Zákon velkých čísel, centrální limitní věta

Lenka Křivánková

142474@mail.muni.cz

Přednáška Statistika I (BKMSTAI)

4. listopad 2016, Brno

Motivace

Při velkém počtu pozorování se ukazuje, že

- **empirické charakteristiky se blíží teoretickým charakteristikám**,
- odhad sledovaných veličin se zpřesňuje přímo úměrně velikosti výběru,
- určité transformované veličiny mají téměř normální rozdělení,
- je možno použít přibližných vzorců pro pozorování, jehož rozdělení není tabelováno.

Konvergence náhodných veličin

Nechť $X_1, X_2, \dots$ je posloupnost náhodných veličin s distribučními funkcemi $\Phi_1(x_1), \Phi_2(x_2), \dots$ a X náhodná veličina s distribuční funkcí $\Phi(x)$. Řekneme, že posloupnost $X_1, X_2, \dots$ konverguje k X

- **jistě**, právě když pro všechna $\omega \in \Omega$ platí

$$\lim_{n \to \infty} X_n(\omega) = X(\omega),$$

- **podle pravděpodobnosti**, právě když pro všechna $\epsilon > 0$ platí

$$\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0,$$

- **v distribuci**, právě když pro všechna $x \in R$ platí

$$\lim_{n \to \infty} \Phi_n(x) = \Phi(x).$$

Čebyševova věta

Nechť $X_1, \dots, X_n$ jsou nekorelované náhodné veličiny, jejichž střední hodnoty splňují vztah

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

a rozptyly jsou shora ohraničené týmž číslem δ. Pak posloupnost aritmetických průměrů

$$\left\{ X_1, \frac{1}{2} \sum_{i=1}^{2} X_i, \dots, \frac{1}{n} \sum_{i=1}^{n} X_i, \dots \right\}$$

konverguje podle pravděpodobnosti k číslu μ.

Bernoulliova věta

Nechť náhodná veličina Y_n udává počet úspěchů v posloupnosti n nezávislých opakovaných pokusů, přičemž úspěch nastává v každém pokusu s pravděpodobností θ, $0 < \theta < 1$. Pak posloupnost relativních četností

$$\{Y_1, Y_2/2, \dots, Y_n/n, \dots\}$$

konverguje podle pravděpodobnosti k pravděpodobnosti úspěchu θ.

Čebyševova nerovnost

Pro jakoukoliv náhodnou veličinu X, která má střední hodnotu $E(X)$ a rozptyl $D(X)$, je pravděpodobnost toho, že absolutní odchylka $|X - E(X)|$ nabude hodnoty menší než libovolné $\epsilon > 0$

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{D(X)}{\epsilon^2}.$$

Této nerovnosti můžeme využít pro odhad uvedené pravděpodobnosti, neznáme-li rozdělení dané náhodné veličiny.

Čebyševova nerovnost – příklad

Víme, že náhodná veličina X má střední hodnotu 3 a rozptyl 4. Máme odhadnout pravděpodobnost, že veličina X nabude hodnoty z intervalu $[-2, 8]$.
Hledáme tedy pravděpodobnost

$$P(-2 < X < 8) = P(|X - E(X)| < 5),$$

která je dle Čebyševovy nerovnosti

$$P(|X - E(X)| < 5) \geq 1 - \frac{4}{25} = 0.84.$$

Lindberg-Lévyova centrální limitní věta

Nechť $X_1, X_2, \ldots$ je posloupnost stochasticky nezávislých náhodných veličin se stejným rozdělením, $E(X_i) = \mu$ a $D(X_i) = \sigma^2$ pro $i = 1, 2, \ldots$. Pak posloupnost standardizovaných součtů

$$\left\{ \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \right\}_{n=1}^{\infty}$$

konverguje v distribuci ke **standardizované normální náhodné veličině**, tj. pro každé $x \in R$ platí

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x \right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Moivre-Laplaceova integrální věta

