Πιθανότητες και Στοιχεία Στατιστικής

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

10 Φεβρουαρίου 2011

1.[10 Βαθμοί] Απο μία κάλπη που περιέχει 50 σφαίρες αριθμημένες $10, 11, 12, \ldots, 59$ επιλέγουμε έξι σφαίρες την μία μετά την άλλη με επανάθεση. Ποιά είναι η πιθανότητα και οι έξι σφαίρες να έχουν το ίδιο ψηφίο δεκάδας;

2.[10 Βαθμοί] Ρίχνουμε μία φορά ένα αμερόληπτο ζάρι, και αν έρθει η ένδειξη $i \in \{1, 2, \ldots, 6\}$, το ρίχνουμε πάλι i φορές. Ποιά είναι η πιθανότητα οι ενδείξεις που θα πάρουμε στο δεύτερο στάδιο του πειράματος να είναι όλες διαφορετικές;

3.[30 Βαθμοί] Έστω X συνεχής τυχαία μεταβλητή με πυκνότητα

$$f(x) := \begin{cases} 2x & \text{αν } x \in (0,1), \\ 0 & \text{αν } x \in \mathbb{R} \setminus (0,1). \end{cases}$$

(α) Να υπολογιστούν οι $E(X), V(X)$.

(β) Να υπολογιστούν οι $P(X = 0.7), P(X > 0.5), E(\sin X)$.

(γ) Έστω ότι έχουμε 1800 ανεξάρτητες τυχαίες μεταβλητές, καθεμία με την παραπάνω πυκνότητα. Ποιά είναι προσεγγιστικά η πιθανότητα το άθροισμά τους να είναι τουλάχιστον 1180;

4.[10 Βαθμοί] Έστω X, Y, Z τυχαίες μεταβλητές (ορισμένες στον ίδιο χώρο πιθανότητας) με $V(X) = V(Y) = 2$, $\mathrm{Cov}(X, Y) = 1$, και την Z ανεξάρτητη από τις X, Y και να ακολουθεί την τυπική κανονική κατανομή $N(0, 1)$. Να υπολογισθεί η διασπορά $V(X - Y - Z)$.

5.[20 Βαθμοί] Έστω (X, Y) διδιάστατη συνεχής τυχαία μεταβλητή με πυκνότητα

$$f(x,y) := \begin{cases} 2 & \text{αν } 0 < y < x < 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

(α) Να υπολογιστούν οι περιθώριες πυκνότητες f_X, f_Y των X, Y.

(β) Να υπολογιστεί η πιθανότητα $P(X > 2Y)$.

6.[15 Βαθμοί] Έστω δείγμα $X_1, X_2, \ldots, X_n$ από πληθυσμό που ακολουθεί την ομοιόμορφη κατανομή στο $[0, r]$. Το $r \in (0, \infty)$ είναι άγνωστη παράμετρος.

(α) Ποιά απο τις

$$\frac{3X_1^2}{r}, 2X_1$$

είναι αμερόληπτη εκτιμήτρια του r;

(β) Για ποιά σταθερά c είναι η

$$c\,\frac{X_1 + X_2 + \cdots + X_n}{n}$$

αμερόληπτη εκτιμήτρια του r;

7.[25 Βαθμοί] Έστω δείγμα $X_1, X_2, \ldots, X_n$ από πληθυσμό που ακολουθεί την ομοιόμορφη κατανομή στο $[r, r + 2]$. Το $r \in \mathbb{R}$ είναι άγνωστη παράμετρος, και το n είναι μεγάλο.

(α) Ποιά είναι η μέση τιμή και ποιά η διασπορά της X_1;

(β) Να βρεθεί προσεγγιστικά διάστημα εμπιστοσύνης για το r με συντελεστή εμπιστοσύνης $1 - a = 0.95$.

