9. cvičení z PSI

28. listopadu - 2. prosince 2016

9.1 (centrální limitní věta, Čebyševova nerovnost - alternativní rozdělení)

Házíme pravidelnou hrací kostkou a počítáme výskyt šestek. Určete, kolik musíme provést hodů, aby se relativní výskyt šestek (tj. poměr počtu šestek ku počtu hodů) lišil od $\frac{1}{6}$ nejvýše o $\varepsilon = 0.05$ s pravděpodobností alespoň $p = 0.95$. Použijte

(a) centrální limitní větu.

(b) Čebyševovu nerovnost.

Řešení:

Připomeneme si, co říká centrální limitní věta. Pro náhodnou veličinu X s konečným rozptylem, položme

$$\mathrm{norm}(X) := \frac{\pi(X)}{\|\pi(X)\|} = \frac{X - E(X)}{\sqrt{D(X)}},$$

kde opět $\pi(X) = X - E(X)$ je projekce na prostor veličin s nulovou střední hodnotou. Speciálně tedy vidíme, že $\|\mathrm{norm}(X)\| = 1$.

Centrální limitní věta (CLV): Nechť $X_i : \Omega \to \mathbb{R}$ pro $i \in \mathbb{N}$ je posloupnost nezávislých náhodných veličin, které mají stejné rozdělení se střední hodnotou μ a (konečným) rozptylem σ^2. Položme

$$Y_n = \sum_{i=1}^{n} X_i$$

a

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{Y_n}{n}\ .$$

Pak

$$\mathrm{norm}(Y_n) = \frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} = \mathrm{norm}(\overline{X}_n)$$

a

$$\lim_{n\to\infty} F_{\mathrm{norm}(Y_n)}(t) = \lim_{n\to\infty} P\Big(\mathrm{norm}(Y_n) \le t\Big) = \Phi(t)$$

pro každé $t \in \mathbb{R}$ (kde Φ je distribuční funkce normovaného normálního rozdělení, tj. pro $N(0,1)$).

Rychlost konvergence v CLV: Pokud pro veličiny X_i v CLV navíc ještě je $\rho := E(|X_i - \mu|^3) < \infty$, pak platí Berry–Esseenův odhad (pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

$$\Big|F_{\mathrm{norm}(Y_n)}(t) - \Phi(t)\Big| \le C_1 \cdot \frac{\rho}{\sigma^3\sqrt{n}}$$

kde C_1 je nějaké konstanta. Nejlepší současný odhad pro C_1 zatím je, že $C_1 < 0.4748$.

Kromě toho platí ještě odhad (opět pro všechna $t \in \mathbb{R}$ a $n \in \mathbb{N}$):

$$\Big|F_{\mathrm{norm}(Y_n)}(t) - \Phi(t)\Big| \le \frac{C_2}{(1+|t|)^3} \cdot \frac{\rho}{\sigma^3\sqrt{n}}$$

kde C_2 je konstanta. Její současný odhad je $C_2 < 30.84$.

Odhad chyby v CLV pro alternativní rozdělení: Pokud mají veličiny X_i alternativní rozdělení s parametrem p, tj. $P(X_i = 1) = p$, pak

$$\mu = E(X_i) = p, \qquad \sigma = \sqrt{D(X_i)} = \sqrt{p(1-p)}$$

$$\rho = E(|X_i - p|^3) = p(1-p)\big(p^2 + (1-p)^2\big) = \sigma^2\big(p^2 + (1-p)^2\big)$$

čímž dostáváme odhady

$$\Big|F_{\mathrm{norm}(Y_n)}(t) - \Phi(t)\Big| \ \le\ C_1 \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}} \ <\ 0.4748 \cdot \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$$

a

$$\left|F_{\mathrm{norm}(Y_n)}(t) - \Phi(t)\right| \;\leq\; \frac{C_2}{(1+|t|)^3} \cdot \frac{p^2+(1-p)^2}{\sqrt{np(1-p)}} \;<\; \frac{30.84}{(1+|t|)^3} \cdot \frac{p^2+(1-p)^2}{\sqrt{np(1-p)}} \ .$$

Druhý odhad je tedy lepší než první pro $|t| > 3.0198$.

Pěkně zpracováno je to v bakalářské práci:
Rastislav Rehák: Rýchlosť konvergencie v centrálnej limitnej vete
`http://is.muni.cz/th/394214/prif_b/BakalarskaPraca.pdf`

Teď tedy vyřešíme zadaný příklad. Pro $i \in \mathbb{N}$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \text{ při } i\text{-tém hodu padla šestka,} \\ \\ 0 & , \text{ při } i\text{-tém hodu nepadla šestka.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělení s parametrem $p = \frac{1}{6}$ (protože $P(X_i = 1) = p$), střední hodnotou $E(X_i) = p = \frac{1}{6}$ a rozptylem $D(X_i) = p(1-p) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$.
Relativní výskyt šestek při n hodech se pak vyjádří jako

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

Zajímá nás teď nejmenší n tak, aby

$$P\left(\left|\overline{X}_n - \frac{1}{6}\right| \leq \varepsilon\right) \geq p = 0.95 \ ,$$

kde $\varepsilon = 0.05$.

