Zadanie nr 3 RPIS

Piotr Pijanowski

21 maja 2025

1 Treść zadań

Rozpatrujemy zmienne losowe X_{30}, X_{150}, X_{600} o rozkładzie $B(n, p)$, gdzie $p = \frac{1}{3}$, natomiast $n = 30, 150, 600$.

1. Obliczyć (wprost) prawdopodobieństwa:
   $P(8 \leq X_{30} \leq 12), \quad P(40 \leq X_{150} \leq 60), \quad P(160 \leq X_{600} \leq 240)$.
2. Wyznaczyć przybliżenia tych prawdopodobieństw wynikające z nierówności Czebyszewa.
3. Wyznaczyć przybliżenia tych prawdopodobieństw zastępując $P(X_n \leq \alpha)$ przez odpowiednio dobrane z, takie aby $P(X_n \leq \alpha) \approx \Phi(z)$.
4. Oszacować te prawdopodobieństwa z pomocą nierówności Chernoffa.

2 Zadanie nr 1

Rozkład zmiennej losowej $X \sim B(n, p)$ opisuje liczbę sukcesów w n niezależnych próbach Bernoulliego, z których każda ma prawdopodobieństwo sukcesu równe p. Jedna próba Bernoulliego to zmienna losowa Y taka, że:

$$P(Y = 1) = p, \quad P(Y = 0) = 1 - p.$$

Jeśli X to suma n takich niezależnych zmiennych Y_i, to X ma rozkład dwumianowy:

$$X \sim B(n, p), \quad \text{czyli:} \quad P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Aby obliczyć prawdopodobieństwo, że zmienna losowa X przyjmie wartość z przedziału domkniętego $[a, b]$, korzystamy z faktu, że:

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a).$$

Ponieważ X jest zmienną losową dyskretną przyjmującą wartości całkowite, to:

$$P(X < a) = P(X \leq a - 1),$$

a więc:

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a - 1).$$

Możemy też zapisać to w formie sumy:

$$P(a \leq X \leq b) = \sum_{k=a}^{b} \binom{n}{k} p^k (1-p)^{n-k}.$$

W naszym przypadku rozpatrujemy:

- $X_{30} \sim B(30, \frac{1}{3})$,
- $X_{150} \sim B(150, \frac{1}{3})$,
- $X_{600} \sim B(600, \frac{1}{3})$.

Zatem odpowiednie prawdopodobieństwa wyrażają się przez sumy:

$$P(8 \leq X_{30} \leq 12) = \sum_{k=8}^{12} \binom{30}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{30-k},$$

$$P(40 \leq X_{150} \leq 60) = \sum_{k=40}^{60} \binom{150}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{150-k},$$

$$P(160 \leq X_{600} \leq 240) = \sum_{k=160}^{240} \binom{600}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{600-k}.$$

Wartości te zostały obliczone przy pomocy funkcji arkusza kalkulacyjnego (np. Excel lub Google Sheets), korzystającego z funkcji dystrybuanty rozkładu dwumianowego.

Uzyskane przybliżone wyniki to:

$$P(8 \leq X_{30} \leq 12) \approx 0{,}667,$$
$$P(40 \leq X_{150} \leq 60) \approx 0{,}932,$$
$$P(160 \leq X_{600} \leq 240) \approx 0{,}9996.$$

3 Zadanie nr 2

Nierówność Czebyszewa mówi, że dla dowolnej zmiennej losowej X o wartości oczekiwanej μ i wariancji σ^2, dla każdej liczby rzeczywistej $a > 0$ zachodzi:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Ponieważ prawdopodobieństwo całkowite wynosi 1, możemy z tej nierówności wyprowadzić:

$$P(|X - \mu| < a) \geq 1 - \frac{\sigma^2}{a^2}.$$

czyli:

$$P(\mu - a \leq X \leq \mu + a) \geq 1 - \frac{\sigma^2}{a^2}.$$

Dla zmiennej $X \sim B(n, p)$ mamy:

$$\mathbb{E}(X) = np, \quad \mathrm{Var}(X) = np(1 - p).$$

Wyznaczmy wartości oczekiwane i wariancje dla rozważanych zmiennych:

- $X_{30} \sim B(30, \frac{1}{3})$: $\quad \mu = 10$, $\sigma^2 = \frac{60}{9} = 6.\overline{6}$,
- $X_{150} \sim B(150, \frac{1}{3})$: $\quad \mu = 50$, $\sigma^2 = 33.\overline{3}$,
- $X_{600} \sim B(600, \frac{1}{3})$: $\quad \mu = 200$, $\sigma^2 = 133.\overline{3}$.

