Πιθανότητες Ι

Εξέταση 4 Φεβρουαρίου 2019

1. (20 Βαθμοί) Τοποθετούμε στην τύχη m (διακεκριμένα) σφαιρίδια σε n κουτιά $K_1, K_2, \dots, K_n$. Κάθε κουτί έχει απεριόριστη χωρητικότητα. Να υπολογιστούν:
(α) Η πιθανότητα το κουτί K_1 να παραμείνει άδειο.
(β) Η μέση τιμή A_n του πλήθους των κουτιών που μένουν άδεια. Αν θέλουμε

$$C_1 < \varliminf_{n\to\infty} A_n \le \varlimsup_{n\to\infty} A_n < C_2$$

για κάποιες πεπερασμένες και θετικές σταθερές C_1, C_2, ποια από τις εξής επιλογές για το m είναι κατάλληλη;

$$\text{(i) } m = n, \quad \text{(ii) } m = n[\log n], \quad \text{(iii) } m = n^2.$$

(γ) Η πιθανότητα να υπάρχει μόνο ένα άδειο κουτί όταν $m = n$.

2. (20 Βαθμοί) Το αποτέλεσμα ενός ιατρικού τέστ είναι είτε θετικό, που σημαίνει ότι ο εξεταζόμενος πάσχει από την ασθένεια, είτε αρνητικό, που σημαίνει ότι ο εξεταζόμενος δεν πάσχει από την ασθένεια. Όμως, το τεστ για μια συγκεκριμένη ασθένεια κάποιες φορές δίνει λάθος αποτελέσματα με αποτέλεσμα 1 στους 100 από αυτούς που δεν έχουν την ασθένεια να λαμβάνουν τεστ θετικό ενώ 2 στους 100 από αυτούς που έχουν την ασθένεια να λαμβάνουν τέστ αρνητικό. Γνωρίζουμε ότι περίπου 1 στους 1000 έχει την ασθένεια. Υπολογίστε τις πιθανότητες:
(α) Ο επόμενος που θα κάνει το τέστ να λάβει θετικό αποτέλεσμα.
(β) Ένας εξεταζόμενος που λαμβάνει θετικό τέστ να έχει πραγματικά την ασθένεια.
(γ) Στους επόμενους 5 με θετικό αποτέλεσμα στο τέστ να έχουμε τουλάχιστον 3 με την ασθένεια.

3. (20 Βαθμοί) Έστω (X, Y) διδιάστατη συνεχής τ.μ. με συνάρτηση πυκνότητας πιθανότητας

$$f_{X,Y}(x, y) = \begin{cases} \alpha(3x^2 + 2y) & \text{αν } (x, y) \in (0, 1) \times (0, 1), \\ 0 & \text{διαφορετικά.} \end{cases}$$

Να υπολογιστούν:
(α) Η σταθερά α.
(β) Οι περιθώριες συναρτήσεις πυκνότητας πιθανότητας f_X και f_Y των X και Y.
(γ) Η δεσμευμένη συνάρτηση πυκνότητας πιθανότητας $f_{X|Y}(x|y)$ της X δοθέντος $Y = y$. Για ποια $x, y \in \mathbb{R}$ ορίζεται;
(δ) Η πιθανότητα $P(X < Y)$.
(ε) Η μέση τιμή $E[3X - 2]$.

4. (15 Βαθμοί) Έστω X και Y ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο θ, δηλαδή έχουν συνάρτηση πυκνότητας πιθανότητας $f_X(x) = \theta e^{-\theta x}\mathbf{1}_{(0,\infty)}(x)$.
(α) Να υπολογιστεί η ροπογεννήτρια $M_X(t) = E(e^{tX})$ της X. Για ποιά $t \in \mathbb{R}$ είναι η $M_X(t)$ πεπερασμένη;
(β) Να προσδιοριστεί η κατανομή της $Z = X + Y$.

5. (20 Βαθμοί) Έστω $a > 0$ και X, Λ δύο τυχαίες μεταβλητές σε κοινό χώρο πιθανότητας ώστε η Λ να ακολουθεί την εκθετική κατανομή με παράμετρο a και δεδομένου ότι $\Lambda = \lambda$, η X ακολουθεί την κατανομή Poisson με παράμετρο λ. Δηλαδή $X|\{\Lambda = \lambda\} \sim \text{Poisson}(\lambda)$.
(α) Να υπολογιστεί η πιθανογεννήτρια, $\mathbf{E}(t^X)$, της X για $|t| < 1$.
(β) Για κάθε $k \in \mathbb{N}$, να υπολογιστεί η πιθανότητα $\mathbf{P}(X = k)$.
(γ) Τι κατανομή ακολουθεί η τυχαία μεταβλητή $X + 1$;

6. (15 Βαθμοί) Συμμετέχουμε σε ένα παιχνίδι σε κάθε γύρο του οποίου κερδίζουμε το ποσό 1 ευρώ με πιθανότητα 3/4 και χάνουμε το ποσό 4 ευρώ με πιθανότητα 1/4. Το αποτέλεσμα κάθε γύρου είναι ανεξάρτητο από τα αποτελέσματα όλων των προηγούμενων. Ποια ανισότητα πρέπει να ικανοποιεί το n ώστε η πιθανότητα μετά από n γύρους του παιχνιδιού να χάνουμε περισσότερα από 30 ευρώ να είναι μικρότερη από το 0.05.