(a) Centrální limitní větu použijeme na normovanou veličinu $\mathrm{norm}(\overline{X}_n) = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{D(\overline{X}_n)}}$. K tomu potřebujeme znát:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = p = \frac{1}{6}$$

$$D(X) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{p(1-p)}{n} = \frac{5}{36n}$$

Takže

$$\mathrm{norm}(\overline{X}_n) = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{D(\overline{X}_n)}} = \frac{6\sqrt{n}}{\sqrt{5}}\Big(\overline{X}_n - \frac{1}{6}\Big)$$

a (pomocí úprav nerovností) můžeme psát :

$$0.95 \leq P\left(\left|\overline{X}_n - \frac{1}{6}\right| \leq 0.05\right) = P\left(\left|\frac{6\sqrt{n}}{\sqrt{5}}\Big(\overline{X}_n - \frac{1}{6}\Big)\right| \leq 0.05 \cdot \frac{6\sqrt{n}}{\sqrt{5}}\right) =$$

$$= P\left(-\frac{0.3}{\sqrt{5}}\sqrt{n} \leq \mathrm{norm}(\overline{X}_n) \leq \frac{0.3}{\sqrt{5}}\sqrt{n}\right) \doteq$$

$$\doteq \Phi\Big(\frac{0.3}{\sqrt{5}}\sqrt{n}\Big) - \Phi\Big(-\frac{0.3}{\sqrt{5}}\sqrt{n}\Big) = 2\Phi\Big(\frac{0.3}{\sqrt{5}}\sqrt{n}\Big) - 1$$

tedy

$$\Phi\Big(\frac{0.3}{\sqrt{5}}\sqrt{n}\Big) \geq \frac{1+0.95}{2} = 0.975$$

$$n \geq \left(\frac{\sqrt{5}}{0.3} \cdot \Phi^{-1}(0.975)\right)^2 \doteq \left(\frac{\sqrt{5}}{0.3} \cdot 1.96\right)^2 \doteq 213.42$$

Musíme tedy provést alespoň $n = 214$ hodů.

(b) Použijeme už upravených výrazů. Z Čebyševovy nerovnosti máme:

$$P\left(\left|\overline{X}_n - \frac{1}{6}\right| \leq 0.05\right) = P\left(\left|\mathrm{norm}(\overline{X}_n)\right| \leq \frac{0.3}{\sqrt{5}}\sqrt{n}\right) \geq 1 - \frac{1}{\left(\frac{0.3}{\sqrt{5}}\sqrt{n}\right)^2} = 1 - \frac{500}{9n} \ .$$

Pokud bude

$$1 - \frac{500}{9n} \geq 0.95$$

neboli

$$n \geq \frac{500}{9(1-0.95)} \doteq 1111.11 \ ,$$

pak máme určitě podmínku $P\left(\left|\overline{X}_n - \frac{1}{6}\right| \leq 0.05\right) \geq 0.95$ splněnu. Čebyševova nerovnost nám tak dává odhad, že potřebujeme udělat alespoň $n = 1112$ hodů. To je podstatně více než u centrální limitní věty, ale zato to víme přesně.

Detaily toho, jak velká byla nepřesnost při použití CLV, nechávám čtenáři ...

9.2 (centrální limitní věta, Čebyševova nerovnost - obecné rozdělení)

Výška mužů (určitého věku) je náhodná veličina o střední hodnotě 180 cm a směrodatnou odchylkou 12 cm. Určete pravděpodobnost, že průměrná výška $n = 150$ mužů bude v intervalu 177.5 cm a 182.5 cm

(a) pomocí centrální limitní věty,

(b) pomocí Čebyševovy nerovnosti.

Řešení:

Výšku i-tého muže (pro $i = 1, \ldots, n$) si označíme jako X_i. Veličiny X_i považujeme za nezávislé, se střední hodnotou $E(X_i) = 180$ cm a směrodatnou odchylkou $\sigma_i = \sqrt{D(X_i)} = 10$ cm.

Průměrná výška se pak vyjádří jako

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \ .$$

Zajímá nás $P(177.5 \leq \overline{X}_n \leq 182.5)$. Budeme potřebovat:

$$E(\overline{X}_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = 180$$

a

$$D(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = \frac{12^2}{n} = \frac{144}{150} = 0.96 \ .$$

Takže

$$\mathrm{norm}(\overline{X}_n) = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{D(\overline{X}_n)}} = \frac{\overline{X}_n - 180}{\sqrt{0.96}} \ .$$

(a) Odhad pomocí centrální limitní věty - úpravami nerovností dostaneme:

$$P(177.5 \le \overline{X}_n \le 182.5) = P\Big(\frac{177.5-180}{\sqrt{0.96}} \le \frac{\overline{X}_n - 180}{\sqrt{0.96}} \le \frac{182.5-180}{\sqrt{0.96}}\Big) =$$
$$= P\Big(-\frac{2.5}{\sqrt{0.96}} \le \mathrm{norm}(\overline{X}_n) \le \frac{2.5}{\sqrt{0.96}}\Big) \doteq \Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - \Phi\Big(-\frac{2.5}{\sqrt{0.96}}\Big) =$$
$$= 2 \cdot \Phi\Big(\frac{2.5}{\sqrt{0.96}}\Big) - 1 \doteq 2 \cdot \Phi\Big(2.5516\Big) - 1 \doteq 2 \cdot 0.99464 - 1 = 0.98928$$

Tedy asi 98.93%.

(b) Odhad pomocí Čebyševovy nerovnosti - použijeme tvar, který už máme:

$$P(177.5 \le \overline{X}_n \le 182.5) = P\left(\big|\mathrm{norm}(\overline{X}_n)\big| \le \frac{2.5}{\sqrt{0.96}}\right) \ge 1 - \frac{1}{\left(\frac{2.5}{\sqrt{0.96}}\right)^2} = \frac{5.29}{6.25} = 0.8464$$

Dostáváme tedy spodní odhad 84.64%.

9.3 (centrální limitní věta - obecnější rozdělení)

Hodíme 420 krát pravidelnou šestistěnnou hrací kostkou a výsledky hodů sčítáme. Pomocí centrální limitní věty odhadněte pravděpodobnost, že součet bude ležet mezi čísly 1400 a 1550.

