

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
Τμήμα Επιστήμης Υπολογιστών

ΗΥ-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2024
Διδάσκων: Π. Τσακαλίδης

Δεύτερη Σειρά Ασκήσεων - Λύσεις

Άσκηση 1

Έστω ότι στις κατατακτήριες εξετάσεις του Τμήματος Επιστήμης Υπολογιστών προσέρχονται 1000 υποψήφιοι αλλά οι διαθέσιμες θέσεις είναι μόνο 70. Για τα αποτελέσματα των εξετάσεων γνωρίζετε ότι ο μέσος βαθμός είναι 60 και η τυπική απόκλιση $\sigma = 6$. Αν ο βαθμός κάποιου υποψήφιου είναι 85, μπορεί να υποθέσει ότι θα είναι ένας από τους επιτυχόντες; Χρησιμοποιήστε την ανισότητα Chebyshev.

Λύση

Για να λύσουμε το πρόβλημα θα χρησιμοποιήσουμε την εναλλακτική μορφή της ανισότητας Chebyshev:

$$P(|X - \mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}$$

που μας λέει ότι: Η πιθανότητα να πάρει η τ.μ. X τιμή μεγαλύτερη από k τυπικές αποκλίσεις μακριά από τη μέση τιμή της, μ , είναι το πολύ $\frac{1}{k^2}$.

Αρχικά, βρίσκουμε το k , δηλαδή πόσες τυπικές αποκλίσεις απέχει το 84 από το μ :

$$k = \frac{84 - \mu}{\sigma} = \frac{84 - 60}{6} = 4$$

Χρησιμοποιώντας την ανισότητα Chebyshev:

$$P(X \geq 84) \leq \frac{1}{4^2} = \frac{1}{16}$$

Επομένως, μπορούμε να συμπεράνουμε ότι **το πολύ** $1000 \times \frac{1}{16} = 62.5 \approx 63$ υποψήφιοι κατάφεραν να γράψουν από 84 και πάνω. Τέλος, αφού οι θέσεις είναι 70, ο υποψήφιος που έγραψε 85 μπορεί να υποθέσει ότι θα είναι ένας από τους επιτυχόντες.

Άσκηση 2

Σε ένα παιχνίδι με βελάκια ένας παίκτης έχει στη διάθεση του n βελάκια. Ανάλογα την περιοχή του στόχου που θα προσγειωθεί το βελάκι ο παίκτης μπορεί να κερδίσει από 1 έως και 6 πόντους. Έστω X_i , οι πόντοι της i -οστής ρίψης και $X_i \in [1, 6]$. Υποθέτουμε ότι οι τ.μ. X_i είναι ανεξάρτητες και ορίζουμε το άθροισμα των ρίψεων: $S_n = X_1 + X_2 + \dots + X_n$. Υπολογίστε ένα άνω όριο για την πιθανότητα $P(S_{100} \geq 400)$:

- (α) Χρησιμοποιώντας την ανισότητα Markov.
- (β) Χρησιμοποιώντας την ανισότητα Chebyshev.
- (γ) Χρησιμοποιώντας προσέγγιση με βάση το Κεντρικό Οριακό Θεώρημα.

Τι παρατηρείτε;

Λύση

(α) Η μέση τιμή των τ.μ. X_i γίνεται:

$$E[X_i] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

Σύμφωνα με την ανισότητα Markov έχουμε:

$$P[S_{100} \geq 400] \leq \frac{100 \cdot 3.5}{400}$$

(β) Υπολογίζουμε την διασπορά των τ.μ. X_i :

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2$$

και

$$E[X_i^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = 15$$

Επομένως, η διασπορά γίνεται:

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2 = 15 - 3.5^2 = 2.75$$

Τέλος, σύμφωνα με την ανισότητα του Chebyshev έχουμε:

$$P[S_{100} \geq 400] = P[|S_{100} - 350| \geq 50] \leq \frac{100 \cdot 2.75}{50^2} = 0.11$$

