Πιθανότητες και Στοιχεία Στατιστικής
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
17 Οκτωβρίου 2011

1.[10 Βαθμοί] Ρίχνουμε ένα αμερόληπτο ζάρι 10 φορές διαδοχικά. Ποιά είναι η πιθανότητα κάθε ένδειξη να είναι διαφορετική από τις δύο προηγούμενές της;

2.[20 Βαθμοί] Μία κάλπη περιέχει τρία νομίσματα N_1, N_2, N_3. Η πιθανότητα να φέρει το καθένα την ένδειξη ''Κεφαλή'' είναι αντίστοιχα $p_1 = 1/100, p_2 = 4/5, p_3 = 4/5$. Επιλέγουμε στην τύχη ένα από τα τρία νομίσματα (πιθανότητα 1/3 για καθένα νόμισμα) και το ρίχνουμε.

(α) Ποιά είναι η πιθανότητα η ρίψη να φέρει ''Κεφαλή'';

(β) Αν η ρίψη έφερε ''Γράμματα'' ποιά είναι η πιθανότητα να έχουμε επιλέξει το N_1;

3.[25 Βαθμοί] Έστω X διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας

$$f(x) := P(X = x) = \begin{cases} 1/6 & \text{αν } x = -1, \\ 1/3 & \text{αν } x = 0, \\ 1/2 & \text{αν } x = 1, \\ 0 & \text{αν } x \in \mathbb{R} \setminus \{-1, 0, 1\}. \end{cases}$$

(α) Να υπολογιστεί η πιθανότητα $P(X \geq 0)$ και οι $E(X), V(X)$.

(β) Έστω ότι έχουμε 720 ανεξάρτητες τυχαίες μεταβλητές, καθεμία με την παραπάνω συνάρτηση πιθανότητας. Ποιά είναι προσεγγιστικά η πιθανότητα το άθροισμά τους να βρίσκεται στο διάστημα [210, 260];

4. [15 Βαθμοί] Δίνονται τυχαίες μεταβλητές X, Y, Z (στον ίδιο χώρο πιθανότητας) με την Z ανεξάρτητη από τις X, Y. Υποθέτουμε ότι η Z ακολουθεί την τυπική κανονική κατανομή $N(0, 1)$ και ότι

$$V(X + Y + 2Z) = 7, \quad V(X) = V(Y) = 1.$$

Να υπολογιστεί η συνδιακύμανση $C(X, Y)$.

5.[25 Βαθμοί] Έστω (X, Y) διδιάστατη συνεχής τυχαία μεταβλητή με πυκνότητα

$$f(x, y) := \begin{cases} cx^2y & \text{αν } (x, y) \in (-1, 1) \times (0, 1), \\ 0 & \text{διαφορετικά,} \end{cases}$$

όπου c είναι ένας πραγματικός αριθμός.

(α) Να υπολογιστεί ο c.

(β) Είναι οι X, Y ανεξάρτητες;

(γ) Να υπολογιστεί η πιθανότητα $P(Y < X)$.

6.[25 Βαθμοί] Έστω δείγμα $X_1, X_2, \ldots, X_n$ από πληθυσμό που ακολουθεί την κατανομή με πυκνότητα

$$f(x) = \begin{cases} \frac{2(r-x)}{r^2} & \text{αν } x \in (0, r), \\ 0 & \text{διαφορετικά .} \end{cases}$$

Το $r > 0$ είναι άγνωστη παράμετρος.

(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 $1 - a = 0.95$. Υποθέτουμε ότι το n είναι αρκετά μεγάλο, π.χ. $n \geq 10^4$.