Řešení:

Pro $i = 1, \ldots, n$ (kde $n = 420$) si zavedeme diskrétní veličiny

$$X_i = \text{"}\textit{hodnota, která padne na kostce při i-tém hodu}\text{"}$$

Veličiny X_i považujeme za nezávislé. Střední hodnota a rozptyl jsou

$$E(X_i) = 3.5$$

$$D(X_i) = \frac{35}{12}\ .$$

Součet hodnot se vyjádří jako

$$X = \sum_{i=1}^{n} X_i\ .$$

Máme teď určit $P(1400 \le X \le 1550)$. To uděláme za pomoci centrální limitní věty použité na normovanou veličinu $\mathrm{norm}(X) = \frac{X - E(X)}{\sqrt{D(X)}}$. K tomu potřebujeme:

$$E(X) = \sum_{i=1}^{n} E(X_i) = 420 \cdot 3.5 = 1470$$

a

$$D(X) = \sum_{i=1}^{n} D(X_i) = 420 \cdot \frac{35}{12} = 1225 = (35)^2\ .$$

Takže

$$\mathrm{norm}(X) = \frac{X - E(X)}{\sqrt{D(X)}} = \frac{X - 1470}{35}$$

a můžeme psát (pomocí úprav nerovností):

$$P(1400 \le X \le 1550) = P\Big(\frac{1400-1470}{35} \le \frac{X-1470}{35} \le \frac{1550-1470}{35}\Big) =$$
$$= P\Big(-2 \le \mathrm{norm}(X) \le \frac{16}{7}\Big) \doteq \Phi\Big(\frac{16}{7}\Big) - \Phi(-2) \doteq$$
$$\doteq \Phi(2.2857) - 1 + \Phi(2) \doteq 0.98886 + 0.97725 - 1 = 0.96611 \ .$$

Postup se dá ještě trochu změnit (protože veličina X má za hodnoty pouze přirozená čísla):

$$P(1400 \le X \le 1550) = P(1399 < X \le 1550) = P\Big(\frac{1399-1470}{35} < \frac{X-1470}{35} \le \frac{1550-1470}{35}\Big) =$$
$$= P\Big(-\frac{71}{35} < \mathrm{norm}(X) \le \frac{16}{7}\Big) = F_{\mathrm{norm}(X)}\Big(\frac{16}{7}\Big) - F_{\mathrm{norm}(X)}\Big(-\frac{71}{35}\Big) \doteq$$
$$\doteq \Phi\Big(\frac{16}{7}\Big) - \Phi\Big(-\frac{71}{35}\Big) \doteq$$
$$\doteq \Phi(2.2857) - 1 + \Phi(2.02857) \doteq 0.98886 + 0.97875 - 1 = 0.96761 \ .$$

Předchozí postup dal 96.611% a tento postup zase 96.761%, což je rozdíl 0.15%.

9.4 (centrální limitní věta - alternativní rozdělení)

Na oboru má studovat 600 studentů. Přibližně jen 2/3 z přijatých studentů se pak nakonec zapíše na studium. Kolik studentů by se mělo přijmout, aby počet zapsaných byl co největší, ale aby překročil 600 s pravděpodobností nejvýše 5%? Jaký pak bude průměrný počet zapsaných studentů?

Řešení:

Nechť $n \in \mathbb{N}$ je počet přijatých studentů. Pro $i = 1, \dots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \ i\text{-tý přijatý student se zapíše na studium,} \\ 0 & , \ i\text{-tý přijatý student se nezapíše na studium.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělením, kde $P(X_i = 1) = \frac{2}{3}$.
Počet zapsaných studentů bude veličina

$$Y_n = \sum_{i=1}^{n} X_i \ ,$$

která má binomické rozdělení $\mathrm{Bi}\Big(n, \frac{2}{3}\Big)$. Zajímá nás teď největší n takové, že

$$P\Big(Y_n > 600\Big) \le 0.05 \ .$$

Opět použijeme centrální limitní větu na normovanou veličinu

$$\mathrm{norm}(Y_n) = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - \frac{2}{3}n}{\sqrt{\frac{2}{3} \cdot \frac{1}{3}n}} = \frac{3Y_n - 2n}{\sqrt{2n}} \ .$$

Pomocí úprav nerovností teď můžeme psát :

$$0.05 \ge P\Big(Y_n > 600\Big) = P\left(\frac{3Y_n - 2n}{\sqrt{2n}} > \frac{3 \cdot 600 - 2n}{\sqrt{2n}}\right) =$$

$$= P\left(\text{norm}(Y_n) > \frac{1800-2n}{\sqrt{2n}}\right) = 1 - P\left(\text{norm}(Y_n) \le \frac{1800-2n}{\sqrt{2n}}\right) =$$

$$= 1 - F_{\text{norm}(Y_n)}\left(\frac{1800-2n}{\sqrt{2n}}\right) \doteq 1 - \Phi\left(\frac{1800-2n}{\sqrt{2n}}\right)$$

tedy

$$\Phi\left(\frac{1800-2n}{\sqrt{2n}}\right) \ge 0.95$$

a

$$\left(\sqrt{2n}\right)^2 + \Phi^{-1}(0.95)\cdot\sqrt{2n} - 1800 \le 0\ .$$

Dostali jsme kvadratickou nerovnost a hledáme největší $n \in \mathbb{N}$, které ji splňuje. Z kořenů kvadratické rovnice si tedy vezmeme ten, co je více vpravo. Označme si ještě pro jednoduchost

$$b = \Phi^{-1}(0.95) \doteq 1.645\ .$$

Dostaneme tak

$$\sqrt{2n} \le \frac{-b + \sqrt{b^2 + 4\cdot 1800}}{2}$$

a

$$n \le \frac{\left(\sqrt{b^2+7200} - b\right)^2}{8} \doteq \frac{\left(\sqrt{1.645^2+7200} - 1.645\right)^2}{8} \doteq 865.774$$

Mělo by se tedy přijmout $n = 865$ studentů. Průměrný počet těch, co se zapíší tak bude

$$E(Y_n) = \frac{2}{3}\cdot n = \frac{2}{3}\cdot 865 \doteq 576,67\ .$$

Opět můžeme udělat ještě trochu jiné řešení (díky tomu, že hodnoty veličiny Y_n jsou pouze přirozená čísla):

$$0.05 \ge P\big(Y_n > 600\big) = P\big(Y_n \ge 601\big) = P\left(\text{norm}(Y_n) \ge \frac{3\cdot 601 - 2n}{\sqrt{2n}}\right) =$$

$$= 1 - P\left(\text{norm}(Y_n) < \frac{1803-2n}{\sqrt{2n}}\right) = 1 - F_{\text{norm}(Y_n)}\left(\frac{1803-2n}{\sqrt{2n}}\right)_{-} \doteq$$

$$\doteq 1 - \Phi\left(\frac{1803-2n}{\sqrt{2n}}\right)$$

Obdobně dostaneme

$$n \le \frac{\left(\sqrt{b^2+7212} - b\right)^2}{8} \doteq \frac{\left(\sqrt{1.645^2+7212} - 1.645\right)^2}{8} \doteq 867.245$$

Tentokrát tedy vyjde $n = 867$ studentů.

