Πιθανότητες και Στοιχεία Στατιστικής A

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

1 Μαρτίου 2012

Θέμα 1. [15 Βαθμοί] Σε μιά πόλη $2n$ κατοίκων $a_1, a_2, \ldots, a_{2n}$, οι $a_1, a_2, \ldots, a_n$ είναι άνδρες (έχουν πλήθος n) και οι υπόλοιποι n είναι γυναίκες. Ο a_1 επιλέγει τυχαία μία γυναίκα από τις n και της λέει μια φημολογία. Έπειτα εκείνη επιλέγει τυχαία έναν άνδρα από τους n και κάνει το ίδιο, και η διαδικασία συνεχίζεται όμοια. Δηλαδή κάθε άτομο επιλέγει στην τύχη ένα άτομο του αντίθετου φύλου και του λέει την φημολογία. Να βρεθεί η πιθανότητα η φημολογία να ειπωθεί $2k+1$ φορές χωρίς να ακουστεί ξανά από κάποιο άτομο που την έχει μεταφέρει σε κάποιο προηγούμενο βήμα. Υποθέτουμε ότι $k \in \{0, 1, \ldots, n-1\}$.

Θέμα 2. [25 Βαθμοί] Μία κάλπη περιέχει 13 σφαιρίδια αριθμημένα $1, 2, \ldots, 13$. Επιλέγουμε ένα στην τύχη, σημειώνουμε το αριθμό του, και το επιστρέφουμε στην κάλπη. Αν ο αριθμός του ήταν i, τότε παίρνουμε από την κάλπη i σφαιρίδια το ένα μετά το άλλο (χωρίς επανάθεση). Ποιά είναι η πιθανότητα το σφαιρίδιο που επιλέξαμε στην αρχή του πειράματος να μήν είναι ανάμεσα σε αυτά που επιλέξαμε στο τέλος;

Θέμα 3. [25 Βαθμοί] **(α)** Ρίχνουμε δύο αμερόληπτα ζάρια συνεχώς μέχρι να εμφανιστεί το αποτέλεσμα (6, 6). Έστω X ο (τυχαίος) αριθμός ρίψεων που κάνουμε ώσπου να συμβεί αυτό. Ποιά η συνάρτηση πιθανότητας της τυχαίας μεταβλητής X και ποιά η μέση τιμή της;

(β) Στο ίδιο πείραμα όπως στο ερώτημα (α) σταματάμε τις ρίψεις όταν έχουμε δει όλες τις μεγάλες διπλές ζαριές, δηλαδή τις $(4,4), (5,5), (6,6)$. Έστω Y ο (τυχαίος) αριθμός ρίψεων που κάνουμε ώσπου να συμβεί αυτό. Ποιά η μέση τιμή της τυχαίας μεταβλητής Y;

Θέμα 4. [20 Βαθμοί] Έστω (X, Y) διδιάστατη συνεχής τυχαία μεταβλητή με πυκνότητα

$$f_{X,Y}(x, y) = \begin{cases} \frac{2}{3} & \text{αν } x \in (1, 2) \text{ και } 0 < y < x, \\ 0 & \text{διαφορετικά.} \end{cases}$$

(α) Ποιά η περιθώρια συνάρτηση πυκνότητας f_X της X;

(β) Ποιά η πιθανότητα $P(X + Y < 2)$;

Θέμα 5. [35 Βαθμοί] Έστω συνεχής τυχαία μεταβλητή X με πυκνότητα

$$f_X(x) = \begin{cases} \frac{1}{r} x^{\frac{1}{r}-1} & \text{αν } x \in (0, 1), \\ 0 & \text{διαφορετικά,} \end{cases}$$

όπου r είναι μιά θετική σταθερά.

(α) Ποιά είναι η συνάρτηση πυκνότητας f_Y της τυχαίας μεταβλητής $Y = -\log X$; Για ποιά $y \in \mathbb{R}$ είναι η $f_Y(y)$ θετική;

(β) Έστω $X_1, X_2, \ldots, X_n$ τυχαίο δείγμα από πληθυσμό με πυκνότητα f_X όπως στην εκφώνηση. Να βρεθεί η εκτιμήτρια μέγιστης πιθανοφάνειας της παραμέτρου r.

(γ) Είναι η εκτιμήτρια του r που προσδιορίστηκε στο (β) αμερόληπτη;

Άριστα είναι το 100. Διάρκεια $2\frac{1}{2}$ ώρες.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!

Απαντήσεις

1. Το πλήθος των δυνατών περιπτώσεων είναι n^{2k+1}, γιατί σε κάθε μετάδοση της φημολογίας, αυτός που την μεταδίδει έχει n επιλογές. Δηλαδή όλα τα άτομα του αντίθετου φύλου.

