

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
Τμήμα Επιστήμης Υπολογιστών

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

Φροντιστήριο 2

Άσκηση 1

Ο μηχανικός της υπηρεσίας συστημάτων και δικτύων του Υπολογιστικού Κέντρου του Πανεπιστημίου Κρήτης θέλει να εκτιμήσει το ποσοστό p , των πακέτων πληροφορίας που ταξιδεύουν μέσα από την οπτική ίνα που συνδέει την Κνωσσό με τις Βούτες, τα οποία πακέτα είναι ψηφιακού video (digital video disk - DVD). Ο μηχανικός γράφει ένα script, για να ελέγξει $n = 1000$ πακέτα, μετράει το πλήθος S_n των DVD και χρησιμοποιεί τη σχέση: $\hat{p} = \frac{S_n}{1000}$ ως τον εκτιμητή του p . Κάθε πακέτο είναι ένα DVD πακέτο με πιθανότητα p , ανεξάρτητο από τα άλλα. Να υπολογισθούν:

- (α) Τι είδους τυχαία μεταβλητή είναι η S_n ?
- (β) Ποια είναι η μέση τιμή και η διασπορά της τυχαίας μεταβλητής $\hat{p} = \frac{S_n}{1000}$?
- (γ) Ποια είναι η πιθανότητα ο εκτιμητής p να κάνει λάθος λιγότερο από 2%?
- (δ) Βρείτε τον αριθμό δ για τον οποίο ισχύει: $P(|\hat{p} - p| \leq \delta) \approx 0.99$, όταν $p = 0.5$, και όταν $p = 0.1$. Δίνεται ότι: $\Phi(2.58) = 0.995$.

Λύση

- (α) Η τυχαία μεταβλητή S_n είναι της μορφής: $S_n = X_1 + \dots + X_{1000}$, όπου η X_i είναι Bernoulli τυχαία μεταβλητή, με $X_i = 1$, όταν το i -οστό πακέτο είναι DVD πακέτο:

$$X_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}.$$

Επομένως, η S_n είναι διωνυμική με παραμέτρους $n = 1000$ και p .

- (β) Η τυχαία μεταβλητή S_n έχει διασπορά $var(S_n) = np(1-p)$. Επομένως, η τ.μ $\hat{p} = \frac{S_n}{n}$ έχει διασπορά:

$$var(\hat{p}) = \frac{1}{n^2} var(S_n) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

και μέση τιμή:

$$E[\hat{p}] = \frac{E[S_n]}{n} = \frac{np}{n} = p$$

(γ) Η ζητούμενη πιθανότητα γίνεται:

$$\begin{aligned}
P(|\hat{p} - p| \leq \delta) &= P(-\delta \leq \hat{p} - p \leq \delta) = P\left(\frac{-\delta}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\delta}{\sqrt{\frac{p(1-p)}{n}}}\right) \\
&= P\left(\frac{-\delta\sqrt{n}}{\sqrt{p(1-p)}} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq \frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&\approx \Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= \Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 + \Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) \\
&= 2 \cdot \Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1
\end{aligned}$$

Για $p = 0.5$, $\delta = 0.02$ έχουμε:

$$P(|\hat{p} - 0.5| \leq 0.02) = 2 \cdot \Phi\left[\frac{0.02 \cdot \sqrt{1000}}{\sqrt{0.5 \cdot 0.5}}\right] - 1 = 2 \cdot \Phi(1.265) - 1 \approx 0.794 = 79.4\%$$

Για $p = 0.1$, $\delta = 0.02$ έχουμε:

$$P(|\hat{p} - 0.1| \leq 0.02) = 2 \cdot \Phi\left[\frac{0.02 \cdot \sqrt{1000}}{\sqrt{0.1 \cdot 0.9}}\right] - 1 \approx 0.965 = 96.5\%$$

Παρατηρούμε ότι όσο πιο μικρό είναι το ποσοστό των πακέτων πληροφορίας, τόσο μεγαλύτερη είναι η πιθανότητα ο εκτιμητής να κάνει λάθος λιγότερο από 2%.