9.5 (centrální limitní věta - alternativní rozdělení)

V letadle je $j = 216$ míst pro cestující. Asi 5% cestujících se k odletu nedostaví. Kolik letenek lze prodat, aby riziko, že se cestující do letadla nevejdou, bylo nejvýše $\alpha = 0.02$?

Řešení:

Příklad je analogický k předchozímu. Nechť $n \in \mathbb{N}$ je počet cestujících, kteří si koupili letenku. Pro $i = 1, \ldots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & ,\ i\text{-tý cestující se dostaví k odletu,} \\ 0 & ,\ i\text{-tý cestující se nedostaví k odletu.} \end{cases}$$

Veličiny X_i považujeme za nezávislé (i když to ve skutečnosti nemusí být až úplně splněno), s alternativním rozdělením, kde $P(X_i = 1) = 0.95$. Počet cestujících, kteří se dostaví k odletu bude veličina

$$Y_n = \sum_{i=1}^{n} X_i \ ,$$

která má binomické rozdělení $\mathrm{Bi}\Big(n, 0.95\Big)$. Zajímá nás teď největší n takové, že

$$P\Big(Y_n > 216\Big) \leq 0.02 \ .$$

Opět použijeme centrální limitní větu na normovanou veličinu

$$\mathrm{norm}(Y_n) = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - 0.95 \cdot n}{\sqrt{0.95 \cdot 0.05 \cdot n}} \ .$$

Pomocí úprav nerovností opět můžeme psát (využijeme toho, že Y_n má za hodnoty pouze přirozená čísla) :

$$0.02 \geq P\Big(Y_n > 216\Big) = P\left(\mathrm{norm}(Y_n) > \frac{216 - 0.95 \cdot n}{\sqrt{0.95 \cdot 0.05 \cdot n}}\right) =$$
$$= 1 - F_{\mathrm{norm}(Y_n)}\left(\frac{216 - 0.95 \cdot n}{\sqrt{0.95 \cdot 0.05 \cdot n}}\right) \doteq 1 - \Phi\left(\frac{216 - 0.95 \cdot n}{\sqrt{0.95 \cdot 0.05 \cdot n}}\right)$$

tedy

$$\Phi\left(\frac{216 - 0.95 \cdot n}{\sqrt{0.95 \cdot 0.05 \cdot n}}\right) \geq 0.98$$

a

$$\Big(\sqrt{0.95 \cdot n}\Big)^2 + \Big(\sqrt{0.05} \cdot \Phi^{-1}(0.98)\Big) \ \cdot \ \sqrt{0.95 \cdot n} \ - \ 216 \leq 0 \ .$$

Dostali jsme kvadratickou nerovnost a hledáme největší $n \in \mathbb{N}$, které ji splňuje. Označme si ještě pro jednoduchost

$$c = \sqrt{0.05} \cdot \Phi^{-1}(0.98) \doteq \sqrt{0.05} \cdot 2.054 \doteq 0.45929 \ .$$

Dostáváme

$$\sqrt{0.95 \cdot n} \leq \frac{-c + \sqrt{c^2 + 4 \cdot 216}}{2}$$

a

$$n \leq \frac{\Big(\sqrt{c^2 + 864} - c\Big)^2}{4 \cdot 0.95} \doteq \frac{\Big(\sqrt{0.45929^2 + 864} - 0.45929\Big)^2}{4 \cdot 0.95} \doteq 220.37$$

Letenek se tak může prodat $n = 220$.

V tomto příkladu můžeme náš přibližný výsledek dokonce přesně ověřit (protože výsledná hodnota $n = 220$ není příliš vzdálená od 216). Máme totiž

$$P\Big(Y_n > 216\Big) = \sum_{i=217}^{n} \binom{n}{n-i} 0.95^i \cdot 0.05^{n-i}.$$

Pro $n = 220$ je

$$P\Big(Y_n > 216\Big) = 0.95^{217} \cdot \left[\binom{220}{3} 0.05^3 + \binom{220}{2} 0.95 \cdot 0.05^2 + \binom{220}{1} 0.95^2 \cdot 0.05 + 0.95^3\right] \doteq 0.0042 \leq 0.02 \ .$$

Tedy podmínka je splněna. Ale je možné, že existuje ještě vyšší n které je řešením (protože jsem počítali pouze s určitou přesností).

Pro $n = 221$ je

$$P\Big(Y_n > 216\Big) = 0.95^{217} \cdot \left[\binom{221}{4} 0.05^4 + \binom{221}{3} 0.95 \cdot 0.05^3 + \binom{221}{2} 0.95^2 \cdot 0.05^2 + \binom{221}{1} 0.95^3 \cdot 0.05 + 0.95^4\right] \doteq$$
$$\doteq 0.0129 \leq 0.02 \ .$$

Opět je podmínka splněna.

Pro $n = 222$ pravděpodobnost vyjde 0.0321, což už přesahuje 0.02. Skutečný počet je tedy $n = 221$, ale i CLV dává dostatečně dobrý odhad (záleží totiž na tom, jak přesně budeme aproximovat).

9.6 (centrální limitní věta - alternativní rozdělení)

Před volbami je v populaci státu 52% příznivců koaliční strany. Jaká je pravděpodobnost, že průzkum veřejného mínění o rozsahu $n = 1500$ ukáže nesprávně převahu opozice?

