A

Πιθανότητες και Στοιχεία Στατιστικής

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

14 Φεβρουαρίου 2013

Θέμα 1. [20 Βαθμοί] Ρίχνουμε ένα αμερόληπτο ζάρι $k \geq 2$ φορές. Ποιά η πιθανότητα τουλάχιστον δύο από τα k αποτελέσματα του ζαριού να είναι ίδια;

Θέμα 2. [30 Βαθμοί] Μιά κάλπη A περιέχει 10 μαύρα σφαιρίδια, ενώ μιά κάλπη B περιέχει 7 άσπρα και 1 μαύρο σφαιρίδιο. Επιλέγουμε μία κάλπη στην τύχη, και εξάγουμε από αυτή την κάλπη με επανάθεση 3 σφαιρίδια το ένα μετά το άλλο. Κάθε φορά, καταγράφουμε το χρώμα του σφαιριδίου που εξάγουμε.

(α) Ποιά είναι η πιθανότητα και τα τρία σφαιρίδια που εξάγουμε να είναι μαύρα;

(β) Από μία πραγματοποίηση του πειράματος, μας φανερώνεται μόνο ότι και τα τρία σφαιρίδια που βγήκαν ήταν μαύρα. Δεδομένου αυτού, ποιά είναι η πιθανότητα η κάλπη που είχε επιλεγεί αρχικά να είναι η B; Χωρίς να υπολογίσουμε αυτή την πιθανότητα, διαισθητικά περιμένουμε να είναι μικρή ή μεγάλη;

Θέμα 3. [25 Βαθμοί] Δίνονται δύο λυχνίες A, B των οποίων ο χρόνος ζωής είναι μιά συνεχής τυχαία μεταβλητή X για την A και Y για την B, με πυκνότητες αντίστοιχα

$$f_X(x) = \begin{cases} e^{-x} & \text{αν } x > 0, \\ 0 & \text{αν } x \leq 0, \end{cases} \qquad f_Y(x) = \begin{cases} xe^{-x} & \text{αν } x > 0, \\ 0 & \text{αν } x \leq 0. \end{cases}$$

(α) Να υπολογιστούν οι μέσες τιμές $E(X), E(Y)$.

(β) Να υπολογιστούν οι πιθανότητες $P(X > t), P(Y > t)$ για $t > 0$.

(γ) Ποιά λυχνία είναι προτιμότερο να χρησιμοποιήσουμε και γιατί;

Θέμα 4. [30 Βαθμοί] Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα από πληθυσμό με πυκνότητα

$$f(x) = \begin{cases} \frac{1}{r}\, e^{-x/r} & \text{αν } x > 0, \\ 0 & \text{αν } x \leq 0, \end{cases}$$

όπου r είναι μια θετική άγνωστη παράμετρος.
(α) Ποιά η πυκνότητα της τυχαίας μεταβλητής $R_n := \min\{X_1, X_2, \ldots, X_n\}$;
(β) Να δειχθεί ότι η $T_n := nR_n/r$ είναι οδηγός ποσότητα (δηλαδή η κατανομή της δεν εξαρτάται από την r), και με βάση αυτήν να κατασκευαστεί διάστημα εμπιστοσύνης για την r με βαθμό εμπιστοσύνης $1 - a = 0.95$

Υπόδειξη: Για το **(α)**, υπολογίζουμε την συνάρτηση κατανομής της R_n ως εξής

$$F_{R_n}(t) = P(R_n \leq t) = 1 - P(R_n > t) = 1 - P(X_1 > t, X_2 > t, \ldots, X_n > t) = ...$$

Άριστα είναι το 100. Διάρκεια 2 ώρες.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!

