

9. Centrální limitní věty

V předchozí kapitole jsme studovali konvergenci součtu řady náhodných veličin ke konstantní hodnotě. V této kapitole se budeme snažit navázat na výsledky části 8.2, v níž jsme se zabývali výsledky silného zákona velkých čísel. Důsledkem bylo, že za dosti

obecných podmínek je $\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i \xrightarrow{s.j.} 0$.

Pokud budeme chtít vyšetřit asymptotické chování výše uvedených součtů, zjistíme, že silný zákon velkých čísel nám žádný způsob stanovení takového chování neumožňuje. Nejdříve se pokusíme vyřešit tento problém pro náš případ hodu kostkou.

Příklad 9.1

Budeme se snažit nalézt asymptotické chování funkcí

$$F_n(x) = P\left[\left|B_n - \frac{n}{6}\right| < \frac{x\sqrt{5n}}{6}, \quad \left(\frac{x\sqrt{5n}}{6} = x\sqrt{\text{VAR}(B_n)}\right) \quad (9.1)$$

při limitním přechodu $n \rightarrow \infty$. Zjistíme nyní hodnotu F_{6n} pro $x > 0$ (důvodem je lepší výpočet, neboť hodnota $\frac{n}{6}$ bude celé číslo):

$$F_{6n}(x) = \sum_{\left\{i: |i-n| < x\sqrt{\frac{5n}{6}}\right\}} \binom{6n}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{6n-i} = \sum_{\left\{i: |i| < x\sqrt{\frac{5n}{6}}\right\}} \binom{6n}{i+n} \left(\frac{1}{6}\right)^{n+i} \left(\frac{5}{6}\right)^{5n-i}$$

výrazy stojící v součtu se budeme snažit napsat jako součin konstantní hodnoty a výrazu závislého na j a n tedy, protože je

$$\binom{6n}{n+j} \left(\frac{1}{6}\right)^{n+j} \left(\frac{5}{6}\right)^{5n-j} = \left(\binom{6n}{n} \left(\frac{1}{6}\right)^n \left(\frac{5}{6}\right)^{5n}\right) \cdot \left(\frac{1}{5}\right)^j \cdot \frac{5n(5n-1)\dots(5n-j+1)}{(n+1)(n+2)\dots(n+j)} = p_n \cdot R_{j,n}.$$

V dalším se budeme nejdříve snažit odhadnout výraz $R_{j,n}$. Výraz p_n odhadneme pomocí Stirlingovy formule-

$$R_{j,n} = \left[\left(1 + \frac{j}{n}\right) \cdot \left(1 + \frac{j - \frac{4}{5}}{n - \frac{1}{5}}\right) \dots \left(1 + \frac{j - \frac{k \cdot 4}{5}}{n - \frac{k}{5}}\right) \dots \left(1 + \frac{j + \frac{4}{5}}{n - \frac{j-1}{5}}\right) \right]^{-1}. \text{ Pro vyjádření použijeme}$$

dále asymptotického rozvoje funkce $\ln(x)$. $\ln(1+x) = x \cdot (1 + \varepsilon(x))$, kde $\varepsilon(x) \rightarrow 0$ když $x \rightarrow 0$. Tedy

$$\ln(R_{j,n}) = - \sum_{i=0}^{j-1} \ln \left(1 + \frac{j - \frac{k \cdot 4}{5}}{n - \frac{k}{5}} \right) = - \sum_{i=0}^{j-1} \frac{j - \frac{k \cdot 4}{5}}{n - \frac{k}{5}} \cdot \left(1 + \varepsilon \left(\frac{j - \frac{k \cdot 4}{5}}{n - \frac{k}{5}} \right) \right) = - \left(1 + \varepsilon_{j,n} \right) \cdot \sum_{i=0}^{j-1} \left(\frac{j - \frac{k \cdot 4}{5}}{n - \frac{k}{5}} \right)$$

V následném kroku se budeme snažit zjistit přibližnou hodnotu $R_{j,n}$.

$$\ln(R_{j,n}) = -\sum_{i=0}^{j-1} \ln \left(1 + \frac{j - \frac{k.4}{5}}{n - \frac{k}{5}} \right) = -\sum_{i=0}^{j-1} \frac{j - \frac{k.4}{5}}{n - \frac{k}{5}} \cdot \left(1 + \varepsilon \left(\frac{j - \frac{k.4}{5}}{n - \frac{k}{5}} \right) \right) = -(1 + \varepsilon_{j,n}) \cdot \sum_{i=0}^{j-1} \frac{j - \frac{k.4}{5}}{n - \frac{k}{5}}$$