Řešení:

Jednotlivý volič volí tedy koalici s pravděpodobností $p = 0.52$. Pro $i = 1, \dots, n$ si zavedeme veličiny

$$X_i = \begin{cases} 1 & , \ i\text{-tý člověk z průzkumu zvolí koalici,} \\ 0 & , \ i\text{-tý člověk z průzkumu zvolí opozici.} \end{cases}$$

Veličiny X_i považujeme za nezávislé, s alternativním rozdělením s parametrem p.

Preference koaliční strany v průzkumu se pak vyjádří jako

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(tj. počet lidí z průzkumu, kteří by volili koalici, ku počtu všech dotázaných). Pravděpodobnost, že se ukáže převaha opozice je

$$P(\overline{X}_n < 0.5)$$

a její hodnotu odhadneme pomocí centrální limitní věty. K tomu si spočítáme

$$D\big(\overline{X}_n\big) = \frac{1}{n^2} \sum_{i=1}^{n} D(X_i) = \frac{p(1-p)}{n} = \frac{0.2496}{1500} = 166.4 \cdot 10^{-6}$$

$$\sqrt{D\big(\overline{X}_n\big)} = \sqrt{166.4} \cdot 10^{-3} \doteq 1.29 \cdot 10^{-2}$$

Použijeme opět normovanou veličinu

$$\mathrm{norm}(\overline{X}_n) = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{D(\overline{X}_n)}} = \frac{\overline{X}_n - 0.52}{\sqrt{166.4}} \cdot 1000 \ .$$

Pomocí úprav nerovností můžeme psát:

$$P(\overline{X}_n < 0.5) = P\left(\frac{\overline{X}_n - 0.52}{\sqrt{166.4}} \cdot 1000 < \frac{0.5 - 0.52}{\sqrt{166.4}} \cdot 1000 \right) = P\left(\mathrm{norm}(\overline{X}_n) < -\frac{20}{\sqrt{166.4}} \right) \doteq$$

$$\doteq \Phi\left(-\frac{20}{\sqrt{166.4}} \right) = 1 - \Phi\left(\frac{20}{\sqrt{166.4}} \right) \doteq 1 - \Phi\Big(1.5504\Big) \doteq 1 - 0.93948 = 0.0652$$

Tedy asi 6.52%.

O něco přesnější postup by byl tento:
Vezmeme veličinu

$$Y_n = \sum_{i=1}^{n} X_i$$

vyjadřující počet lidí z průzkumu, kteří by volili koalici. V průzkumu se ukáže převaha opozice, pokud ta bude mít alespoň o 1 hlas více než koalice, tedy ptáme se na $P(Y_n \leq 749)$. To odpovídá situaci kdy $P(\overline{X}_n \leq \frac{749}{1500})$. Opravený výpočet (s použitím $0.52 = \frac{780}{1500}$) pak dává:

$$P\left(\overline{X}_n \leq \frac{749}{1500} \right) = P\left(\frac{\overline{X}_n - 0.52}{\sqrt{166.4}} \cdot 1000 \leq \frac{749/1500 - 780/1500}{\sqrt{166.4}} \cdot 1000 \right) = P\left(\mathrm{norm}(\overline{X}_n) \leq -\frac{31}{\sqrt{166.4}} \cdot \frac{2}{3} \right) \doteq$$

$$\doteq \Phi\left(-\frac{62}{3\sqrt{166.4}} \right) = 1 - \Phi\left(\frac{62}{3\sqrt{166.4}} \right) \doteq 1 - \Phi\left(1.6021\right) \doteq 1 - 0.94543 = 0.05457$$

tedy asi $5,46\%$, což je skoro o 1% méně než v předchozím postupu.

9.7 (intervalový odhad pomocí centrální limitní věty)

Počet X ryb, které rybář uloví za den je popsán Poissonovým rozdělením s parametrem $\lambda = 3$. Na ryby jde $n = 100$-krát za rok. Najděte (co nejmenší) symetrický interval, v němž se počet ulovených ryb za rok nachází s pravděpodobností alespoň 95%.

Řešení:
Pro $i = 1, \ldots, n$ máme veličiny

$$X_i = \text{"}\textit{počet ryb ulovených za i-tý den}\text{"}$$

které jsou za nezávislé a mají Poissonovo rozdělení s parametrem $\lambda = 3$.
Takže

$$E(X_i) = \lambda = 3 \quad \text{a} \quad D(X_i) = \lambda = 3 \ .$$

Počet ryb ulovených za $n = 100$ dnů je

$$Y_n = \sum_{i=1}^{n} X_i \ .$$

Hledaný symetrický interval $\langle u_1, u_2 \rangle$, do kterého padnou hodnoty Y_n s pravděpodobností alespoň $0.95 = 1 - \alpha$, je určen požadavkem na doplňkovou množinu

$$P(Y_n < u_1) = \frac{\alpha}{2} \quad \text{a} \quad P(Y_n > u_2) = \frac{\alpha}{2} \ .$$

To přepíšeme pro normovanou veličinu $\mathrm{norm}(Y_n) = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - 300}{\sqrt{300}}$ jako

$$P\left(\mathrm{norm}(Y_n) < \frac{u_1 - 300}{10\sqrt{3}}\right) = \frac{\alpha}{2} \quad \text{a} \quad P\left(\mathrm{norm}(Y_n) > \frac{u_2 - 300}{10\sqrt{3}}\right) = \frac{\alpha}{2} \ .$$

Rozdělení veličiny $\mathrm{norm}(Y_n)$ aproximujeme pomocí CLV rozdělením $N(0, 1)$. Takže dostaneme

$$\Phi\left(\frac{u_1 - 300}{10\sqrt{3}}\right) = \frac{\alpha}{2} \quad \text{a} \quad 1 - \Phi\left(\frac{u_2 - 300}{10\sqrt{3}}\right) = \frac{\alpha}{2}$$

neboli

$$\frac{u_1 - 300}{10\sqrt{3}} = \Phi^{-1}\left(\frac{\alpha}{2}\right) = -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \text{a} \quad \frac{u_2 - 300}{10\sqrt{3}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