(γ) Με βάση το Κεντρικό Οριακό Θεώρημα έχουμε:

$$P[S_{100} \geq 400] = P\left[\frac{S_{100} - 350}{\sqrt{2.75 \cdot 100}} \geq \frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right] \approx$$

$$1 - \Phi\left(\frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right) \approx$$

$$1 - \Phi(3.015) \approx$$

$$0.0009$$

Άσκηση 3

Η ποσότητα γατοτροφής που πωλείται εβδομαδιαία από ένα pet shop ακολουθεί κανονική κατανομή με μέση τιμή $\mu = 50$ κιλά και τυπική απόκλιση $\sigma = 10$ κιλά. Η αρχική ποσότητα γατοτροφής που διαθέτει το pet shop είναι 74 κιλά και εφοδιάζεται εβδομαδιαία με 47 κιλά την φορά.

(α) Ποια είναι η πιθανότητα μετά από 11 εβδομάδες το συνολικό απόθεμα γατοτροφής του pet shop να είναι λιγότερο από 20 κιλά;

(β) Με πόσα κιλά θα έπρεπε να εφοδιάζεται το pet shop κάθε εβδομάδα έτσι ώστε η πιθανότητα μετά από 11 εβδομάδες το συνολικό απόθεμα να είναι λιγότερο από 20 κιλά, να είναι ίση μόνο με 0.5%;

Λύση

Μας δίνεται ότι $\mu = 50$, $\sigma = 10$, $n = 11$.

Το αρχικό απόθεμα γατοτροφής που διαθέτει το κατάστημα είναι 74 κιλά και εφοδιάζεται εβδομαδιαία με 47, επομένως, το συνολικό απόθεμα στο διάστημα των 11 εβδομάδων θα είναι: $74 + 11 \times 47 = 591$ κιλά.

- (α) Το συνολικό απόθεμα του pet shop θα είναι λιγότερο από 20 κιλά μετά από 11 εβδομάδες, αν πουληθούν περισσότερα από $591 - 20 = 571$ κιλά.

Επομένως,

$$P(X > 571) = P\left(z > \frac{571 - 11 \cdot 50}{\sqrt{11} \cdot 10}\right) = P(z > 0.63) = 1 - \Phi(0.63) = 1 - 0.7357 = 0.2643$$

- (β) Έστω A η ζητούμενη εβδομαδιαία ποσότητα εφοδιασμού. Τώρα, το συνολικό απόθεμα του pet shop θα είναι λιγότερο από 20 κιλά, αν πουληθούν περισσότερα από $74 + 11 \times A - 20$ κιλά.

Και θέλουμε:

$$P(X > 74 + 11 \times A - 20) = 0.005$$

Η z -value που αντιστοιχεί σε αυτήν την πιθανότητα είναι 2.575. Επομένως,

$$\frac{74 + 11 \times A - 20 - 11 \cdot 50}{\sqrt{11} \cdot 10} = 2.575$$

$$A \approx 52.85$$

Άρα, το pet shop θα πρέπει να εφοδιάζεται με 52.85 κιλά γατοτροφής την εβδομάδα.

Άσκηση 4

Κατά τη διάρκεια της καλοκαιρινής πρακτικής άσκησης σας, εργάζεστε σ'ένα εργοστάσιο παραγωγής μικρο-επεξεργαστών. Ο επιβλέπων σας ζητά να εκτιμήσετε την ποιότητα της γραμμής παραγωγής, δηλαδή την πιθανότητα p ένας μικρο-επεξεργαστής να μην είναι ελαττωματικός. Υποθέτουμε ότι οι μικρο-επεξεργαστές είναι καλοί με την ίδια πιθανότητα p , και ανεξάρτητοι ο ένας από τον άλλο.