Rozważane przedziały są symetryczne względem wartości oczekiwanej μ, tj.:

$$\begin{aligned} P(8 \leq X_{30} \leq 12) & \quad \text{jest symetryczny wokół } \mu = 10, \\ P(40 \leq X_{150} \leq 60) & \quad \text{jest symetryczny wokół } \mu = 50, \\ P(160 \leq X_{600} \leq 240) & \quad \text{jest symetryczny wokół } \mu = 200. \end{aligned}$$

Dlatego możemy zastosować wersję nierówności Czebyszewa dla przedziału $[\mu - a, \mu + a]$, co daje:

$$P(\mu - a \leq X \leq \mu + a) \geq 1 - \frac{\sigma^2}{a^2}.$$

Wyznaczmy dolne oszacowania dla poszczególnych przypadków:

- Dla X_{30}: $\mu = 10$, $a = 2$, $\sigma^2 = 6.\overline{6}$:

$$P(8 \leq X_{30} \leq 12) \geq 1 - \frac{6.\overline{6}}{4} = 1 - 1.666... \quad \Rightarrow \text{brak sensu (ujemna wartość).}$$

- Dla X_{150}: $\mu = 50$, $a = 10$, $\sigma^2 = 33.\overline{3}$:

$$P(40 \leq X_{150} \leq 60) \geq 1 - \frac{33.\overline{3}}{100} = 1 - 0.333 = 0.667.$$

- Dla X_{600}: $\mu = 200$, $a = 40$, $\sigma^2 = 133.\overline{3}$:

$$P(160 \leq X_{600} \leq 240) \geq 1 - \frac{133.\overline{3}}{1600} = 1 - 0.0833 = 0.917.$$

Wnioski

Z nierówności Czebyszewa wynika, że:

- Dla X_{30} oszacowanie jest bezużyteczne – wynik wychodzi ujemny, ponieważ przedział jest zbyt wąski w stosunku do wariancji.
- Dla X_{150} uzyskaliśmy oszacowanie $\geq$ 0,667, które jest zgodne z intuicją i znacząco mniejsze niż wartość rzeczywista.
- Dla X_{600} oszacowanie wynosi $\geq$ 0,917, co już dobrze przybliża rzeczywiste prawdopodobieństwo (0,9996).

4 Zadanie nr 3

Dla dużych wartości n, rozkład dwumianowy $B(n, p)$ można przybliżyć rozkładem normalnym, korzystając z **centralnego twierdzenia granicznego (CLT)**. Zmienna losowa $X \sim B(n, p)$ ma wartość oczekiwaną $\mu = np$ i wariancję $\sigma^2 = np(1 - p)$, więc:

$$X \approx \mathcal{N}(\mu, \sigma^2).$$

Aby obliczyć przybliżone prawdopodobieństwo $P(a \leq X \leq b)$, przekształcamy zmienną X do postaci standaryzowanej $Z \sim \mathcal{N}(0, 1)$, stosując wzór:

$$z = \frac{x - \mu}{\sigma},$$

gdzie:

- x — konkretna wartość graniczna (lewa lub prawa),
- $\mu = np$ — wartość oczekiwana zmiennej X,
- $\sigma = \sqrt{np(1 - p)}$ — odchylenie standardowe.

Aby poprawić dokładność przybliżenia, stosujemy tzw. **poprawkę ciągłości**, czyli:

$$z_1 = \frac{a - 0.5 - \mu}{\sigma}, \quad z_2 = \frac{b + 0.5 - \mu}{\sigma}.$$

Wówczas:

$$P(a \leq X \leq b) \approx \Phi(z_2) - \Phi(z_1),$$

gdzie $\Phi(z)$ to dystrybuanta standardowego rozkładu normalnego.

