RPIS zadanie domowe 3

Kornel Orawczak 346010

15 maja 2025

1 Obliczanie ppb wprost

Zauważmy, że skoro wszystkie X_i są o rozkładzie B(n, $\frac{1}{3}$) to P(X = k) = $\binom{n}{k}p^k(1-p)^{n-k}$, zatem P(a <= X <= b) = $\sum_{k=a}^{b}\binom{n}{k}p^k(1-p)^{n-k}$.
Zatem dla:

a) X_{30} o rozkładzie B(30, $\frac{1}{3}$)
$P(8 \leqslant X_{30} \leqslant 12) = \sum_{k=8}^{12}\binom{30}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{30-k} \approx 0{,}6672$

b) X_{150} o rozkładzie B(150, $\frac{1}{3}$)
$P(40 \leqslant X_{150} \leqslant 60) = \sum_{k=40}^{60}\binom{150}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{150-k} \approx 0{,}9315$

c) X_{600} o rozkładzie B(600, $\frac{1}{3}$)
$P(160 \leqslant X_{600} \leqslant 240) = \sum_{k=160}^{240}\binom{600}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{600-k} \approx 0{,}9996$

2 Przybliżanie tych ppb za pomocą nierówności Czebyszewa

Dla zmiennej losowej X o wartości oczekiwanej μ oraz wariancji σ^2:
$P(|X-\mu| \geqslant k) \leqslant \frac{\sigma^2}{k^2}$ czyli $P(|X-\mu| < k) \geqslant 1 - \frac{\sigma^2}{k^2}$
Zauważmy, że dla rozkładu dwumianowego: $\mu = np$ oraz $\sigma^2 = np(1-p)$
Zatem dla:

a) X_{30} o rozkładzie B(30, $\frac{1}{3}$)
$\mu = 30 \cdot \frac{1}{3} = 10$ oraz $\sigma^2 = 30 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{20}{3}, \quad \sigma = \sqrt{\frac{20}{3}} \approx 2.582$
$P(8 \leqslant X \leqslant 12) = P(|X-10| \leqslant 2)$
Z nierówności Czebyszewa: $P(|X-\mu| < 2) \geqslant 1 - \frac{\sigma^2}{2^2} = 1 - \frac{20/3}{4} = -\frac{2}{3}$

Łatwo zauważyć, że nie daje nam to żadnej interesującej informacji

b) X_{150} o rozkładzie B(150, $\frac{1}{3}$)

$\mu = 150 \cdot \frac{1}{3} = 50$ oraz $\sigma^2 = 150 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{300}{9} = \frac{100}{3}, \quad \sigma = \sqrt{\frac{100}{3}} \approx 5.774$

$P(40 \leqslant X \leqslant 60) \Rightarrow k = \frac{10}{5.774} \approx 1.732$

Czebyszew: $P(|X-50| \leqslant 10) \geqslant 1 - \frac{1}{(1.732)^2} = 1 - \frac{1}{3} = \frac{2}{3} \approx 0.6667$

c) X_{600} o rozkładzie B(600, $\frac{1}{3}$)

$\mu = 600 \cdot \frac{1}{3} = 200$ oraz $\sigma^2 = 600 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1200}{9} = \frac{400}{3}, \quad \sigma = \sqrt{\frac{400}{3}} \approx 11.547$

$P(160 \leqslant X \leqslant 240) = P(|X-200| \leqslant 40) \Rightarrow k = \frac{40}{11.547} \approx 3.464$

Czebyszew: $P(|X-200| \leqslant 40) \geqslant 1 - \frac{1}{(3.464)^2} = 1 - \frac{1}{12} = \frac{11}{12} \approx 0.9167$

3 Przybliżanie rozkładu dwumianowego przez rozkład normalny

Dla dużych n możemy przybliżać rozkład dwumiany X_i przez N(np, np(1-p)).