- (α) Υποθέτοντας ότι τεστάρετε ένα σύνολο από επιλεγμένους μικρο-επεξεργαστές, ορίστε μια εκτιμήτρια Z_n , για την παράμετρο p , ώστε η Z_n να συγκλίνει στο p ως προς την πιθανότητα.
- (β) Ο επιβλέπων σας ζητά η εκτιμήτρια Z_n , να βρίσκεται στη γειτονιά τιμών $p \pm 0.1$, με πιθανότητα 0.95. Αρκεί να τεστάρετε 27 επιλεγμένους μικρο-επεξεργαστές για να ικανοποιήσετε αυτήν την απαίτηση; Χρησιμοποιήστε τόσο την ανισότητα Chebyshev, όσο και το Κεντρικό Οριακό Θεώρημα, για να απαντήσετε στο ερώτημα.

Λύση

- (α) Έστω X_i τ.μ. που χαρακτηρίζει την ποιότητα του i -οστού επεξεργαστή (0: για ελαττωματικός, 1: για μη-ελαττωματικός). Οι $X_i, i = 1, \dots, n$ είναι ανεξάρτητες Bernoulli τ.μ.

Ορίζουμε Z_n ως:

$$Z_n = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Η ανισότητα του Chebyshev ορίζεται ως:

$$P(|Z_n - E[Z_n]| \geq \epsilon) \leq \frac{\text{var}(Z_n)}{\epsilon^2}.$$

Εφόσον $E[Z_n] = \frac{n \cdot E[X_i]}{n} = \frac{n \cdot p}{n} = p$ και $\text{var}(Z_n) = \frac{n \cdot \text{var}(X_i)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$, αφού οι X_i είναι ανεξάρτητες τ.μ. Bernoulli, έχουμε ότι:

$$P(|Z_n - p| \geq \epsilon) \leq \frac{p(1-p)}{\epsilon^2},$$

άρα

$$\lim_{n \rightarrow \infty} P(|Z_n - p| \geq \epsilon) = 0$$

αφού

$$\lim_{n \rightarrow \infty} \frac{p(1-p)}{n \cdot \epsilon^2} = 0.$$

Οπότε, αυτό σημαίνει ότι η Z_n συγχλίνει στο p ως προς την πιθανότητα.

(β) **Χρησιμοποιώντας την ανισότητα του Chebyshev:**

Εφόσον θέλουμε η Z_n , να βρίσκεται στη γειτονιά τιμών $p \pm 0.1$, με πιθανότητα 0.95, θέλουμε

$$P(|Z_n - p| \geq 0.1) = 1 - 0.95 = 0.05$$

Επίσης, εφαρμόζοντας την ανισότητα Chebyshev για $\epsilon = 0.1$ προκύπτει ότι:

$$P(|Z_n - p| \geq 0.1) \leq \frac{p(1-p)}{n \cdot 0.1^2} \leq \frac{1/4}{n \cdot 0.01}$$

αφού η διασπορά $\text{var}(X_i) = p(1-p)$ της X_i μεγιστοποιείται στο $p = \frac{1}{2}$, και άρα $p(1-p) \leq \frac{1}{4}$.
Άρα τελικά

$$0.05 \leq \frac{1/4}{n \cdot 0.01}$$

και

$$n \leq 500.$$

Επομένως, χρησιμοποιώντας την ανισότητα του Chebyshev δε μπορούμε να πούμε αν λιγότεροι από 500 μικρο-επεξεργαστές επαρκούν για το τεστ, μιας και δε γνωρίζουμε τη διασπορά. Αν για παράδειγμα η διασπορά είναι μικρή, το οποίο μπορεί να συμβεί όταν το p είναι αρκετά μικρό, 27 μικρο-επεξεργαστές θα μπορούσαν να επαρκούν στην πράξη. Συνεπώς, η απάντηση στο ερώτημα είναι: "εξαρτάται"!