(δ) Έχουμε ότι:

$$2 \cdot \Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 = 0.99$$

$$\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) = \frac{1.99}{2}$$

$$\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) = 0.995$$

$$\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) = \Phi(2.58)$$

$$\delta = 2.58 \cdot \sqrt{\frac{p(1-p)}{1000}}$$

$$\text{Για } p = 0.5, \delta = 2.58 \cdot \sqrt{\frac{0.5 \cdot 0.5}{1000}} = 0.04.$$

$$\text{Για } p = 0.1, \delta = 2.58 \cdot \sqrt{\frac{0.1 \cdot 0.9}{1000}} = 0.0025.$$

Άσκηση 2

Ψηφοφόροι κάποιας περιφέρειας ψήφισαν για έναν από τους υποψηφίους Α και Β. Η αρχική καταμέτρηση έδειξε ότι ο υποψήφιος Β προηγείται του Α, κατά 537 ψήφους. Όμως, υπάρχουν 25.600 ψηφοδέλτια τα οποία εκκρεμούν στο εκλοδικείο και δεν έχουν καταχωρηθεί σε κανέναν από τους δύο υποψηφίους. Υποθέστε ότι τελικά και τα 25.600 ψηφοδέλτια, θα βρεθούν έγκυρα και θα αποδοθούν ισοπίθानα στους δύο υποψηφίους, κάθε ένα ανεξάρτητα από τα υπόλοιπα. Έστω X και Y το πλήθος των ψηφοδελτίων (από τα 25.600) που θα αποδοθούν στους Α και Β αντίστοιχα.

- (α) Τι είδους μεταβλητή είναι η X ?
- (β) Να βρεθούν η μέση τιμή και η διασπορά της τ.μ X .
- (γ) Είναι οι τ.μ X και Y ανεξάρτητες?
- (δ) Από τη στιγμή που και οι 25.600 ψήφοι προσμετρηθούν στο τελικό αποτέλεσμα, ποια είναι η μέση τιμή και η διασπορά της τ.μ $Z = X - Y$, του καθαρού κέρδους του υποψηφίου Α στην ψηφοφορία?
- (ε) Από τη στιγμή που και οι 25.600 ψήφοι προσμετρηθούν στο τελικό αποτέλεσμα, χρησιμοποιήστε το Κεντρικό Οριακό Θεώρημα για να βρείτε την πιθανότητα: "Να κερδίσει ο Β την εκλογή".

Λύση

- (α) Η τυχαία μεταβλητή X είναι διωνυμική με παραμέτρους: $(n, p) = (25.600, 1/2)$.
- (β) Η μέση τιμή και η διασπορά της τ.μ X γίνονται:

$$E[X] = np = 25.600 \cdot \frac{1}{2} = 12.800$$

και

$$\text{var}(X) = np(1 - p) = 25.600 \cdot \left(\frac{1}{2}\right)^2 = 6.400$$

- (γ) Έχουμε ότι: $X + Y = 25.600$. Άρα, οι τ.μ X, Y δεν μπορούν να είναι ανεξάρτητες.
- (δ) Έχουμε ότι: $Z = X - Y$. Όμως: $X + Y = 25.600 \Rightarrow Y = 25.600 - X$.
Η μέση τιμή της τ.μ Z γίνεται:

$$E[Z] = E[X] - E[Y] = E[X] - E[25.600 - X] = 2 \cdot E[X] - 25.600 = 2 \cdot 12.800 - 25.600 = 0$$

Για τη διασπορά της τ.μ Z έχουμε:

$$\text{var}(Z) = \text{var}(X - Y) = \text{var}(X - 25.600 + X) = \text{var}(2X - 25.600) = 4 \cdot \text{var}(X) = 4 \cdot 6.400 = 25.600$$

- (ε) Η Z μπορεί να προσεγγιστεί από μια κανονική τ.μ με $Z \sim N(0, 25.600)$. Σύμφωνα με το Κεντρικό Οριακό Θεώρημα, η πιθανότητα γίνεται:

$$P(\text{"Ο Β κερδίζει"}) = P(Z \leq 536) \approx \Phi\left(\frac{536 - 0}{\sqrt{25.600}}\right) = \Phi(3.35) \approx 0.9996$$

Άσκηση 3

Το ημερήσιο εισόδημα χαρτοπαίκτη ακολουθεί την κατανομή $\text{Uniform}([-5,5])$ σε χιλιάδες ευρώ. Να υπολογίσετε προσεγγιστικά:

- a) την πιθανότητα σε 48 μέρες να κερδίσει τουλάχιστον 30 χιλιάδες ευρώ
- b) το ποσό s ώστε με πιθανότητα τουλάχιστον 95% το εισόδημα 48 μερών να είναι μικρότερο ή ίσο με s .