Wtedy $P(a \leqslant X \leqslant b) \approx \phi(\frac{b+0.5-np}{\sqrt{np(1-p)}}) - \phi(\frac{a-0.5-np}{\sqrt{np(1-p)}})$,

gdzie $\phi(z)$ to dystrybuanta rozkładu normalnego, a z to znormalizowana zmienna o tym rozkładzie na bazie rozkładu dwumianowego. Offset +-0.5 to tzw. continuity correction, przydatny przy przybliżaniu rozkładu dyskretnego przez ciągły. Zatem dla:

a) X_{30} o rozkładzie B(30, $\frac{1}{3}$), $\mu = 10, \sigma \approx 2.582$

$z_1 = \frac{8-0.5-10}{\sqrt{20/3}} \approx -0.968, \quad z_2 = \frac{12+0.5-10}{\sqrt{20/3}} \approx 0.968$

$P(8 \leqslant X \leqslant 12) \approx \Phi(0.968) - \Phi(-0.968) = 2 \cdot \Phi(0.968) - 1 \approx 2 \cdot 0.833 - 1 = 0.666$

b) X_{150} o rozkładzie B(150, $\frac{1}{3}$), $\mu = 50, \sigma = 5.774$

$z_1 = \frac{40-0.5-50}{5.774} = -1.818, \quad z_2 = \frac{60+0.5-50}{5.774} = 1.818$

$P(40 \leqslant X \leqslant 60) \approx \Phi(1.818) - \Phi(-1.818) \approx 0.0965 = 0.034 = 0.931$

c) X_{600} o rozkładzie B(600, $\frac{1}{3}$), $\mu = 200, \sigma = 11.547$

$z_1 = \frac{160-0.5-200}{11.547} = -3.505, \quad z_2 = \frac{240+0.5-200}{11.547} = 3.505$

$P(160 \leqslant X \leqslant 240) \approx \Phi(3.505) - \Phi(-3.505) \approx 0.99977 - 0.00023 = 0.99954$

4 Przybliżanie ppb za pomocą nierówności Chernoffa

Nierówność Chernoffa:

$P(X \geqslant (1+\delta)\mu) \leqslant exp(-\frac{\delta^2\mu}{2+\delta})$
$P(X \leqslant (1-\delta)\mu) \leqslant exp(-\frac{\delta^2\mu}{2})$.
Wykorzystajmy to dla naszych przykładów:

a) X_{30} o rozkładzie B(30, $\frac{1}{3}$), $\mu = 10$,

$\delta = \frac{10-8}{10} = 0.2 \Rightarrow P(X < 8) \leqslant \exp\left(-\frac{(0.2)^2 \cdot 10}{2}\right) = \exp(-0.2) \approx 0.8187$
$P(X > 12) \leqslant \exp\left(-\frac{(0.2)^2 \cdot 10}{2+0.2}\right) = \exp(-0.1818) \approx 0.8339$

Zatem: $P(8 \leqslant X \leqslant 12) \geqslant 1 - 0.8187 - 0.8339 = -0.6526$ (b. słabe oszacowanie)

b) X_{150} o rozkładzie B(150, $\frac{1}{3}$), $\mu = 50$

$\delta = \frac{50-40}{50} = 0.2 \Rightarrow P(X < 40) \leqslant \exp(-1) \approx 0.3679$

$P(X > 60) \leqslant \exp\left(-\frac{(0.2)^2 \cdot 50}{2.2}\right) = \exp(-0.909) \approx 0.4028$

Zatem: $P(40 \leqslant X \leqslant 60) \geqslant 1 - 0.3679 - 0.4028 = 0.2293$ (słabe oszacowanie)

c) X_{600} o rozkładzie B(600, $\frac{1}{3}$), $\mu = 200$

$\delta = \frac{200-160}{200} = 0.2 \Rightarrow P(X < 160) \leqslant \exp(-4) \approx 0.0183$

$P(X > 240) \leqslant \exp\left(-\frac{(0.2)^2 \cdot 200}{2.2}\right) = \exp(-3.636) \approx 0.0265$

Zatem: $P(160 \leqslant X \leqslant 240) \geqslant 1 - 0.0183 - 0.0265 = 0.9552$ (całkiem dobre oszacowanie)

5 Konkluzje

Na przykładzie prób oszacowania wprost wyliczonych ppb rozkładu dwumianowego (2, 3, 4) możemy ocenić ich skuteczność w tym właśnie przybliżaniu. Próba nr.2 i nr.4 wykazała skuteczność jedynie dla większej ilości obserwacji (dla n = 30, n = 150 była dość bezużyteczna). Natomiast próba nr.3 okazała się bardzo skuteczna dla dowolnej ilości prób, co było do przewidzenia, ponieważ jest ona po prostu translacją rozkładu dwumianowego na znormalizowany rozkład normalny, zachowując dobrą wariancję i wartość oczekiwaną.