

Připomenutí teorie

- **Čebyševova nerovnost:** Nechť X má konečnou střední hodnotu μ a rozptyl σ^2 . Pak

$$P[|X - \mu| \geq a\sigma] \leq 1/a^2.$$

- **Centrální limitní věta:** Nechť $X_1, \dots, X_n$ jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2 . Označme $Y_n := \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}}$. Pak $Y_n \rightarrow^d \mathcal{N}(0, 1)$. Neboli

$$\lim_{n \rightarrow \infty} F_{Y_n}(x) = \Phi(x) \text{ pro každé } x \in \mathbb{R}.$$

- **De Moivre-Laplaceova věta** říká o něco silnější věc: v okolí pn je i pravděpodobnostní funkce $\text{Bin}(n, p)$ dobře aproximována hustotou $\mathcal{N}(np, np(1-p))$, tedy vhodně přeskálovanou Gaussovou funkcí φ . Přesněji:

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{(k-np)^2}{2np(1-p)}}$$

pro k blízké np . Ještě přesněji: pokud pro nějaké c je $|k - np| < c\sqrt{np(1-p)}$ a n se blíží k nekonečnu, tak poměr dvou výrazů nahoře se blíží k jedné.

- **Tabulka Φ :**

x	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5
$\Phi(x)$	0.01	0.02	0.07	0.16	0.31	0.5	0.69	0.84	0.93	0.98	0.99

Aplikace nerovností a Centrální Limitní Věty

1. Měření výšky

Statistik chce odhadnout průměrnou výšku h (v metrech) lidí v nějaké populaci, pomocí n nezávislých vzorků $X_1, \dots, X_n$, které vybíráme uniformně náhodně ze všech možných lidí. Pro odhad použije výběrový průměr $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$. Odhaduje, že směrodatná odchylka jednoho měření je nejvýše 1 metr.

- Jak velké n má volit, aby směrodatná odchylka $\bar{X}_n$ byla nejvýše 1 cm?
- Pro jaké n zajistí Čebyševova nerovnost, že $\bar{X}_n$ se liší od h nejvýše o 5 cm s pravděpodobností alespoň 99%?
- Statistik si všimne, že všichni měření lidé mají výšku v intervalu (1.4, 2.1). Jak má upravit odhad směrodatné odchylky? Jak se změní odpovědi na předchozí otázky?

2. Odhad binomického koeficientu

Odhadněte $\binom{100}{30}$ pomocí CLV. Náповěda: použijte CLV pro odhad $P[29.5 < X < 30.5]$ pro vhodnou n.v. X . Na druhou stranu pro $P[X = 30]$ máme vzorec $\binom{100}{30} \cdot 2^{-100}$ z binomického rozdělení. Alternativně, můžete použít Moivre-Laplaceho větu. Pokud máte po ruce vhodný stroj, vyčíslete.

3. Stahování souborů

Měříme rychlost stahování souborů z cloudového úložiště. Každý čas stahování jednoho souboru je náhodná veličina s průměrem $\mu = 5$ minut a standardní odchylkou $\sigma = 2$ minuty. Předpokládejme, že časy stahování jednotlivých souborů jsou na sobě nezávislé, stahování probíhá jedno po druhém (tj. vždy se stahuje jen jeden soubor, hned po jeho dokončení začneme stahovat další).

- Pokud stáhneme 50 souborů, jaká je přibližná pravděpodobnost, že celková doba stahování přesáhne 270 minut?
- Jaká je přibližná pravděpodobnost, že průměrná doba stahování na soubor je kratší než 4,5 minuty?

Použijte Centrální limitní větu. Napište přesnou formuli pomocí funkce Φ a použijte tabulku pro odhad.

4. Odhad binomické kumulativní distribuce

Označme $S = \sum_{k=0}^{30} \binom{100}{k}$. Označme dále $X = \sum_{i=1}^{100} X_i$, kde X_i je 0 nebo 1, obojí s pravděpodobností $1/2$ a veličiny $X_1, \dots, X_n$ jsou nezávislé. Je tedy $X \sim \text{Bin}(100, 1/2)$.

- Vyjádřete S pomocí distribuční funkce F_X .
- Použijte CLV na odhad této pravděpodobnosti.
- Případně vyčíslete S vhodným softwarem a srovnajte.

5. Test hypotézy

Chceme odhadnout, zda naše mince (a způsob jak s ní házíme) je spravedlivá. Pokud ze sta hodů padne orel více než 55-krát, řekneme, že spravedlivá není. Jaká je pravděpodobnost, že se zmýlíme?