Λύση

$P(S_n) \leq \eta \geq x \geq y$, όπου S_n, x, y είναι οι μέρες, το ποσό και το επίπεδο βεβαιότητας αντίστοιχα και

X_i είναι το εισόδημα την μέρα i

$$E[X_i] = \int_{-5}^5 x \frac{1}{b-a} dx = 0$$

$$E[S_n] = 0n = 0$$

$$\text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = \int_{-5}^5 x^2 \frac{1}{b-a} dx = \frac{1}{10} \left[\frac{x^3}{3} \right]_{x=-5}^5 = 25/3$$

$$\text{Var}[S_n] = n25/3$$

a)

$$\begin{aligned} P(S_{48}) \geq 30 &= P\left(\frac{S_{48} - E[S_{48}]}{\sqrt{\text{Var}[S_{48}]}} \geq \frac{30 - 0}{\sqrt{48 \cdot 25/3}}\right) \\ &= P(Z \leq 30/20) \\ &= 1 - \Phi(1.5) \end{aligned}$$

b)

$$\begin{aligned} P(|S_{48}| \leq s) &\geq 0.95 \\ P(-s \leq S_{48} \leq s) &\geq 0.95 \\ P\left(\frac{-s - E[S_{48}]}{\sqrt{\text{Var}[S_{48}]}} \leq \frac{S_{48} - E[S_{48}]}{\sqrt{\text{Var}[S_{48}]}} \leq \frac{-s - E[S_{48}]}{\sqrt{\text{Var}[S_{48}]}}\right) &\geq 0.95 \\ P(-s/20 \leq Z \leq s/20) &\geq 0.95 \\ \Phi(s/20) - \Phi(-s/20) &\geq 0.95 \\ 2\Phi(s/20) - 1 &\geq 0.95 \\ \Phi(s/20) &\geq 0.975 = \Phi(1.96) \\ s &\geq 20 * 1.96 \end{aligned}$$

Άσκηση 4

Σε μια πόλη 4000 κατοίκων, κατά μέσο όρο 10 άτομα/μέρα χρειάζονται νοσηλεία. Ποιός είναι ο προσεγγιστικός υπολογισμός ελαχίστου αριθμού κρεβατιών, ώστε η πόλη να εξυπηρετείται χωρίς διακομιδές ασθενών σε άλλες πόλεις με πιθανότητα 95%;

Λύση

$I_k = 1$ εάν ο κάτοικος k χρειάζεται νοσηλεία μια συγκεκριμένη μέρα, $I_k = 0$ εάν όχι
 $S_{4000} = \sum_{k=1}^{4000} I_k$: πλήθος κατοίκων που χρειάζεται νοσηλεία μια συγκεκριμένη μέρα

$$\begin{aligned} P(S_{4000} \leq x) &\geq 0.95 \\ P(S_{4000} \leq x + 1/2) &\geq 0.95 \\ P\left(\frac{S_{4000} - E[S_{4000}]}{\sqrt{Var[S_{4000}]}} \leq \frac{x + 1/2 - E[S_{4000}]}{\sqrt{Var[S_{4000}]}}\right) &\geq 0.95 \\ P\left(Z \leq \frac{x + 1/2 - 1}{\sqrt{0.99}}\right) &\geq 0.95 \\ \Phi\left(\frac{x - 9.5}{\sqrt{9.9}}\right) &\geq 0.95 = \Phi(1.96) \\ x &\geq 0.5 + 1.96 * \sqrt{0.99} \end{aligned}$$

δεδομένου ότι

$$E[I_k] = 10/4000,$$

$$E[S_{4000}] = 4000E[I_k],$$

$$Var[I_k] = E[I_k^2] - E[I_k]^2 = 399/400^2,$$

$$Var[S_{4000}] = 4000 - 399/400^2 = 3.9.$$

Άσκηση 5

Η μέση τιμή και η τυπική απόκλιση του σιδηρού του αίματος του πληθυσμού ενός χωριού είναι 120mg/100ml και 15mg/100ml αντίστοιχα. Ποιά είναι η πιθανότητα ένα τυχαίο δείγμα 50 ατόμων από τον πληθυσμό αυτόν να έχει μέση τιμή σιδηρού του αίματος:

a) μεγαλύτερη από 125 mg/100ml;

b) μεταξύ 115 και 125 mg/100ml;

Λύση

$$\begin{aligned} \text{a) } P(X_n > 125) &= P\left(\frac{X_n - 120}{2.12} > \frac{125 - 120}{2.12}\right) = P(Z > 2.36) \\ &= 1 - P(Z \leq 2.36) = 1 - 0.991 = 0.009 \end{aligned}$$

$$\begin{aligned} \text{b) } P(115 < X_n < 125) &= P\left(\frac{115 - 120}{2.12} \geq \frac{X_n - 120}{2.12} \leq \frac{125 - 120}{2.12}\right) \\ &= P(-2.36 \leq Z \leq 2.36) = 0.991 - 0.009 = 0.982 \end{aligned}$$