Výsledný interval (pro $\alpha = 0.05$) je pak

$$\langle u_1, u_2 \rangle = \left\langle 300 \ - \ 10\sqrt{3} \cdot \Phi^{-1}(0.975), \quad 300 \ + \ 10\sqrt{3} \cdot \Phi^{-1}(0.975) \right\rangle \doteq$$

$$\doteq \langle 300 \ - \ 33.948, \quad 300 \ + \ 33.948 \rangle \ .$$

K řešení také můžeme intuitivně přistoupit tak, že veličina $\mathrm{norm}(Y_n) = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - 300}{\sqrt{300}}$ má přibližně rozdělení $N(0, 1)$. Takže zpětně můžeme usoudit, že veličina $Y_n = \sqrt{300} \cdot \mathrm{norm}(Y_n) + 300$ má přibližně rozdělení $N\big(300, 300\big) =: N(\mu, \sigma^2)$. Pro takovouto veličinu má symetrický intervalový odhad tvar

$$\left\langle \mu \ - \ \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \quad \mu \ + \ \sigma \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\rangle \ .$$

9.8 (intervalový odhad střední hodnoty při známém rozptylu)

Předpokládejme, že chyba, se kterou tachometr ukazuje rychlost, má normální rozdělení se směrodatnou odchylkou $\sigma = 1.5\ \frac{\text{km}}{\text{h}}$. Jaká musí být střední hodnota chyby, aby pravděpodobnost, že tachometr ukazuje menší než skutečnou rychlost, byla nejvýše $\alpha = 0.001$?

Řešení:
Máme veličinu

$$X = \text{"}\textit{chyba, se kterou tachometr ukazuje rychlost}\text{"}$$

s rozdělením $N(\mu, \sigma^2)$. Zajímá nás pro jaké hodnoty parametru μ platí, že

$$P(X < 0) \leq \alpha = 0.001\ .$$

Stačí opět přejít k normalizované veličině $\text{norm}(X) = \frac{X-\mu}{\sigma}$ s rozdělením $N(0,1)$:

$$\alpha \geq P\big(X < 0\big) = P\left(\text{norm}(X) < -\frac{\mu}{\sigma}\right) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

neboli

$$\frac{\mu}{\sigma} \geq \Phi^{-1}(1-\alpha)$$

a tedy

$$\mu \geq \sigma \cdot \Phi^{-1}(1-\alpha) = 1.5 \cdot \Phi^{-1}(0.999) \doteq 1.5 \cdot 3.09 = 4.635\ .$$

Tachometr tak musí mít střední hodnotu chyby alespoň $4.635\ \frac{\text{km}}{\text{h}}$.

9.9 (normální rozdělení)
Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě 67 a 75 m a směrodatné odchylky 6 a 3 m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.

Řešení:
Máme veličiny X_A (X_B, resp.), které znamenají délku hodu Anny (Barbory, resp.) a mají rozdělení $N(67, 6^2)$ ($N(75, 3^2)$, resp.).

Zajímá nás hodnota pravděpodobnosti

$$P(X_A > X_B) = P(X_A - X_B > 0) = ?\ .$$

Protože veličiny X_A a X_B jsou nezávislé, má veličina $X_A - X_B$ opět normální rozdělení a sice

$$N(67 - 75, 6^2 + 3^2) = N\big(-8,\ (\sqrt{45})^2\big)\ .$$

Proč to tak je? Protože pro nezávislé veličiny X a Y máme

$$E\big(X + (-Y)\big) = E(X) + (-1) \cdot E(Y)$$
$$D\big(X + (-Y)\big) = D(X) + (-1)^2 \cdot D(Y)\ .$$

Takže

$$P\big(X_A - X_B > 0\big) = P\left(\text{norm}(X_A - X_B) > \frac{8}{3\sqrt{5}}\right) = 1 - \Phi\left(\frac{8}{3\sqrt{5}}\right) \doteq$$

$$\doteq 1 - \Phi(1.1926) \doteq 1 - 0.88349 = 0.11651\ .$$

Šance, že Anna hodí dál než Barbora, je tak asi 11.65%.

9.10 (normální rozdělení)

Rozvodné závody dodávaly elektřinu, jejíž napětí ve voltech mělo normální rozdělení $\mathrm{N}(\mu_1, \sigma_1^2)$, kde $\mu_1 = 230$ a $\sigma_1^2 = 25$. Horní mez U_0 dodávaného napětí je nejnižší mez, která je překročena s pravděpodobností nejvýše $\alpha = 10^{-4}$. Nyní se závodům podařilo snížit rozptyl na $\sigma_2^2 = 10$. O kolik mohou zvýšit střední hodnotu μ_2, aby byla zachována horní mez?

Řešení:
Máme tedy veličinu

$$U_1 = \text{"}\textit{hodnota původního dodávaného napětí}\text{"}$$

s rozdělením $N(\mu_1, \sigma_1^2) = N(230, 25)$ a veličinu

$$U_2 = \text{"}\textit{hodnota nově dodávaného napětí}\text{"}$$

s rozdělením $N(\mu_2, \sigma_2^2) = N(\mu_2, 10)$. Horní mez U_0 je v obou případech nejnižší hranice vzhledem k podmínce

$$P(U_i > U_0) \le \alpha = 10^{-4} \quad \text{pro} \;\; i = 1, 2 \; .$$

Takže

$$\alpha \ge P(U_i > U_0) = P\left(\mathrm{norm}(U_i) > \frac{U_0 - \mu_i}{\sigma_i}\right) = 1 - \Phi\left(\frac{U_0 - \mu_i}{\sigma_i}\right)$$

a U_0 tudíž splňuje rovnosti

$$U_0 = \mu_1 + \sigma_1 \cdot \Phi^{-1}(1 - \alpha)$$
$$U_0 = \mu_2 + \sigma_2 \cdot \Phi^{-1}(1 - \alpha) \; .$$

Jejich odečtením dostaneme

$$\mu_2 - \mu_1 = (\sigma_1 - \sigma_2) \cdot \Phi^{-1}(1 - \alpha) = (5 - \sqrt{10}) \cdot \Phi^{-1}(0.9999) \doteq$$

$$\doteq (5 - \sqrt{10}) \cdot 3.719 \doteq 6.8345 \; .$$

Středního hodnota nově dodávaného napětí tedy může být až o 6.8345 V vyšší oproti původnímu.