Τιμές από τον Πίνακα της Τυποποιημένης Κανονικής, $N(0, 1)$**:**

$$\begin{array}{lll} \Phi(0.5) = 0.6915, & \Phi(1) = 0.8413, & \Phi(1.5) = 0.9332, \\ \Phi(2) = 0.9773, & \Phi(2.5) = 0.9938, & \Phi(3) = 0.9987, \\ \Phi(1.65) = 0.95, & \Phi(1.96) = 0.975 & \end{array}$$

Άριστα είναι το 100. Διάρκεια $2\frac{1}{2}$ ώρες. **ΚΑΛΗ ΕΠΙΤΥΧΙΑ!**

1.

$$P(\text{όλες το ίδιο ψηφίο δεκάδας}) = \sum_{i=1}^{5} P(\text{όλες ψηφίο δεκάδας } i) = 5\frac{10^6}{50^6} = 5\frac{1}{5^6} = \frac{1}{5^5}$$

2. Για $i = 1, 2, ..., 6$ έστω $A_i := \{\text{η πρώτη ρίψη φέρνει } i\}$, και

$$B := \{\text{όλες οι ενδείξεις της δεύτερης φάσης είναι διαφορετικές}\}.$$

Τότε

$$P(B) = \sum_{i=1}^{6} P(B \cap A_i) = \sum_{i=1}^{6} P(A_i)P(B \mid A_i) = \sum_{i=1}^{6} \frac{1}{6}\frac{(6)_i}{6^i}.$$

3. (α) $E(X) = \int_0^1 2x^2\,dx = 2/3$, $E(X^2) = \int_0^1 2x^3\,dx = 1/2$, $V(X) = 1/2 - 4/9 = 1/18$.

(β) Η είναι συνεχής τυχαία μεταβλητή (η συνάρτηση κατανομής της, έστω F, είναι συνεχής), άρα $P(X = 0.7) = F(0.7) - F(0.7-) = 0$. Έπειτα,

$$P(X > 0.5) = \int_{1/2}^{1} 2x\,dx = 1 - (1/4) = 3/4,$$
$$E(\sin X) = \int_0^1 2x \sin x\,dx = ... = 2(\sin 1 - \cos 1).$$

Παραγοντική ολοκλήρωση στο τελευταίο.

(γ) Έστω $S_{1800} = X_1 + \cdots + X_{1800}$. Από το ερώτημα (α) και το κεντρικό οριακό θεώρημα, έχουμε ότι η τυχαία μεταβλητή

$$\frac{S_{1800} - 1800(2/3)}{\sqrt{1800/18}} = \frac{S_{1800} - 1200}{10}$$

ακολουθεί προσεγγιστικά την τυπική κανονική κατανομή $N(0, 1)$. Άρα

$$P(S_{1800} \geq 1180) = P\left(\frac{S_{1800} - 1200}{10} \geq -2\right) \approx 1 - \Phi(-2) = \Phi(2) = 0.9773$$

4. Από τον γνωστό τύπο για διασπορά αθροίσματος

$$\begin{aligned} V(X - Y - Z) &= V(X) + V(Y) + V(Z) - 2\mathrm{Cov}(X,Y) - 2\mathrm{Cov}(X,Z) + 2\mathrm{Cov}(Y,Z) \\ &= 2 + 2 + 1 - 2 - 0 - 0 = 3. \end{aligned}$$

Η συνδιακύμανση ανεξάρτητων τυχαίων μεταβλητών είναι 0.

5. (α) Η $f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy$ ισούται με 0 για $x \notin (0,1)$, ενώ για $x \in (0,1)$ έχουμε

$$f_X(x) = \int_{\mathbb{R}} 2 \times \mathbf{1}_{0<y<x<1}\,dy = \int_0^x 2\,dy = 2x.$$

Όμοια, $f_Y(y) = 0$ για $y \notin (0,1)$, ενώ για $y \in (0,1)$ έχουμε

$$f_Y(y) = \int_{\mathbb{R}} 2 \times \mathbf{1}_{0<y<x<1}\,dx = \int_y^1 2\,dy = 2(1-y).$$

(β) [Εδώ ένα σχήμα βοηθάει.]

$$P(X > 2Y) = \iint_{\mathbb{R}^2} f(x,y)\mathbf{1}_{x>2y}\,dxdy = \int_0^1 \int_0^{x/2} 2\,dydx = \int_0^1 x\,dx = \frac{1}{2}.$$

6. Και στα δύο ερωτήματα χρησιμοποιούμε ότι η μέση τιμή μιας τυχαίας μεταβλητής που ακολουθεί την ομοιόμορφη κατανομή στο $[0, r]$ είναι $r/2$ (το μέσο του διαστήματος).