Απαντήσεις

1. (Αυτό είναι το ίδιο με το πρόβλημα των γενεθλίων) Έστω το ενδεχόμενο

$$A := \{\text{υπάρχουν τουλάχιστον δύο ίδιες ενδείξεις από τις } k\}.$$

Τότε

$$P(A) = 1 - P(A^c) = 1 - \frac{(6)_k}{6^k}.$$

Αυτό γιατί για τη διατεταγμένη k-άδα των αποτελεσμάτων υπάρχουν 6^k ισοπίθανες επιλογές. Από αυτές, ευνοϊκές είναι αυτές χωρίς επαναλήψεις, άρα είναι διατάξεις των 6 ανά k.

Ο τύπος βέβαια ισχύει και για τις τετριμένες περιπτώσεις όπου $k \geq 7$. Δεν χρειάζεται να τις εξετάσουμε χωριστά.

2. (α) Θεώρημα ολικής πιθανότητας.
(β) Τύπος Bayes.

4. (α) Ακολουθώντας την υπόδειξη, υπολογίζουμε για $t \in \mathbb{R}$ τη συνάρτηση κατανομής της R_n.

$$\begin{aligned} F_{R_n}(t) &= P(R_n \leq t) = 1 - P(R_n > t) = 1 - P(X_1 > t, X_2 > t, \dots, X_n > t) \\ &= 1 - P(X_1 > t)P(X_2 > t) \cdots P(X_n > t) = 1 - P(X_1 > t)^n. \end{aligned}$$

Χρησιμοποιήσαμε την ανεξαρτησία των $X_1, X_2, \dots, X_n$. Τώρα εύκολα βρίσκουμε ότι

$$P(X_1 > t) = \int_t^\infty f(x)\,dx = \begin{cases} 1 & \text{αν } t \leq 0, \\ e^{-t/r} & \text{αν } t > 0. \end{cases}$$

Άρα

$$F_{R_n}(t) = \begin{cases} 0 & \text{αν } t \leq 0, \\ e^{-nt/r} & \text{αν } t > 0. \end{cases}$$

Επομένως παραγωγίζοντας βρίσκουμε ότι η πυκνότητα της R_n είναι $f_{R_n}(t) = (n/r)e^{-nt/r}\mathbf{1}_{t>0}$. Δηλαδή η R_n ακολουθεί την εκθετική κατανομή με παράμετρο n/r.

(β) Η συνάρτηση κατανομής της T_n ισούται με

$$F_{T_n}(t) = P(T_n \leq t) = P(R_n \leq tr/n) = F_{R_n}(rt/n),$$

και παραγωγίζοντας βρίσκουμε την πυκνότητα της ως

$$f_{T_n}(t) = f_{R_n}\left(\frac{rt}{n}\right) \cdot \frac{r}{n} = e^{-t}\mathbf{1}_{t>0}.$$

Δηλαδή η T_n ακολουθεί την εκθετική κατανομή με παράμετρο 1. Επομένως είναι οδηγός ποσότητα.

Βρίσκουμε σταθερές $0 < c < d$ ώστε $P(T_n < c) = P(T_n > d) = a/2$. Συγκεκριμένα[1] , με βάση τα παραπάνω, αυτές οι δύο σχέσεις ισοδυναμούν με

$$1 - e^{-c} = e^{-d} = a/2,$$

και άρα $c = -\log(1 - (a/2)), d = -\log(a/2)$. Τότε

$$a = P(c < T_n < d) = P\left(c < \frac{n}{r}R_n < d\right) = P\left(\frac{n}{d}R_n < r < \frac{n}{c}R_n\right).$$

Οι $nR_n/c, nR_n/d$ είναι στατιστικές συναρτήσεις. Επομένως η τελευταία σχέση δίνει ότι ένα διάστημα εμπιστοσύνης για την r με βαθμό εμπιστοσύνης $1 - a$ είναι το

$$\left[\frac{n}{d}R_n, \frac{n}{c}R_n\right]$$

[1] Ο ακριβής υπολογισμός τους δεν ήταν απαραίτητος για την εξέταση. Είναι σαφές ότι υπάρχουν, και δεν εξαρτώνται από την r αφού η T_n είναι συνεχής τυχαία μεταβλητή της οποίας η κατανομή δεν εξαρτάται από την r.