Otázkou je, co na to řeknou přístroje, které jsou dimenzovány na 230 V ...

9.11 (normální rozdělení)

Nechť veličina X má normované normální rozdělení $N(0, 1)$. Určete $P(X^2 < 3X - 2)$ a najděte takové číslo ε, že $P(|X| < \varepsilon) = 0.95$.

Řešení:
Máme

$$P(X^2 < 3X - 2) = P(X^2 - 3X + 2 < 0) = P\Big((X - 1)(X - 2) < 0\Big) = P\Big(1 < X < 2\Big) =$$

$$= \Phi(2) - \Phi(1) \doteq 0.97725 - 0.84134 = 0.13591 \; .$$

A dále je

$$0.95 = P(|X| < \varepsilon) = P(-\varepsilon < X < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1$$

a tedy

$$\varepsilon = \Phi^{-1}(\frac{1 + 0.95}{2}) = \Phi^{-1}(0.975) \doteq 1.96 \; .$$

9.12 (metody odhadů parametru)

Náhodná veličina X nabývá hodnot s pravděpodobnostmi dle tabulky, kde c, q jsou reálné parametry rozdělení. Z četností hodnot v náhodném výběru, uvedených v tabulce, odhadněte parametry c a q.

hodnota i	1	2	3
pravděpodobnost $p_X(i)$	$c-q$	c	$c+q$
četnost n_i	8	10	5

Řešení:
Protože součet pravděpodobností všech hodnot je 1, musí být

$$1 = (c-q) + c + (c+q) = 3c$$

tedy $c = \frac{1}{3}$. Současně musí být pravděpodobnosti nezáporné, tj. $0 \le c-q = \frac{1}{3} - q$ a $0 \le c+q = \frac{1}{3} + q$, takže $|q| \le \frac{1}{3}$. Zbývá tedy odhadnout parametr q.

Metoda maximální věrohodnosti:

Hledáme hodnotu q, která maximalizuje funkci věrohodnosti

$$L(q) = P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) = \left(\frac{1}{3} - q\right)^8 \cdot \left(\frac{1}{3}\right)^{10} \cdot \left(\frac{1}{3} + q\right)^5$$

kde X_i jsou jednotlivé nezávislé veličiny (pokusy) a x_i naměřené hodnoty. Funkce L je nezáporná a spojitá na uzavřené množině $\left\langle -\frac{1}{3}, \frac{1}{3} \right\rangle$, takže zde nabývá maxima. To nemůže být v krajních bodech a proto je nabyto uvnitř dané množiny. To odpovídá hledání maxima funkce

$$\ell(q) = \ln\big(L(q)\big) = 8 \cdot \ln\left(\frac{1}{3} - q\right) + 10 \cdot \ln\frac{1}{3} + 5 \cdot \ln\left(\frac{1}{3} + q\right)$$

na intervalu $(-\frac{1}{3}, \frac{1}{3})$. Protože maximum existuje, musí pro něj platit

$$0 = \ell'(q) = \frac{-8}{\frac{1}{3} - q} + \frac{5}{\frac{1}{3} + q}$$

Odhad parametru q je

$$q = -\frac{1}{13} \doteq -0.07692\ .$$

Odhady pravděpodobností hodnot $1, 2, 3$ jsou tedy

$$p_X(1) = \frac{16}{39} \doteq 0.4103 \qquad p_X(2) = \frac{1}{3} \doteq 0.3333 \qquad p_X(3) = \frac{10}{39} \doteq 0.2564$$

což vyhovuje zadání.

Metoda momentů:

Střední hodnota je

$$E(X) = \left(\frac{1}{3} - q\right) + 2 \cdot \frac{1}{3} + 3 \cdot \left(\frac{1}{3} + q\right) = 2 + 2\,q$$

její odhad z realizace je

$$\overline{x} = \frac{1}{n} \sum_i i\, n_i = \frac{1}{8+10+5} \cdot (1 \cdot 8 + 2 \cdot 10 + 3 \cdot 5) = \frac{43}{23}\ .$$

Porovnáním dostaneme

$$2 + 2q = E(X) = \overline{x} = \frac{43}{23}$$

což odpovídá hodnotě

$$q = -\frac{3}{46} \doteq -0.06522 \ .$$

Odhady pravděpodobností hodnot $1, 2, 3$ jsou tedy

$$p_X(1) = \frac{55}{138} \doteq 0.3986 \qquad p_X(2) = \frac{1}{3} \doteq 0.3333 \qquad p_X(3) = \frac{37}{138} \doteq 0.2681$$

což opět vyhovuje zadání.

9.13 (metody odhadů parametru)

Náhodná veličina X nabývá hodnot $0, 1, 2$. Její rozdělení je závislé na parametrech p a q (viz tabulka). Bylo provedeno 20 pokusu. Četnosti jednotlivých hodnot jsou uvedeny v tabulce:

hodnota	0	1	2
teoretická pravděpodobnost	p	q	q^2
pozorovaná četnost	2	12	6

Odhadněte parametry p a q metodou momentů i metodou maximální věrohodnosti.

Řešení:

Součet pravděpodobností všech hodnot musí být 1, takže

$$p = 1 - q - q^2$$

tedy $c = \frac{1}{3}$. Současně musí být pravděpodobnosti nezáporné (dokonce kladné, protože v pokusu byly dané hodnoty skutečně realizovány), tj. $1 - q - q^2 > 0$ a $q > 0$, takže

$$0 < q < \frac{\sqrt{5} - 1}{2} \doteq 0.618 \ .$$

Zbývá tedy odhadnout parametr q.