$$\ln(R_{j,n}) = -(1 + \varepsilon_{j,n}) \cdot \sum_{i=0}^{j-1} \left(\frac{j - \frac{k.4}{5}}{n - \frac{k}{5}} \right) = -(1 + \alpha_{j,n}) \cdot \left(\frac{j}{n} \right) \cdot \sum_{i=0}^{j-1} \left(\frac{1 - \frac{k.4}{5.j}}{1 - \frac{k}{5.n}} \right). \text{ Protože opět } 1 - \frac{k}{5.n} \rightarrow 1.$$

můžeme dále psát

$$\ln(R_{j,n}) = -(1 + \alpha_{j,n}) \cdot \left(\frac{j}{n} \right) \cdot \sum_{i=0}^{j-1} \left(\frac{1 - \frac{k.4}{5.j}}{1 - \frac{k}{5.n}} \right) = -(1 + \beta_{j,n}) \cdot \left(\frac{j}{n} \right) \cdot \sum_{i=0}^{j-1} \left(1 - \frac{k.4}{5.j} \right) = -(1 + \beta_{j,n}) \cdot \left(\frac{3.j^2 + 2.j}{5.n} \right)$$

$$R_{j,n} = (1 + \mathcal{G}_{j,n}) \cdot e^{-\frac{3.j^2 + 2.j}{5.n}}.$$

Spočteme – li p_n zjistíme (využíváme Stirlingovu formuli (8.10)):

$$p_n = \binom{6n}{n} \cdot \left(\frac{1}{6} \right)^n \cdot \left(\frac{5}{6} \right)^{5n} \sim \sqrt{\frac{3}{5.\pi.n}}, \text{ celkem po dosazení získáváme:}$$

$$F_{6n}(x) = (1 + \mathcal{G}_n) \cdot \sum_{\left\{ i: |i| < x \cdot \sqrt{\frac{5n}{6}} \right\}} e^{-\frac{3.j^2 + 2.j}{5.n}}, \text{ zavedeme nyní substituci } t_j = j \cdot \sqrt{\frac{6}{5.n}}, \text{ potom je}$$

$$F_{6n}(x) = (1 + \mathcal{G}_n) \cdot \frac{1}{\sqrt{2.\pi}} \cdot \sum_{|t_j| < x} e^{-\frac{t_j^2}{2}} \cdot (t_{j+1} - t_j), \text{ po provedení limitního přechodu}$$

v integrálních součtech máme celkem

$$\lim_{n \rightarrow \infty} F_{6n}(x) = \frac{1}{\sqrt{2.\pi}} \cdot \int_{-x}^x e^{-\frac{t^2}{2}} dt = \Phi(x) - \Phi(-x). \text{ Podle výpočtu je zřejmé, že limitní}$$

distribuční funkce je distribuční funkcí $\mathbf{N(0,1)}$. Tento výsledek se budeme snažit dokázat v další větě.

Věta 9.2

Centrální limitní věta pro stejně rozdělené náhodné veličiny

Nechť $\{X_n\}$ je posloupnost nezávislých stejně rozdělených náhodných veličin se střední hodnotou $E(X_n) = \mu$ a rozptylem $\text{VAR}(X_n) = \sigma^2$. Označme dále jako dříve $S_n = \sum_{i=1}^n X_i$.

Potom pro všechny $x \in \mathbb{R}_1$ platí

$$\lim_{n \rightarrow \infty} P \left(\frac{S_n - n.\mu}{\sigma.\sqrt{n}} < x \right) = \frac{1}{2.\pi} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x) \quad (9.2)$$

Důkaz:

Důkaz provedeme pomocí aparátu charakteristických funkcí z kapitoly . Symbolem H_n označíme náhodnou veličinu $H_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Je zřejmé, že takovéto náhodné veličiny mají střední hodnotu nula a rozptyl rovný jedné. Důkaz bude dokončen, jestliže dokážeme, že posloupnost charakteristických funkcí náhodných veličin H_n bude konvergovat k charakteristické funkci $N(0,1)$. Tedy

$$\lim_{n \rightarrow \infty} \varphi_{H_n}(t) = e^{-\frac{t^2}{2}}.$$

Označme dále G_n náhodnou veličinu $G_n = X_n - \mu$, charakteristická funkce této náhodné veličiny bude označena φ_{G_n} . Podle předpokladu jsou náhodné veličiny X_n nezávislé, jsou tedy nezávislé i náhodné veličiny G_n , navíc jsou stejně rozdělené. Podle věty 6.9 a věty 6.10 je možno určit charakteristickou funkci náhodné veličiny H_n takto:

$$\varphi_{H_n}(t) = \prod_{i=1}^n \varphi_{G_i}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(\varphi_{G_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \quad (9.3)$$