Wyznaczmy teraz wartości μ, σ, z-score'y oraz przybliżone prawdopodobieństwa dla naszych przypadków:

- **Dla** $X_{30} \sim B(30, \frac{1}{3})$:

$$\mu = 10, \quad \sigma \approx 2.58, \quad z_1 \approx -0.58, \quad z_2 \approx 0.97,$$

$$P(8 \leq X_{30} \leq 12) \approx \Phi(0.97) - \Phi(-0.58) \approx 0.834 - 0.281 = 0.553.$$

- **Dla** $X_{150} \sim B(150, \frac{1}{3})$:

$$\mu = 50, \quad \sigma \approx 5.77, \quad z_1 \approx -1.82, \quad z_2 \approx 1.82,$$

$$P(40 \leq X_{150} \leq 60) \approx \Phi(1.82) - \Phi(-1.82) \approx 0.9656 - 0.0344 = 0.931.$$

- **Dla** $X_{600} \sim B(600, \frac{1}{3})$:

$$\mu = 200, \quad \sigma \approx 11.55, \quad z_1 \approx -3.49, \quad z_2 \approx 3.49,$$

$$P(160 \leq X_{600} \leq 240) \approx \Phi(3.49) - \Phi(-3.49) \approx 0.99976 - 0.00024 = 0.99952.$$

Wnioski

Dla większych wartości n, przybliżenie rozkładem normalnym jest coraz dokładniejsze. Dla X_{150} i X_{600} wyniki są bardzo bliskie rzeczywistym wartościom.

5 Zadanie nr 4

W tym zadaniu wykorzystujemy nierówność Chernoffa do oszacowania prawdopodobieństw:

$$P(8 \leq X_{30} \leq 12), \quad P(40 \leq X_{150} \leq 60), \quad P(160 \leq X_{600} \leq 240),$$

dla zmiennych losowych $X_n \sim B(n, \frac{1}{3})$.

Nierówność Chernoffa — wersja symetryczna

Dla zmiennej losowej $X \sim B(n, p)$, wartości oczekiwanej $\mu = np$ oraz $0 < \delta < 1$, zachodzi:

$$P(|X - \mu| \geq \delta\mu) \leq 2 \cdot e^{-\frac{\mu\delta^2}{3}}.$$

Oznacza to, że:

$$P((1-\delta)\mu \leq X \leq (1+\delta)\mu) \geq 1 - 2 \cdot e^{-\frac{\mu\delta^2}{3}}.$$

Zastosowanie do trzech przypadków

Dla każdego przedziału wyznaczamy:

$$\mu = np, \quad \delta = \frac{b-\mu}{\mu}.$$

- **Dla** $X_{30} \sim B(30, \frac{1}{3})$:
  $\mu = 10, \quad b = 12 \Rightarrow \delta = \frac{2}{10} = 0.2$

  $$P(8 \leq X_{30} \leq 12) \geq 1-2e^{-\frac{10\cdot 0.2^2}{3}} = 1-2e^{-0.133} \approx 1-2(0.875) = -0.75 \quad \text{(brak sensu)}.$$

- **Dla** $X_{150} \sim B(150, \frac{1}{3})$:
  $\mu = 50, \quad b = 60 \Rightarrow \delta = \frac{10}{50} = 0.2$

  $$P(40 \leq X_{150} \leq 60) \geq 1-2e^{-\frac{50\cdot 0.2^2}{3}} = 1-2e^{-0.666} \approx 1-2(0.513) = -0.026 \quad \text{(również brak sensu)}.$$

- **Dla** $X_{600} \sim B(600, \frac{1}{3})$:
  $\mu = 200, \quad b = 240 \Rightarrow \delta = \frac{40}{200} = 0.2$

  $$P(160 \leq X_{600} \leq 240) \geq 1-2e^{-\frac{200\cdot 0.2^2}{3}} = 1-2e^{-2.666} \approx 1-2(0.069) = 0.862.$$

Wnioski

Nierówność Chernoffa daje użyteczne dolne oszacowania dla prawdopodobieństw przebywania w przedziale symetrycznym wokół średniej, ale tylko wtedy, gdy wartość oczekiwana μ jest dostatecznie duża.

Dla X_{30} i X_{150} wyniki są ujemne, czyli nie dają sensu probabilistycznego. Dla X_{600} uzyskane oszacowanie $\geq 0{,}862$ jest już poprawne i użyteczne.

W porównaniu z nierównością Czebyszewa, Chernoff zapewnia silniejsze ograniczenia, ale jego stosowanie ma sens dopiero przy dużych n.