(α) Η πρώτη δεν είναι αμερόληπτη εκτιμήτρια γιατί δεν είναι στατιστική συνάρτηση καθώς περιέχει την άγνωστη παράμετρο r (έχει βέβαια την σωστή μέση τιμή, δηλ. $3E(X_1^2/r) = ... = r$, αλλά αυτό δεν φτάνει). Η δεύτερη είναι στατιστική συνάρτηση, και έχει μέση τιμή $2E(X_1) = 2(r/2) = r$. Άρα είναι αμερόληπτη εκτιμήτρια του r.

(β) Χρησιμοποιώντας την γραμμικότητα της μέσης τιμής, βρίσκουμε ότι η μέση τιμή της δοσμένης τυχαίας μεταβλητής είναι

$$c\frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n} = cE(X_1) = cr/2.$$

Με την επιλογή $c=2$ έχουμε ότι η

$$2\frac{X_1+X_2+\cdots+X_n}{n}$$

είναι στατιστική συνάρτηση με μέση τιμή r. Άρα είναι αμερόληπτη εκτιμήτρια του r.

7. (α) Είναι γνωστό ότι για μία τυχαία μεταβλητή $X \sim U(a,b)$ με $a<b$ ισχύει

$$E(X) = \frac{a+b}{2},\ V(X) = \frac{(b-a)^2}{12}.$$

Άρα $E(X_1) = r+1, V(X_1) = 4/12 = 1/3$

(β) [Αυτό είναι παραλλαγή στο παράδειγμα 4.1, σελ. 138 στις σημειώσεις του κ. Χαραλαμπίδη. Άγνωστη μέση τιμή, γνωστή διασπορά.]

Έστω $S_n = X_1+\cdots+X_n$. Από το κεντρικό οριακό θεώρημα, η τυχαία μεταβλητή

$$\frac{S_n - n(r+1)}{\sqrt{n/3}}$$

ακολουθεί προσεγγιστικά την τυπική κανονική κατανομή $N(0,1)$. Βρίσκουμε το $z_{a/2} = z_{0.025}$ το οποίο είναι ο αριθμός που ικανοποιεί $P(Z>x) = 0.025$ όπου $Z \sim N(0,1)$, δηλαδή $\Phi(x) = 0.975$. Από τις τιμές της Φ που μας δίνονται, βρίσκουμε $z_{0.025} = 1.96$.
Από το κεντρικό οριακό θεώρημα, έχουμε

$$P\left(-z_{0.025} \le \frac{S_n - n(r+1)}{\sqrt{n/3}} \le z_{0.025}\right) \approx \Phi(z_{0.025}) - \Phi(-z_{0.025}) = 2\Phi(z_{0.025}) - 1 = 0.95$$

Η διπλή ανισότητα μέσα στην πιθανότητα ισοδυναμεί με

$$\frac{S_n}{n} - z_{0.025}\frac{1}{\sqrt{3n}} - 1 < r < \frac{S_n}{n} + z_{0.025}\frac{1}{\sqrt{3n}} - 1.$$

Άρα ένα διάστημα εμπιστοσύνης για το r με συντελεστή εμπιστοσύνης $1-a = 0.95$ είναι το

$$I = \left[\frac{S_n}{n} - z_{0.025}\frac{1}{\sqrt{3n}} - 1,\ \frac{S_n}{n} + z_{0.025}\frac{1}{\sqrt{3n}} - 1\right].$$