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

$\Phi(0.5) = 0.6915$, $\Phi(1) = 0.8413$, $\Phi(1.5) = 0.9332$,
$\Phi(2) = 0.9773$, $\Phi(2.5) = 0.9938$, $\Phi(3) = 0.9987$,
$\Phi(1.65) = 0.95$, $\Phi(1.96) = 0.975$

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

1. Επειδή το ζάρι είναι αμερόληπτο, μπορούμε να χρησιμοποιήσουμε τον κλασικό ορισμό της πιθανότητας, και βρίσκουμε ότι η ζητούμενη πιθανότητα είναι

$$\frac{6 \times 5 \times 4^8}{6^{10}} = \frac{5 \times 2^7}{3^9}.$$

Ο αριθμητής μετράει τις ευνοϊκές περιπτώσεις γιατί η πρώτη ζαριά δεν έχει κανένα περιορισμό, η δεύτερη έχει 5 επιλογές καθώς απαγορεύεται να συμπέσει με την πρώτη. Έπειτα, καθεμία από τις επόμενες ζαριές έχει 4 επιλογές γιατί απαγορεύεται να συμπέσει με κάποια από τις δύο που προηγήθηκαν, και αυτές οι δύο έχουν διαφορετικά αποτελέσματα.

2. (α) Για $i = 1, 2, 3$ έστω $A_i := \{\text{επιλέγουμε το νόμισμα } N_i\}$, και

$$B := \{\text{η ρίψη φέρνει κεφαλή}\}.$$

Τότε

$$P(B) = \sum_{i=1}^{3} P(B \cap A_i) = \sum_{i=1}^{6} P(A_i) P(B \,|\, A_i) = \frac{1}{3} \times \frac{1}{100} + \frac{1}{3} \times \frac{4}{5} + \frac{1}{3} \times \frac{4}{5} = \frac{161}{300}.$$

(β) Ζητάμε την πιθανότητα

$$P(A_1 \,|\, B^c) = \frac{P(A_1 \cap B^c)}{P(B^c)} = \frac{P(B^c \,|\, A_1) P(A_1)}{P(B^c)} = \frac{\frac{99}{100} \times \frac{1}{3}}{\frac{139}{300}} = \frac{99}{139} \approx 0.71$$

Και βέβαια περιμέναμε αυτή την πιθανότητα να είναι μεγάλη αφού τα νομίσματα N_2, N_3 φέρνουν Γράμματα με μικρή πιθανότητα, $1/5$, ενώ το N_1 σχεδόν με σιγουριά, $99/100$. Δηλαδή το αποτέλεσμα Γράμματα είναι πιθανότερο να οφείλεται στην επιλογή του N_1 στο πρώτο στάδιο του πειράματος.

3. (α) $P(X \geq 0) = \sum_{x \geq 0} f(x) = f(0) + f(1) = 5/6$. Επίσης

$$E(X) = \sum_{x \in \mathbb{R}} x f(x) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3},$$

$$E(X^2) = \sum_{x \in \mathbb{R}} x^2 f(x) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3},$$

και άρα $V(X) = E(X^2) - E(X)^2 = 5/9$.

(β) Έστω $X_1, X_2, \ldots, X_{720}$ αυτές οι τυχαίες μεταβλητές και $S_{720} = X_1 + \cdots + X_{720}$ το άθροισμά τους. Από το ερώτημα (α) και το κεντρικό οριακό θεώρημα, έχουμε ότι η τυχαία μεταβλητή

$$\frac{S_{720} - 720(1/3)}{\sqrt{720 \times 5/9}} = \frac{S_{720} - 240}{20}$$

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

$$P(S_{720} \in [210, 260]) = P\left(\frac{S_{720} - 240}{20} \in [-1.5,\, 1]\right) \approx \Phi(1) - \Phi(-1.5) = \Phi(1) + \Phi(1.5) - 1.$$

4. Επειδή η $2Z$ είναι ανεξάρτητη από την $X + Y$, χρησιμοποιώντας τον γνωστό τύπο για διασπορά αθροίσματος βρίσκουμε

$$7 = V(X + Y + 2Z) = V(X + Y) + V(2Z) = V(X) + V(Y) + 2C(X, Y) + 4V(Z)$$

Άρα $C(X, Y) = 1/2$.