Nechť $Y_1, Y_2, \dots$ je posloupnost stochasticky nezávislých náhodných veličin, $Y_i \sim Bi(n, \theta)$ pro $i = 1, 2, \dots$. Pak posloupnost standardizovaných náhodných veličin

$$\left\{ \frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \right\}_{n=1}^{\infty}$$

konverguje v distribuci k náhodné veličině $U \sim N(0, 1)$, tj. pro každé $x \in R$ platí

$$\lim_{n\to\infty} P\left(\frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq x \right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Moivre-Laplaceova integrální věta

Na základě této věty se používá přibližného vzorce, který nahrazuje pracný výpočet distribuční funkce binomického rozdělení tabelovanou hodnotou distribuční funkce standardizovaného normálního rozdělení:

$$P(Y_n \leq y) = \sum_{t=0}^{y} \binom{n}{t} (1-\theta)^{n-t} \theta^t =$$

$$= P\left(\frac{Y_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{y - n\theta}{\sqrt{n\theta(1-\theta)}} \right) \approx \Phi\left(\frac{y - n\theta}{\sqrt{n\theta(1-\theta)}} \right).$$

Aproximace se považuje za vyhovující, jsou-li splněny podmínky

$$n\theta(1-\theta) > 9 \quad \text{a} \quad \frac{1}{n+1} < \theta < \frac{n}{n+1}.$$

Moivre-Laplaceova integrální věta – příklad 1

Pravděpodobnost, že určitý typ výrobku má výrobní vadu, je 0.05. Jaká je pravděpodobnost, že ze série 1000 výrobků bude mít výrobní vadu nejvýše 70?
Označíme Y_n náhodnou veličinu, která udává počet vadných výrobků ze série n výrobků. Zřejmě je

$$Y_n \sim Bi(n, 0.05), \quad n\theta(1-\theta) = 1000 \cdot 0.05 \cdot 0.95 = 47.5 > 9$$

$$\text{a} \quad \frac{1}{n+1} = \frac{1}{1001} < 0.05 < \frac{n}{n+1} = \frac{1000}{1001}.$$

Spočteme

$$P(Y_{1000} \leq 70) = P\left(\frac{Y_{1000} - 1000 \cdot 0.05}{\sqrt{1000 \cdot 0.05(1-0.05)}} \leq \frac{70 - 1000 \cdot 0.05}{\sqrt{1000 \cdot 0.05(1-0.05)}}\right) \approx$$

$$\approx \Phi\left(\frac{70 - 1000 \cdot 0.05}{\sqrt{1000 \cdot 0.05(1-0.05)}}\right) = \Phi\left(\frac{20}{\sqrt{47.5}}\right) = \Phi(2.90) = 0.99813.$$

Pravděpodobnost, že výrobek má 1. jakost, je $\theta = 0.9$. Kolik výrobků je třeba zkontrolovat, aby s pravděpodobností aspoň 0.99 bylo zaručeno, že rozdíl relativní četnosti počtu výrobků 1. jakosti a pravděpodobnosti $\theta = 0.9$ byl v absolutní hodnotě menší než 0.03?
Hledáme pravděpodobnost

$$P\left(\left|\frac{X}{n} - 0.9\right| < 0.03\right) \geq 0.99$$

Výpočet:

$$0.99 \leq P(0.87n < X < 0.93n) =$$

$$= P\left(\frac{0.87n - 0.9n}{\sqrt{0.09n}} < \frac{X - 0.9n}{\sqrt{0.09n}} < \frac{0.93n - 0.9n}{\sqrt{0.09n}}\right) =$$

$$= P\left(-0.1\sqrt{n} < \frac{X - 0.9n}{\sqrt{0.09n}} < 0.1\sqrt{n}\right) \approx$$

$$\approx \Phi(0.1\sqrt{n}) - \Phi(-0.1\sqrt{n}) = 2\Phi(0.1\sqrt{n}) - 1$$

Moivre-Laplaceova integrální věta – příklad 2

První a poslední člen této nerovnosti je

$$\begin{aligned} 0.99 &\leq 2\Phi(0.1\sqrt{n}) - 1 \\ 0.995 &\leq \Phi(0.1\sqrt{n}) \qquad\qquad /\Phi^{-1} \\ 2.57583 &\leq 0.1\sqrt{n} \\ 664.76 &\leq n \end{aligned}$$

Abychom dosáhli požadované pravděpodobnosti, musíme zkontrolovat alespoň 665 výrobků.