Pro výpočet charakteristické funkce G_1 použijeme větu 6.17. Ověříme nejdříve její předpoklady: $E(G_1) = 0$, $VAR(G_1) = \sigma^2$. Tedy podle věty 6.17 můžeme hodnotu φ_{G_1} psát:

$$\varphi_{G_1}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2 \cdot \sigma^2}{2 \cdot \sigma^2 \cdot n} + r\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{-\frac{t^2}{2} + n \cdot r\left(\frac{t}{\sigma\sqrt{n}}\right)}{n}, \quad (9.4)$$

dosadíme – li hodnotu (9.4) do vztahu (9.3) máme celkem

$$\varphi_{G_n}(t) = \left(1 + \frac{-\frac{t^2}{2} + n \cdot r\left(\frac{t}{\sigma\sqrt{n}}\right)}{n}\right)^n. \quad (9.5)$$

Zřejmě limita výrazu $-\frac{t^2}{2} + n \cdot r\left(\frac{t}{\sigma\sqrt{n}}\right)$ pro t pevné je rovna $-\frac{t^2}{2}$. Použijeme – li dále

klasickou limitu $\lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^{\lim_{n \rightarrow \infty} a_n}$, za předpokladů konvergence posloupnosti $\{a_n\}$.

Závěrem je tedy $\lim_{n \rightarrow \infty} \varphi_{H_n}(t) = e^{-\frac{t^2}{2}}$. Tím jsme prokázali, že posloupnost distribučních funkcí náhodných veličin H_n konverguje k distribuční funkci $N(0,1)$. Protože je výsledná distribuční funkce Φ spojitá ve všech reálných číslech, platí konvergence pro libovolné reálné číslo.

Q.E.D.

Poznámka 9.3

Jestliže zvolíme za výše uvedené náhodné veličiny nám známé alternativní rozdělení s parametrem p , získáme náhodnou veličinu S_n , která je typu binomické rozdělení s parametry n a p . Předchozí věta ukazuje, že distribuční funkce znormované hodnoty S_n konverguje k distribuční funkci $N(0,1)$. Této věty lze potom využít pro případ přibližného výpočtu hodnot $P(S_n)$ při vysokých n .

Příklad 9.4

Nechť $n=1000$ a $p=0,6$. Naším úkolem je zjistit $P(|S_n - 700| < 50)$!

Řešení:

Podle věty 9.2 je v našem případě $\lim_{n \rightarrow \infty} P\left(\frac{S_n - n \cdot 0,7}{\sqrt{0,7 \cdot 0,3 \cdot n}} < x\right) = \Phi(x)$. Přibližný výpočet

provedeme tak, že $P\left(\frac{S_n - n \cdot 0,7}{\sqrt{0,7 \cdot 0,3 \cdot n}} < x\right) \sim \Phi(x)$, v tomto výrazu zvolíme za $n=1000$, tedy:

$$P\left(\frac{S_{1000} - 700}{\sqrt{210}} < x\right) \sim \Phi(x) \Leftrightarrow P\left(\frac{S_{1000} - 700}{\sqrt{210}} < x\right) = P(S_{1000} < 700 + x \cdot \sqrt{210}), \text{ podobně}$$
$$P\left(\frac{S_{1000} - 700}{\sqrt{210}} < -x\right) \sim \Phi(-x) \Leftrightarrow P\left(\frac{S_{1000} - 700}{\sqrt{210}} < -x\right) = P(S_{1000} < 700 - x \cdot \sqrt{210}) \sim \Phi(-x).$$

Zřejmě budeme počítat pravděpodobnost

$$P(650 < S_n < 750) = P(-50 < S_n - 700 < 50) = P\left(-\frac{50}{\sqrt{210}} < \frac{S_n - 700}{\sqrt{210}} < \frac{50}{\sqrt{210}}\right) \sim$$
$$\sim \Phi\left(\frac{50}{\sqrt{210}}\right) - \Phi\left(-\frac{50}{\sqrt{210}}\right) = 2 \cdot \Phi\left(\frac{50}{\sqrt{210}}\right) - 1 = 2 \cdot \Phi(3,4503) - 1 = 2 \cdot 0,99972 - 1 = 0,99944.$$

Hledaná pravděpodobnost je přibližně 99,944%.

Příklad 9.5

Nechť $n=1000$ a $p=0,6$. Naším úkolem je nalézt hodnotu a takovou, aby $P(|S_n - 700| < a) = 95\%$!

