Pravděpodobnost a statistika, přednáška 23. 3. 2020

Doporučená literatura: Likeš J., Machek J.: Počet pravděpodobosti. SNTL, Praha 1982.

Zákony velkých čísel

Mají zásadní význam pro matematickou statistiku (zejména Glivenkova věta).

Čebyševova nerovnost

Nechť X je náhodná veličina se střední hodnotou $E(X)$ a konečným rozptylem $var(X) > 0$. Potom $\forall\lambda > 1$ platí

$$P\left(|X - E(X)| \geq \lambda\sqrt{var(X)}\right) \leq {}^{1}\!/_{\lambda^2}$$

Poznámka

Čebyševova nerovnost se uplatňuje při odhadech pravděpodobnosti náhodných veličin s neznámým rozdělením („vystačíme" pouze se základními číselnými charakteristikami).

Příklad

Doba potřebná k montáži agregátu má $E(X) = 25$ minut a směrodatnou odchylku $\sigma = 1$ minuta. Jaká je pravděpodobnost, že se skutečná doba montáže bude od střední doby lišit o více než $\pm$10 % střední doby?

Řešení

Jelikož 10 % z $E(X)$ je 2,5 minuty, s ohledem na $\sigma = 1$, dostáváme $\lambda = 2{,}5$. Odtud

$$P(|X - 25| \geq 2{,}5) \leq {}^{1}\!/_{2{,}5^2} = 0{,}16.$$

Bernoulliho zákon velkých čísel

Nechť A je náhodný jev reprezentující výsledek náhodného pokusu, $P(A) > 0$ a $h_A(n)$ označuje relativní četnost jevu A v n nezávislých opakováních náhodného pokusu. Potom

$$\forall\varepsilon > 0 \ \lim_{n\to\infty} P(|h_A(n) - P(A)| < \varepsilon) = 1$$

Poznámky

- Výše uvedený zákon velkých čísel potvrzuje možnost experimentálního odhadu pravděpodobnosti na základě relativních četností - s rostoucím počtem pokusů je odchylka relativní četnosti od pravděpodobnosti velmi nepravděpodobná.
- Dále viz Glivenkova věta (později).

Centrální limitní věta

Zjednodušená interpretace - součet velkého počtu nezávislých (resp. omezeně závislých) náhodných veličin s „libovolným“ rozdělením má za velmi obecných podmínek přibližně normální rozdělení (t.j. normální rozdělení je limitním rozdělením celé řady rozdělení).

Platí - nechť náhodná veličina X má binomické (resp. Poissonovo, hypergeometrické) rozdělení s dostatečně velkým rozptylem ($var(X) > 9$) (pro hypergeometrické rozdělení navíc $^n/_N < 0{,}1$). Potom pro $0 < a < b$ platí

$$P(a \leq X \leq b) \cong \Phi\left(\frac{b + {}^1/_2 - E(X)}{\sqrt{var(X)}}\right) - \Phi\left(\frac{a - {}^1/_2 - E(X)}{\sqrt{var(X)}}\right),$$

kde Φ označuje distribuční funkci rozdělení $N(0,1)$.

Poznámky

- Výše uvedená rozdělení aproximujeme normálním rozdělením s odpovídající střední hodnotou a rozptylem, tj.:

  $Bi(n,p)$ aproximujeme $N(\mu = np, \sigma^2 = np(1-p))$,

  $Po(\lambda)$ aproximujeme $N(\mu = \lambda, \sigma^2 = \lambda)$,

  $HGe(N,M,n)$ aproximujeme $N\left(\mu = n\frac{M}{N}, \sigma^2 = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}\right)$.
- Centrální limitní větu lze v případě binomického rozdělení experimentálně „potvrdit“ na tzv. Galtonově deska.

Platí (Poissonovo rozdělení jako limitní rozdělení binomického)
Nechť $\{p_n\}_{n=0}^{\infty}$ $(0 < p_n < 1)$ je posloupnost taková, že $\lim_{n\to\infty} np_n = \lambda > 0$, potom

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!},\ k \in N.$$

(V praxi se tato aproximace doporučuje pro $p < 0{,}1$ a $n \geq 30$)

Konvoluce, transformace náhodných veličin

Poznámky

- Jsou-li X, Y náhodné veličiny, potom i jejich součet $X + Y$ je zřejmě také náhodná veličina. V případě, kdy X, Y jsou nezávislé, mluvíme o konvoluci náhodných veličin.
  (samozřejmě i $X - Y, XY, {}^X/_Y$ jsou náhodné veličiny)
- Je-li X náhodná veličina, $\varphi(x)$ funkce, potom $\varphi(X)$ je také náhodná veličina. V tomto případě mluvíme o transformaci náhodné veličiny X.

