Příklad (3) ze vzorové písemky kolegy Korbeláře

20.11.2014

.

(3) Házíme 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 0.95.

Řešení Označíme-li X_i náhodnou veličinu nabývající 1, pokud v i-tém hodu padne šestka a 0, pokud nepadne, pak má X_i alternativní rozdělení s parametrem $p = \frac{1}{6}$. Tedy $\mathbb{E}X = \frac{1}{6}$ a $DX = \frac{5}{36}$. Označíme-li

$$R_n = \frac{\sum_{i=1}^n X_i}{n}$$

relativní výskyt šestek v n hodech, potom je

$$\mathbb{E}R_n = \frac{\sum_{i=1}^n \mathbb{E}X_i}{n} = \frac{n\frac{1}{6}}{n} = \frac{1}{6}, \qquad DR_n = \frac{\sum_{i=1}^n DX_i}{n^2} = \frac{5}{36},$$

neboť veličiny X_i jsou navzájem nezávislé.

Čebyševova nerovnost nám dává

$$\mathbf{P}(|R_n - \frac{1}{6}| > 0.05) \leq \frac{DR_n}{(0.05)^2} = \frac{2000}{36n}.$$

Na druhou stranu jsme ze zadání chtěli mít

$$\mathbf{P}(|R_n - \frac{1}{6}| > 0.05) \leq 0.05,$$

tedy $\frac{2000}{36n} = 0.05$ a z toho plyne $n \geq 1112$.

Pomocí CLV bychom měli

$$\mathbf{P}(|R_n - \frac{1}{6}| \leq 0.05) = \mathbf{P}(R_n - \frac{1}{6} \leq 0.05) - \mathbf{P}(R_n - \frac{1}{6} < -0.05) = \mathbf{P}\left(\frac{R_n - \frac{1}{6}}{\sqrt{\frac{5}{36n}}} \leq \frac{0.05}{\sqrt{\frac{5}{36n}}}\right) - \mathbf{P}\left(\frac{R_n - \frac{1}{6}}{\sqrt{\frac{5}{36n}}} \leq \frac{-0.05}{\sqrt{\frac{5}{36n}}}\right) =$$

$$= \Phi\left(\frac{3\sqrt{n}}{10\sqrt{5}}\right) - \Phi\left(-\frac{3\sqrt{n}}{10\sqrt{5}}\right) = 2\Phi\left(\frac{3\sqrt{n}}{10\sqrt{5}}\right) - 1.$$

Dle zadání chceme, aby platilo $2\Phi(\frac{3\sqrt{n}}{10\sqrt{5}}) - 1 \geq 0.95$, což snadno dopočteme jako

$$\begin{aligned} 2\Phi\left(\frac{3\sqrt{n}}{10\sqrt{5}}\right) - 1 &\geq 0.95 \\ \Phi\left(\frac{3\sqrt{n}}{10\sqrt{5}}\right) &\geq 0.975 \\ \frac{3\sqrt{n}}{10\sqrt{5}} &\geq q(0.975) = 1.96 \\ n &\geq 213.422222. \end{aligned}$$

Tedy hodů kostkou musí být alespoň 214.

Poznámka k výpočtu rozptylu Rozptyl veličiny R_n lze spočítat jednoduše tak, jak je výše popsáno. Poněkud komplikovaněji jsem se jej pokusil spočítat já na konzultaci, uvádím znovu zde:

$$DR_n = \mathbb{E}(R_n)^2 - (\mathbb{E}R_n)^2$$

Platí $\mathbb{E}R_n = \frac{1}{6}$, vypočítáme tedy hodnotu výrazu $\mathbb{E}(R_n)^2$:

$$\mathbb{E}(R_n)^2 = \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right)^2 = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^n X_i^2 + \sum_{i\neq j} X_i X_j\right) \overset{(1)}{=} \frac{1}{n^2}\left(\sum_{i=1}^n \mathbb{E}X_i^2 + \sum_{i\neq j} \mathbb{E}X_i\mathbb{E}X_j\right) = \frac{1}{n^2}\left(n\frac{1}{6} + n(n-1)\frac{1}{36}\right)$$

$$= \frac{1}{6n} + \frac{n-1}{36n} = \frac{n+5}{36n},$$

Kde rovnost (1) platí, protože X_i a X_j pro $i \neq j$ jsou navzájem nezávislé. Dohromady tedy máme

$$DR_n = \mathbb{E}(R_n)^2 - (\mathbb{E}R_n)^2 = \frac{n+5}{36n} - \frac{1}{36} = \frac{5}{36n}.$$