[Δίνονται: $\Phi(0.05) = 0.52, \Phi(-1.65) = 0.05, \Phi(1.65) = 0.95$]

Άριστα είναι το 100. Η διάρκεια της εξέτασης είναι $2\frac{1}{2}$ ώρες. **Καλή επιτυχία!**

Μερικές απαντήσεις

1. (α) $p_{n,m} = \frac{(n-1)^m}{n^m}$. Για τις ευνοϊκές επιλογές, κάθε ένα από τα m σφαιρίδια έχει $n-1$ επιλογές, άρα με βάση την πολλαπλασιαστική αρχή έχουμε $(n-1)^m$ επιλογές. Πιο αναλυτικά, επειδή τα σφαιρίδια (και τα κουτιά) είναι διακεκριμένα, κάθε κατανομή αντιστοιχεί σε μια διατεταγμένη m-αδα $(k_1, k_2, \ldots, k_m)$ όπου $k_i \in \{1, 2, \ldots, n\}$ και αντίστροφα κάθε τέτοια m-αδα δίνει μια κατανομή. Το k_i δίνει τον αριθμό του κουτιού που περιέχει το i σφαιρίδιο. Έτσι δικαιολογείται η εφαρμογή της πολλαπλασιαστικής αρχής για την καταμέτρηση των δυνατών και ευνοϊκών αποτελεσμάτων.
(β) Έστω E_i το ενδεχόμενο το κουτί i να είναι άδειο. Τότε το πλήθος των άδειων κουτιών είναι

$$N = \mathbf{1}_{E_1} + \mathbf{1}_{E_2} + \cdots + \mathbf{1}_{E_n}$$

και

$$A_n = \mathbf{E}(N) = \mathbf{P}(E_1) + \mathbf{P}(E_2) + \cdots + \mathbf{P}(E_n) = np_{n,m}.$$

Στο σεναριο (i),

$$A_n = n\left(1 - \frac{1}{n}\right)^n \to \infty \cdot e^{-1} = \infty.$$

Για τα άλλα δύο σενάρια, γράφουμε

$$\log A_n = \log n + m \log\left(1 - \frac{1}{n}\right) = \log n + m\left(-\frac{1}{n} + \frac{g(n)}{n^2}\right)$$

με $g(n)$ φραγμένη ακολουθία (θεώρημα Taylor). Βλέπουμε λοιπόν ότι στο σενάριο (iii) το $\lim_{n\to\infty} A_n = -\infty$ (και άρα $A_n \to 0$) ενώ στο σενάριο (ii),

$$\log A_n = \log n - [\log n] + \frac{g(n) \log n}{n^2}.$$

Είναι σαφές ότι η τελευταία ποσότητα έχει $\limsup \le 1$ και $\liminf \ge 0$ (μάλιστα αυτά ισχύουν ως ισότητες). Άρα η A_n έχει $\liminf \ge 1$ και $\limsup \le e$.
(γ)

$$\frac{n\binom{n}{2}\{(n-1)!\}}{n^n} = \frac{\binom{n}{2}\{(n-1)!\}}{n^{n-1}}.$$

Για να κατασκευάσουμε μια ευνοϊκή κατανομή έχουμε n επιλογές για το ποιο κουτί θα μείνει άδειο. Έπειτα $\binom{n}{2}$ επιλογές για το ποιο ζευγάρι σφαιριδίων θα τοποθετηθούν μαζί σε ένα κουτί και τέλος $(n-1)!$ επιλογές για την τοποθέτηση των $n-1$ γκρούπ σφαιριδίων ($n-2$ σφαιρίδια το καθένα μόνο του και 1 ζευγάρι σφαιριδίων που τοποθετούνται στο ίδιο κουτί).