5. (α) Επειδή η f είναι πυκνότητα, έχουμε

$$1 = \iint_{\mathbb{R}^2} f(x, y) dx dy = c \int_0^1 \int_{-1}^1 x^2 y \, dx \, dy = c \frac{2}{3} \frac{1}{2}.$$

Άρα $c = 3$.

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

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = 3x^2 \int_0^1 y \, dy = 3x^2/2.$$

Άρα $f_X(x) = (3x^2/2)\,\mathbf{1}_{x\in(-1,1)}$. Όμοια βρίσκουμε $f_Y(y) = 2y\mathbf{1}_{y\in(0,1)}$. Παρατηρούμε ότι $f(x,y) = f_X(x)f_Y(y)$ για κάθε $(x,y)\in\mathbb{R}^2$. Με βάση γνωστή πρόταση, οι X, Y είναι ανεξάρτητες.

(γ) [Εδώ ένα σχήμα βοηθάει.]

$$P(Y<X) = \iint_{\mathbb{R}^2} f(x,y)\mathbf{1}_{y<x}\,dxdy = \int_0^1\int_0^x 3x^2y\,dydx = (3/2)\int_0^1 x^4\,dx = \frac{3}{10}.$$

6. (α) Έχουμε

$$E(X_1) = \int_{\mathbb{R}} xf(x)\,dx = \frac{2}{r^2}\int_0^r x(r-x)\,dx = \frac{2}{r^2}\left(\frac{r^3}{2}-\frac{r^3}{3}\right) = \frac{r}{3},$$
$$E(X_1^2) = \int_{\mathbb{R}} x^2f(x)\,dx = \frac{2}{r^2}\int_0^r x^2(r-x)\,dx = \frac{2}{r^2}\left(\frac{r^4}{3}-\frac{r^4}{4}\right) = \frac{r^2}{6},$$

και άρα $V(X_1) = E(X_1^2) - E(X_1)^2 = r^2/18$.

(β) Έστω $S_n = X_1 + \cdots + X_n$. Οι υπολογισμοί από το ερώτημα (α) και το κεντρικό οριακό θεώρημα δίνουν ότι η τυχαία μεταβλητή

$$\frac{S_n - n(r/3)}{\sqrt{nr^2/18}}$$

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

$$P\left(-z_{0.025} \le \frac{S_n - n(r/3)}{\sqrt{nr^2/18}} \le z_{0.025}\right) \approx \Phi(z_{0.025}) - \Phi(-z_{0.025}) = 2\Phi(z_{0.025}) - 1 = 0.95$$

Η διπλή ανισότητα μέσα στην πιθανότητα ισοδυναμεί με

$$-z_{0.025} \le \frac{S_n\sqrt{18}}{r\sqrt{n}} - \sqrt{2n} \le z_{0.025} \Leftrightarrow \sqrt{2n} - z_{0.025} \le \frac{S_n\sqrt{18}}{r\sqrt{n}} \le \sqrt{2n} + z_{0.025}$$
$$\Leftrightarrow \frac{S_n\sqrt{18}}{\sqrt{n}(\sqrt{2n}+z_{0.025})} \le r \le \frac{S_n\sqrt{18}}{\sqrt{n}(\sqrt{2n}-z_{0.025})}.$$

Για την τελευταία ισοδυναμία χρησιμοποιούμε το γεγονός ότι $\sqrt{2n} - z_{0.025} > 0$, που ισχύει γιατί $z_{0.025} = 1.96$ ενώ $\sqrt{n} \ge 100$ αφού $n \ge 10^4$. Άρα ένα προσεγγιστικό διάστημα εμπιστοσύνης για το r με συντελεστή εμπιστοσύνης $1-a = 0.95$ είναι το

$$I = \left[\frac{S_n\sqrt{18}}{\sqrt{n}(\sqrt{2n}+z_{0.025})}, \frac{S_n\sqrt{18}}{\sqrt{n}(\sqrt{2n}-z_{0.025})}\right]$$