Řešení:

Provedeme rychleji, tedy podobně jako v předchozím případě podle věty 9.2 máme:

$$P\left(\left|\frac{S_n - n \cdot 0,7}{\sqrt{0,7 \cdot 0,3 \cdot n}}\right| < x\right) \sim \Phi(x) - \Phi(-x) = 2 \cdot \Phi(x) - 1 \quad (9.6)$$

Do tohoto výrazu budeme dosazovat za $n=1000$ a budeme hledat x tak, aby pravá strana byla rovna 0,95. Nejdříve řešíme rovnici $2 \cdot \Phi(x) - 1 = 0,95 \Leftrightarrow \Phi(x) = 0,975$. Nalezneme hodnotu 97,5% kvantilu normálního rozdělení, jak víme je to číslo 1,96. Tedy toto číslo nyní dosadíme do vztahu (9.6). Číslo a nyní určíme jako součin čísla x a $\sqrt{0,7 \cdot 0,3 \cdot n}$. Tedy $a = 1,96 \cdot \sqrt{210} = 28,4$.

Praktická zkušenost s binomickým rozdělením umožnila zkonstruovat dvě věty, které jsou verzi věty 9.2.

Věta 9.6

Lokální Moivre – Laplaceova věta

Nechť $\{S_n\}$ je posloupnost nezávislých náhodných veličin typu binomického rozdělení s parametry (n, p) . Označme $x_{n,k} = \frac{k - n \cdot p}{\sqrt{n \cdot p \cdot q}}$, $n \geq 1$, $k = 0, 1, \dots, n$.

Nechť $\lim_{n \rightarrow \infty} \sqrt{n \cdot p \cdot q} \rightarrow +\infty$ a nechť dále je $|x_{n,k}| \leq A$, kde $0 < A < +\infty$. Potom

$$\lim_{n \rightarrow \infty} \frac{P(S = k)}{\frac{1}{\sqrt{2\pi n p q}} \cdot e^{-\frac{x_{n,k}^2}{2}}} = 1$$

stejněměrně na kruhu se středem v počátku souřadnic a s poloměrem A.

Věta 9.7

Integrální Moivre – Laplaceova věta

Nechť jsou splněny předpoklady věty 9.6. Nechť dále $a \leq b$ jsou reálná čísla. Potom

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} < b\right) = \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-\frac{t^2}{2}} dt = \Phi(b) - \Phi(a). \quad (9.7)$$

Důkaz :

Podle zadání jsou splněny předpoklady věty 9.2, tedy vztah (9.7) je pravdivý.

Q.E.D.

Věta 9.6 nám dává dostatečně rychlou aproximaci binomického rozdělení za předpokladu, že hodnota parametru p je blízká hodnotě $\frac{1}{2}$, jestliže je hodnota p blízká buď jedné nebo nule použijeme následující tvrzení, které náš problém s rychlostí aproximace řeší.

Věta 9.8

Poissonova

Nechť $\{S_n\}$ je posloupnost nezávislých náhodných veličin s binomickým rozdělením s parametry n, p_n . Nechť dále je $\lim_{n \rightarrow \infty} E(S_n) = \lim_{n \rightarrow \infty} n \cdot p_n = \lambda$, kde $\lambda > 0$. Potom pro $k=0,1,\dots$

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}. \quad (9.8)$$

Důkaz:

Nejdříve ze vztahu $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda$, vyplývá, že hodnotu parametru p_n můžeme asymptoticky zapsat takto :

$$p_n = \frac{\lambda}{n} + \alpha_n, \text{ kde } \lim_{n \rightarrow \infty} \alpha_n = 0,$$

jistě platí :

$$P(S_n = k) = \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = \frac{n!}{(n-k)! \cdot k!} \cdot p_n^k \cdot (1 - p_n)^{n-k}, \quad (9.9)$$

tyto výrazy uspořádáme tak, abychom mohli provést limitu $n \rightarrow \infty$. Nejdříve tedy upravíme p_n^k :

$$p_n^k = \left(\frac{\lambda}{n} + \alpha_n\right)^k = \frac{\lambda^k}{n^k} + \beta_n,$$

stejně tak upravíme $(1 - p_n)^{n-k}$:

$$(1 - p_n)^{n-k} = \left(1 - \frac{\lambda}{n} - \alpha_n\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^{n-k} + \gamma_n.$$

Funkce β_n, γ_n mají stejné vlastnosti jako α_n . Nyní již můžeme přepsat výraz (9.9):

$$\left(\frac{n!}{(n-k)! \cdot n^k} \right) \cdot \left(\frac{\lambda^k}{k!} \right) \cdot \left(\left(1 - \frac{\lambda}{n} \right)^{n-k} \right) + \delta_n \quad (9.10)$$

Nyní provedeme limitu tohoto výrazu . První část má evidentně limitu 1 , druhá část na n nezávisí tedy její limita je $\left(\frac{\lambda^k}{k!} \right)$ a konečně limita poslední třetí části je rovna $e^{-\lambda}$ (připomínáme, že číslo k je pevné přirozené číslo).

Q.E.D.