2. (α) $1097/10^5$.
(β) 98/1097.
(γ) Έστω $p = 98/1097$ η πιθανότητα ένας που παίρνει θετικό απολεσμα στο τεστ να είναι πράγματι άρρωστος. Το πλήθος N των αρρώστων στους επόμενους 5 με θετικό αποτέλεσμα ακολουθεί την διωνυμική κατανομή με παραμέτρους 5 και p. Έτσι,

$$\begin{aligned}\mathbf{P}(N \ge 3) &= 1 - \mathbf{P}(N \le 2) = 1 - \{\mathbf{P}(N = 0) + \mathbf{P}(N = 1) + \mathbf{P}(N = 2)]\} \\ &= 1 - \left\{\binom{5}{0}p^0(1-p)^5 + \binom{5}{1}p^(1-p)^4 + \binom{5}{2}p^2(1-p)^3\right\}.\end{aligned}$$

3. (α) $a = 1/2$.
(β) $f_X(x) = \frac{3x^2+1}{2}\mathbf{1}_{(0,1)}(x)$ για κάθε $x \in \mathbb{R}$ και $f_Y(y) = (y + \frac{1}{2})\mathbf{1}_{(0,1)}(y)$. για κάθε $y \in \mathbb{R}$.
(γ) Η $f_{X|Y}(x|y)$ ορίζεται για $y \in \mathbb{R}$ ώστε $f_Y(y) \ne 0$, δηλαδή για $y \in (0, 1)$, και για κάθε $x \in \mathbb{R}$ (όλες οι πυκνότητες τυχαίων μεταβλητών με τιμές στο $\mathbb{R}$ έχουν πεδίο ορισμού το $\mathbb{R}$. Και για κάθε $y \in (0, 1)$ η $x \mapsto f_{X|Y}(x|y)$ είναι μια πυκνότητα). Ο τύπος για τη δεσμευμένη πυκνότητα είναι

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{3x^2 + 2y}{1 + 2y}\mathbf{1}_{(0,1)}(x)$$

για κάθε $x \in \mathbb{R}$ και $y \in (0, 1)$.
(δ) $\mathbf{P}(X < Y) = 11/24$.

(ε) $\mathbf{E}(3X-2) = 3\mathbf{E}(X) - 2 = -1/8$.

4. (α) Η $M_X(t)$ είναι πεπερασμένη ακριβώς για $t \in (-\infty, \theta)$.

5. (α) Για $|t| < 1$ έχουμε $\mathbf{E}(t^X) = \mathbf{E}(\mathbf{E}(t^X|\Lambda)) = \mathbf{E}(m(\Lambda))$ όπου η m ορίζεται για $\lambda \in (0, \infty)$ ως

$$m(\lambda) := \mathbf{E}(t^X|\Lambda = \lambda) = \sum_{k=0}^{\infty} t^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{\lambda t - \lambda}.$$

Άρα

$$\mathbf{E}(m(\Lambda)) = \mathbf{E}(e^{(t-1)\Lambda}) = \frac{1}{1 - \frac{t-1}{a}}$$

με βάση το θέμα 4(α) πιο πάνω (αφού $t - 1 < a$).
(β) Αναπτύσσουμε την πιθανογεννήτρια του (α) ερωτήματος σε δυναμοσειρά με κέντρο το 0. Ισούται με

$$\frac{1}{1 + \frac{1}{a} - \frac{t}{a}} = \frac{a}{a+1} \frac{1}{\left(1 - \frac{t}{a+1}\right)} = \frac{a}{a+1} \sum_{k=0}^{\infty} \frac{1}{(a+1)^k} t^k.$$

Έπεται ότι $\mathbf{P}(X = k) = \frac{a}{a+1} \left(\frac{1}{a+1}\right)^k$ για κάθε $k \in \mathbb{N}$. Η ίδια σχέση προκύπτει και από την

$$\mathbf{P}(X = k) = \int_0^{\infty} f_\Lambda(\lambda) \mathbf{P}(X = k|\Lambda = \lambda)\, d\lambda = \int_0^{\infty} a e^{-a\lambda} e^{-\lambda} \frac{\lambda^k}{k!}\, d\lambda = ...$$

(γ) Η $X + 1$ ακολουθεί τη γεωμετρική κατανομή με πιθανότητα επιτυχίας $a/(a+1)$.

6. Έστω X_i το κέρδος μας στο παιχνίδι i. Οι $(X_i)_{i\geq 1}$ είναι ανεξάρτητες, ισόνομες με $\mathbf{P}(X_1 = 1) = 3/4, \mathbf{P}(X_1 = -4) = 1/4$. Η X_1 έχει μέση τιμή $\mu = -1/4$ και διασπορά $\sigma^2 = 75/16$. Έστω $S_n := X_1 + \cdots + X_n$ για κάθε $n \in \mathbb{N}^+$. Θέλουμε το n να είναι τέτοιο ώστε $\mathbf{P}(S_n \leq -30) \leq 0.05$. Εφαρμόζουμε το κεντρικό οριακό θεώρημα και δείχνουμε (δεχόμενοι την προσέγγιση για την πιθανότητα που μας δίνει το ΚΟΘ) ότι αυτή η σχέση ισοδυναμεί με

$$\frac{-120 + n}{5\sqrt{3n}} \leq -1.65$$

Αυτό ισχύει για $n \leq 35$.