Σχετικά με τις ευνοϊκές περιπτώσεις, εφαρμόζουμε την πολλαπλασιαστική αρχή. Η φημολογία θα μεταδοθεί από $k+1$ άνδρες και από k γυναίκες. Ο πρώτος άνδρας $A_1 = a_1$ έχει k επιλογές, η γυναίκα η οποία αυτός επιλέγει (έστω Γ_1) έχει $k-1$ επιλογές (δεν μπορεί να επιλέξει τον A_1), ο άνδρας που επιλέγει η Γ_1, έστω A_2, έχει ομοίως $k-1$ επιλογές. Η k γυναίκα σε αυτή την αλυσίδα έχει $n-k$ επιλογές, και ο $k+1$ άνδρας έχει $n-k$ επιλογές (αφού k γυναίκες έχουν μεταδώσει την φημολογία).

Άρα η ζητούμενη πιθανότητα είναι

$$\frac{n(n-1)(n-1)(n-2)(n-2)\dots(n-k)(n-k)}{n^{2k+1}} = \frac{((n-1)_k)^2}{n^{2k}}.$$

2. Έχουμε πείραμα σε δύο στάδια. Πάμε ως συνήθως με θεώρημα ολικής πιθανότητας. Για $i = 1, 2, ..., 13$ έστω $A_i := \{\text{αρχικά επιλέγουμε το σφαιρίδιο } i\}$, και

$$B := \{\text{το αρχικό σφαιρίδιο δεν περιέχεται σε αυτά που επιλέγουμε έπειτα}\}.$$

Τότε

$$\begin{aligned} P(B) &= \sum_{i=1}^{13} P(B \cap A_i) = \sum_{i=1}^{6} P(A_i)P(B \mid A_i) = \sum_{i=1}^{13} \frac{1}{13}\frac{(12)_i}{(13)_i} \\ &= \sum_{i=1}^{13} \frac{1}{13}\frac{13-i}{13} = \frac{1}{13^2}\sum_{k=1}^{12} k = \frac{1}{13^2}\frac{12 \times 13}{2} = \frac{6}{13}. \end{aligned}$$

Η $P(B \cap A_i) = (12)_i/(13)_i$ ισχύει γιατί επιλέγουμε διαδοχικά χωρίς επανάθεση i σφαιρίδια, άρα έχουμε διάταξη χωρίς επανάληψη, και στα ευνοϊκά σενάρια απαγορεύεται να πάρουμε το i σφαιρίδιο. Γιάυτό και το 12 στον αριθμητή.

3. (α) Κατά τα γνωστά η X ακολουθεί την Γεωμετρική κατανομή με παράμετρο $p = 1/36$. Άρα έχει συνάρτηση πιθανότητας

$$f_X(k) = P(X = k) = \begin{cases} (1-p)^{k-1}p & \text{για } k \in \mathbb{N}, k \geq 1, \\ 0 & \text{για } k \in (\mathbb{R} \setminus \mathbb{N}) \cup \{0\}. \end{cases}$$

Από θεωρία, ξέρουμε ότι $E(X) = 1/p = 36$.

(β) Ίδιο σκεπτικό όπως στο πρόβλημα του συλλέκτη κουπονιών. Έστω X_1 ο αριθμός των δοκιμών που απαιτείται ώσπου να δούμε μία από τις ενδείξεις $(4,4), (5,5), (6,6)$, και έστω ότι αυτή είναι (a,a). Έπειτα έστω X_2 ο αριθμός των επιπλέον δοκιμών που απαιτείται ώσπου να δούμε μία από τις ενδείξεις $(4,4), (5,5), (6,6)$ που να είναι όμως διαφορετική από την (a,a). Έστω ότι αυτή είναι η (b,b). Τέλος, έστω X_3 ο αριθμός των επιπλέον δοκιμών που απαιτείται ώσπου να δούμε την ένδειξη από τις $(4,4), (5,5), (6,6)$ που είναι διαφορετική από τις $(a,a), (b,b)$. Οι X_1, X_2, X_3 ακολουθούν την γεωμετρική κατανομή με παραμέτρους $p_1 = 3/36, p_2 = 2/36, p_3 = 1/36$. Για τον αριθμό στην εκφώνηση έχουμε $Y = X_1 + X_2 + X_3$. Η γραμμικότητα της μέσης τιμής δίνει

$$E(Y) = EX_1 + EX_2 + EX_3 = \frac{1}{p_1} + \frac{1}{p_3} + \frac{1}{p_3} = \frac{36}{3} + \frac{36}{2} + \frac{36}{1} = 66.$$

4. Σε αυτή την άσκηση επιβάλλεται να κάνουμε ένα σχήμα του χωρίου στο $\mathbb{R}^2$ που η $f_{X,Y}$ είναι διαφορετική από το 0.