Metoda maximální věrohodnosti:

Hledáme hodnotu q, která maximalizuje funkci věrohodnosti

$$L(q) = P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) = \left(1 - q - q^2\right)^2 \cdot q^{12+2\cdot 6}$$

kde X_i jsou jednotlivé nezávislé veličiny (pokusy) a x_i naměřené hodnoty. Funkce L je nezáporná a spojitá na uzavřené množině $\left\langle 0, \frac{\sqrt{5}-1}{2} \right\rangle$, takže zde nabývá maxima. To nemůže být v krajních bodech a proto je nabyto uvnitř dané množiny. To tedy odpovídá hledání maxima funkce

$$\ell(q) = \ln\left(L(q)\right) = 2 \cdot \ln\left(1 - q - q^2\right) + 24 \cdot \ln(q)$$

na intervalu $\left(0, \frac{\sqrt{5}-1}{2}\right)$. Protože maximum existuje, musí pro něj platit

$$0 = \ell'(q) = \frac{-2 - 4q}{1 - q_2 - q_2^2} + \frac{24}{q}$$

neboli

$$14q^2 + 13q - 12 = 0 \ .$$

Rovnice má řešení $q_1 = -\frac{3}{2}$ (které nevyhovuje zadání) a řešení

$$q_2 = \frac{4}{7} \doteq 0.57143 \ ,$$

které leží v požadovaném intervalu $\left(0, \frac{\sqrt{5}-1}{2}\right)$. Odpovídající hodnota druhého parametru je

$$p = 1 - q - q^2 = \frac{5}{49} \doteq 0.10204 \ .$$

Metoda momentů:

Střední hodnota je

$$E(X) = 0 \cdot \left(1 - q - q^2\right) + 1 \cdot q + 2 \cdot q^2 = q + 2q^2$$

a její odhad z realizace je

$$\overline{x} = \frac{1}{n} \sum_i i \, n_i = \frac{1}{2 + 12 + 6} \cdot (0 \cdot 2 + 1 \cdot 12 + 2 \cdot 6) = \frac{6}{5} \ .$$

Porovnáním dostaneme

$$q + 2q^2 = E(X) = \overline{x} = \frac{6}{5}$$

Rovnice má řešení $q_1 = -\frac{\sqrt{265}}{20} - \frac{1}{4}$ (které nevyhovuje zadání) a řešení

$$q_2 = \frac{\sqrt{265}}{20} - \frac{1}{4} \doteq 0.56394,$$

které leží v požadovaném intervalu $\left(0, \frac{\sqrt{5}-1}{2}\right)$. Odpovídající hodnota druhého parametru je

$$p = 1 - q_2 - q_2^2 \doteq 0.0.11803 \ .$$

9.14 (metody odhadů parametru)
Odhadněte parametr w geometrického rozdělení

$$p_i = w^i(1 - w), \quad i \in \mathbb{N}_0$$

na základě realizace s následujícími četnostmi výsledků:

hodnota	0	1	2	3
pozorovaná četnost	20	10	7	3

Použijte metodu momentů i metodu maximální věrohodnosti.

Řešení:

Součet pravděpodobností všech hodnot je 1. Pravděpodobnosti hodnot v pokusu musí být nenulové, protože dané hodnoty byly skutečně realizovány, takže

$$0 < w < 1 \ .$$

Metoda maximální věrohodnosti:

Hledáme hodnotu q, která maximalizuje funkci věrohodnosti

$$L(w) = \prod_{i=1}^{n} P(X_i = x_i) = \Big(1-w\Big)^{20}\Big(w(1-w)\Big)^{10}\Big(w^2(1-w)\Big)^{7}\Big(w^3(1-w)\Big)^{3} =$$

$$= w^{33}(1-w)^{40}$$

kde X_i jsou jednotlivé nezávislé veličiny (pokusy) a x_i naměřené hodnoty. Funkce L je nezáporná a spojitá na uzavřené množině $\langle 0, 1\rangle$, takže zde nabývá maxima. To nemůže být v krajních bodech a proto je nabyto uvnitř dané množiny. To tedy odpovídá hledání maxima funkce

$$\ell(w) = \ln\big(L(w)\big) = 33 \cdot \ln(w) + 40 \cdot \ln(1-w)$$

na intervalu $(0, 1)$. Protože maximum existuje, musí pro něj platit

$$0 = \ell'(q) = \frac{33}{w} - \frac{40}{1-w}$$

neboli

$$w = \frac{33}{73} \doteq 0.45205$$

což vyhovuje zadání.

Metoda momentů:

Střední hodnota je

$$E(X) = \sum_{i=0}^{\infty} iw^i(1-w) = \sum_{i=1}^{\infty} iw^i - \sum_{i=1}^{\infty} iw^{i+1} = \sum_{i=1}^{\infty} iw^i - \sum_{i=2}^{\infty}(i-1)w^i =$$

$$= w + \sum_{i=2}^{\infty}(i-i+1)w^i = \sum_{i=1}^{\infty} w^i = w\sum_{i=1}^{\infty} w^{i-1} = \frac{w}{1-w}$$

její odhad z realizace je

$$\overline{x} = \frac{1}{n}\sum_{i} i\, n_i = \frac{1}{20+10+7+3} \cdot (0 \cdot 20 + 1 \cdot 10 + 2 \cdot 7 + 3 \cdot 3) = \frac{33}{40} \ .$$

Porovnáním dostaneme

$$\frac{w}{1-w} = E(X) = \overline{x} = \frac{33}{40}$$

což dává opět řešení

$$w = \frac{33}{73} \doteq 0.45205$$

jako v předchozí metodě.

Jak je snadno vidět, v případě geometrického rozdělení dostáváme pro jeho parametr w vždy stejné výsledky pro obě metody.