Χρησιμοποιώντας το Κεντρικό Οριακό Θεώρημα:

Καθώς $n \rightarrow \infty$, η $Z_n = \frac{X_1 + X_2 + \dots + X_n}{n}$, συμπεριφέρεται σαν κανονικά κατανοημένη με μέση τιμή $E[Z_n] = \frac{n \cdot p}{n} = p$ και διασπορά $var(Z_n) = \frac{1}{n^2} \cdot p \cdot (1-p) \cdot n = \frac{p \cdot (1-p)}{n} = \frac{\sigma^2}{n}$.
Θέλουμε να βρούμε τέτοιο n ώστε:

$$P(|Z_n - p| \leq 0.1) = 0.95$$

$$P(-0.1 \leq Z_n - p \leq 0.1) = 0.95$$

$$P\left(-\frac{0.1}{\sigma/\sqrt{n}} \leq \frac{Z_n - p}{\sigma/\sqrt{n}} \leq \frac{0.1}{\sigma/\sqrt{n}}\right) = 0.95$$

$$\Phi\left(\frac{0.1}{\sigma/\sqrt{n}}\right) - \Phi\left(-\frac{0.1}{\sigma/\sqrt{n}}\right) = 2 \cdot \Phi\left(\frac{0.1}{\sigma/\sqrt{n}}\right) - 1 = 0.95$$

Στη χειρότερη περίπτωση, η διασπορά της X_i είναι μεγάλη, οπότε απαιτούνται περισσότερα δείγματα n για να ικανοποιήσουμε το κριτήριο.

Έχουμε ότι $\sigma^2 = p(1-p) \leq \frac{1}{4}$.

Άρα, στη χειρότερη περίπτωση έχουμε:

$$2 \cdot \Phi\left(\frac{0.1}{\sigma/\sqrt{n}}\right) - 1 = 0.95$$

$$2 \cdot \Phi\left(\frac{0.1 \cdot \sqrt{n}}{1/2}\right) - 1 = 0.95$$

$$2 \cdot \Phi(0.2 \cdot \sqrt{n}) - 1 = 0.95$$

$$\Phi(0.2 \cdot \sqrt{n}) = 0.9750$$

$$\Phi(0.2 \cdot \sqrt{n}) = \Phi(1.96)$$

$$0.2 \cdot \sqrt{n} = 1.96$$

$$\sqrt{n} = 9.8 \Leftrightarrow n = 96.04$$

Επομένως, $n \geq 96$, δηλαδή χρειαζόμαστε περισσότερες από 96 προσπάθειες.

Λύση

Για κάθε πεταλούδα που επιλέγεται για το δείγμα έχουμε μία Bernoulli τ.μ., με $X_i = 1$ όταν η i -οστή πεταλούδα ανήκει στο είδος προς μελέτη.

Μπορούμε να υπολογίσουμε την τιμή του n χρησιμοποιώντας το Κεντρικό Οριακό Θεώρημα ως εξής:

$$P\left(\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0.9$$

όπου $\frac{S_n}{n} = 0.5$ αφού θέλουμε το 50% των πεταλούδων του δείγματος να ανήκει στο είδος που θέλουν να μελετήσουν οι ερευνητές.

Επιπλέον, γνωρίζουμε ότι $\mu = 0.45 (= p)$ και $\sigma = \sqrt{p(1-p)} = \sqrt{0.55 \cdot 0.55} = 0.55$.

Και από τον z-table:

$$\Phi(z) = 0.9 \Rightarrow z = 1.2816$$

Επομένως, έχουμε (για το minimum n):

$$\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = 1.2816 \Leftrightarrow$$

$$\frac{0.5 - 0.45}{\frac{0.55}{\sqrt{n}}} = 1.2816 \Leftrightarrow$$

$$n^2 = \frac{1.2816 \cdot 0.55}{0.5 - 0.45} \Leftrightarrow n \approx 199$$