(α)

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \begin{cases} \frac{2}{3}x & \text{αν } x \in (1,2) \\ 0 & \text{αν } x \in \mathbb{R} \setminus (1,2). \end{cases}$$

Πράγματι, για $x \in (1,2)$,

$$\int_{\mathbb{R}} f_{X,Y}(x,y)\,dy = \int_{\mathbb{R}} \frac{2}{3}\mathbf{1}_{0<y<x}\,dy = \int_0^x \frac{2}{3}\,dy = \frac{2}{3}x.$$

Ενώ για $x \in \mathbb{R} \setminus (1,2)$, έχουμε $f_{X,Y}(x,y) = 0$ για κάθε $y \in \mathbb{R}$.

(β)

$$P(X+Y<2) = \int_1^2 \int_0^{2-x} \frac{2}{3}\,dy\,dx = \frac{2}{3}\int_1^2 (2-x)\,dx = \frac{1}{3}$$

5. (α) Βρίσκουμε πρώτα την συνάρτηση κατανομής της Y. Για $y \in \mathbb{R}$, έχουμε

$$\begin{aligned} F_Y(y) &= P(-\log X \le y) = P(\log X \ge -y) = P(X \ge e^{-y}) = 1 - P(X < e^{-y}) \\ &= 1 - F_X(e^{-y}). \end{aligned}$$

Η $P(X < e^{-y}) = P(X \le e^{-y}) (= F_X(e^{-y}))$ ισχύει γιατί η X είναι συνεχής τυχαία μεταβλητή. Παραγωγίζοντας την πιο πάνω ισότητα, παίρνουμε $f_Y(y) = F_Y'(y) = -f_X(e^{-y})(-e^{-y}) = f_X(e^{-y})e^{-y}$. Από την μορφή της f_X, προκύπτει ότι η τελευταία ποσότητα είναι θετική ακριβώς όταν $e^{-y} \in (0,1)$, δηλαδή $y > 0$, και τότε ισούται με

$$\frac{1}{r}(e^{-y})^{\frac{1}{r}-1}e^{-y} = \frac{1}{r}e^{-\frac{1}{r}y}.$$

Άρα

$$f_Y(y) = \frac{1}{r}e^{-\frac{1}{r}y}\mathbf{1}_{y>0}.$$

Δηλαδή η Y ακολουθεί την εκθετική με παράμετρο $1/r$. $f_Y(y) > 0$ μόνο για $y > 0$.

(β) Για $x_1, x_2, \ldots, x_n \in (0,1)$, η συνάρτηση πιθανοφάνειας του δείγματος είναι

$$L(r) := f_X(x_1)f_X(x_2)\cdots f_X(x_n) = \left(\frac{1}{r}\right)^n (x_1x_2\cdots x_n)^{\frac{1}{r}-1},$$

και ο λογάριθμός της είναι

$$\ell(r) := \log L(r) = \left(\frac{1}{r} - 1\right)\log(x_1x_2\cdots x_n) - n\log r.$$

Η ℓ είναι παραγωγίσιμη συνάρτηση του r στο $(0,\infty)$, με παράγωγο

$$\ell'(r) = -\frac{1}{r^2}\log(x_1x_2\cdots x_n) - \frac{n}{r} = -\frac{n}{r^2}\left(r + \frac{1}{n}\log(x_1x_2\cdots x_n)\right) = -\frac{n}{r^2}(r-s)$$

με $s := -(1/n)\log(x_1x_2\cdots x_n)$ το οποίο είναι > 0 αφού $x_1, x_2, \ldots, x_n \in (0,1)$. Άρα η $\ell(r)$ είναι αύξουσα στο $(0,s]$ και φθίνουσα στο $[s,\infty)$. Παρουσιάζει ολικό μέγιστο το s. Άρα η εκτιμήτρια μέγιστης πιθανοφάνειας του r είναι η

$$T(X_1,\ldots,X_n) := \frac{-\log(X_1X_2\cdots X_n)}{n} = \frac{-\log X_1 - \log X_2 - \cdots - \log X_n}{n}.$$

(γ) Η $Y = -\log X$ από το (α) έχει μέση τιμή $1/(1/r) = r$ (αυτό από τον πίνακα των κατανομών ή το υπολογίζουμε με βάση τον ορισμό, δηλαδή $E(-\log X) = -\int_0^1 (\log x) r^{-1} x^{\frac{1}{r}-1}\,dx = \ldots$). Χρησιμοποιώντας την γραμμικότητα της μέσης τιμής και το ότι οι X_i έχουν την ίδια κατανομή, βρίσκουμε

$$E\big(T(X_1,\ldots,X_n)\big) = \frac{nE(-\log X_1)}{n} = \frac{nr}{n} = r.$$

Άρα η $T(X_1,\ldots,X_n)$ είναι αμερόληπτη εκτιμήτρια του r.