = \sum_i X_i \sim \text{Poisson}(n\lambda)$. Ισχύει $\mathbb{E}_\lambda(B_n) = e^{-\lambda}$; Για τη σύγκριση μπορείτε να χρησιμοποιήσετε $e^{-u} > 1 - u$ για $u > 0$.