Platí

- Nechť X, Y jsou nezávislé diskrétní náhodné veličiny nabývající hodnot z oboru přirozených čísel. Potom pravděpodobnostní funkce náhodné veličiny $Z = X + Y$ má tvar
  $$P(Z = k) = \sum_{i=0}^{k} P(X = i) \cdot P(Y = k - i).$$
- Nechť X, Y jsou nezávislé spojité náhodné veličiny s hustotou $f(x)$, resp. $g(y)$. Potom hustota náhodné veličiny $Z = X + Y$ má tvar

$$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx = \int_{-\infty}^{\infty} g(y)f(z-y)dy.$$

Poznámky

- Snadno nahlédneme, že:
  - Pokud $X \sim Bi(n,p)$, $Y \sim Bi(m,p)$ jsou nezávislé náhodné veličiny, potom jejich konvoluce $Z = X + Y$ má zřejmě rozdělení $Bi(n+m,p)$.
  - Pokud $X \sim Po(\lambda)$, $Y \sim Po(\mu)$ jsou nezávislé náhodné veličiny, potom jejich konvoluce $Z = X + Y$ má zřejmě rozdělení $Po(\lambda + \mu)$.
- Jak demonstruje níže uvedený příklad, konvoluce tzv. „vyhlazuje" rozdělení.

Příklad

Nechť X_1, X_2, X_3 jsou nezávislé náhodné veličiny s rozdělením $R(0,1)$. Pokud označíme $h_2(z)$ hustotu konvoluce $Z = X_1 + X_2$ a $h_3(z)$ hustotu konvoluce $Z = X_1 + X_2 + X_3$ snadno dostáváme:

$$h_2(z) = \begin{cases} 0, \ (z \leq 0) \ \vee \ (z \geq 2) \\ z, \ 0 < z < 1 \\ 2 - z, \ 1 \leq z < 2 \end{cases}, \text{ resp. } h_3(z) = \begin{cases} 0, \ (z \leq 0) \ \vee \ (z \geq 3) \\ \frac{z^2}{2}, \ 0 < z \leq 1 \\ -z^2 + 3z - \frac{3}{2}, \ 1 < z < 2 \\ \frac{z^2}{2} - 3z + \frac{9}{2}, \ 2 \leq z < 3 \end{cases}.$$

Přesto, že hustoty veličin X_1, X_2, X_3 jsou nespojité v bodech 0 a 1, vidíme, že hustota $h_2(z)$ je již spojitá, ale nemá v bodech 0, 1, 2 derivaci. Hustota $h_3(z)$ je nejen spojitá, ale má všude derivaci.

Platí

Nechť X je spojitá náhodná veličina s distribuční funkcí $F(x)$ a hustotou $f(x)$, dále $\varphi(x)$ je absolutně spojitá ryze monotonní funkce taková, že $\forall x \ \varphi'(x) \neq 0$. Označíme-li $G(y), g(y)$ distribuční funkci, resp. hustotu náhodné veličiny $Y = \varphi(X)$, potom platí

$G(y) = F\left(\varphi^{-1}(y)\right)$ a $g(y) = {f\left(\varphi^{-1}(y)\right)} \Big/ {\left|\varphi'\left(\varphi^{-1}(y)\right)\right|}$ pro $y \in (\inf \varphi(x), \sup \varphi(x))$ a $g(y) = 0$ jinde.

Příklad

Určete distribuční funkci $G(y)$ hustotu $g(y)$, střední hodnotu a rozptyl rozdělení náhodné veličiny $Y = e^X$, kde $X \sim N(\mu, \sigma^2)$.

$$G(y) = \begin{cases} F(\ln y), \ \text{pro } y > 0 \\ 0, \ \text{pro } y \leq 0 \end{cases}$$, kde F je distibuční funkce rozdělení $N(\mu, \sigma^2)$,

$$g(y) = \begin{cases} \frac{1}{\sigma y \sqrt{2\pi}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}, \ \text{pro } y > 0 \\ 0, \ \text{pro } y \leq 0 \end{cases},$$

$$E(Y) = e^{\mu + \sigma^2/2}, \ var(Y) = \left(e^{\sigma^2} - 1\right)e^{2\mu + \sigma^2}.$$

Výše popsané rozdělení náhodné veličiny Y se nazývá log-normální (Galtonovo) rozdělení s parametry μ a σ^2, značíme $LN(\mu, \sigma^2)$.