10. cvičení

Pravděpodobnost a statistika 1, 23. 4. 2026 `https://iuuk.mff.cuni.cz/~chmel/2526/past/`

Nerovnosti

Úloha 1 (These two are the same picture)
Čebyševova nerovnost byla na přednášce v jiném znění: $\Pr(|X - \mathbb{E}(X)| \geq a) \leq \operatorname{var}(X)/a^2$. Rozmyslete si, že znění v taháku je ekvivalentní.

Úloha 2 (Zase kostky?)
Házíme kostkou, za 1 a 2 dostaneme bod. Označme X počet bodů, které dostaneme po n (nezávislých) hodech. Odhadněte pravděpodobnost, že $X \geq n/2$.

a) Pomocí Markovovy nerovnosti.

b) Pomocí Čebyševovy nerovnosti.

c) Pro konkrétní n, jak lze tuto hodnotu určit přesně?

Úloha 3 (Statistika výšky)
Statistik chce odhadnout průměrnou výšku h (v metrech) lidí v nějaké populaci, pomocí n nezávislých vzorků $X_1, \ldots, X_n$, které vybíráme uniformně náhodně ze všech možnych lidí. Pro odhad použije výběrový průměr $S_n = (X_1 + \cdots + X_n)/n$. Odhaduje, že směrodatná odchylka jednoho výběru je nejvýše 1 metr.

a) Jak velké n má volit, aby směrodatná odchylka S_n byla nejvýše 1 cm?

b) Pro jaké n zajistí Čebyševova nerovnost, že pravděpodobnost, že S_n se liší od h nejvýše o 5 cm s pravděpodobností alespoň 99 %?

c) 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?

Zákony velkých čísel

Úloha 4 (Počítání obsahu kruhu náhodným samplováním)
Vygenerujeme náhodný bod v jednotkovém čtverci (obě souřadnice budou mít rozdělení $U(0,1)$). Označíme X_i indikátor jevu „i-tý bod leží ve vepsaném kruhu“.

a) Určete $\mathbb{E}(X_i)$, $\operatorname{var}(X_i)$.

b) Položte $\overline{X}_n = (X_1 + \cdots + X_n)/n$. Určete $\mathbb{E}(\overline{X}_n)$ a $\operatorname{var}(\overline{X}_n)$.

c) Rozmyslete si, co říká slabý a silný zákon velkých čísel o aproximaci π pomocí X_n?

d) Pro jaké n čekáte, že dostaneme výsledek správně na jedno desetinné místo? Na dvě, tři, ...?

e) Jiný výpočet obsahu kruhu: $Y_i = \sqrt{1 - U_i^2}$, kde $U_i \sim U(0,1)$. Uvědomte si, že $\mathbb{E}(Y_i)$ je obsah čtvrtkruhu, tedy $\pi/4$. Jaké je $\operatorname{var}(Y_i)$? Jaké je $\operatorname{var}(\overline{Y}_n)$?

f) Která metoda je přesnější?

Centrální limitní věta

Úloha 5 (Standardizace)
Připomeňme, že standardizací n.v. X myslíme $\operatorname{stand}(X) = (X - \mathbb{E}(X))/\sigma_X$.

a) $\operatorname{stand}(X)$ má střední hodnotu 0 a rozptyl 1

b) Y_n v CLV je rovna $\operatorname{stand}(\overline{X}_n)$ a také $\operatorname{stand}(X_1 + \cdots + X_n)$.

Úloha 6 (Downloading...)
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).

a) Pokud stáhneme 50 souborů, jaká je přibližná pravděpodobnost, že celková doba stahování přesáhne 270 minut?

b) 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 na předchozí straně pro odhad.

Úloha 7 (Rozdíl počtu hodů mincí)
Označme $S = \sum_{k=0}^{30} \binom{100}{k}$. Označme dále $X = \sum_{i=1}^{100} X_i$, kde X_i je $0, 1$ s pravděpodobností $1/2$ (tedy $X_i \sim \mathrm{Bern}(1/2)$) a veličiny $X_1, \ldots, X_n$ jsou nezávislé.

a) Vyjádřete S pomocí distribuční funkce X.

b) Použijte CLV na odhad této pravděpodobnosti.

c) Zkuste S vyčíslit vhodným softwarem a srovnejte.

Kvantilová funkce

Úloha 8 (Kvantilová funkce)
Kvantilovou funkci jsme definovali předpisem $Q_X(p) = F_X^{-1}(p)$.

a) Jaký je obor hodnot Q_X? Kdy dává definice smysl?

b) Nahlédněte, že taková funkce má (zjevně) tu vlastnost, že $Q_X(p) = x \iff p = F_X(x)$.

c) Rozmyslete si, jak je rozumné definovat Q_X pro diskrétní n.v. (a v čem je vlastně problém).

Tahák

- **Markovova nerovnost:** $P(X \geq a\mathbb{E}(X)) \leq \frac{1}{a}$ pro $X \geq 0$.
- **Čebyševova nerovnost:** $P(|X - \mathbb{E}(X)| \geq t\sigma_X) \leq \frac{1}{t^2}$.
- Nechť $X_1, \ldots, X_n$ jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2. Definujeme $\overline{X}_n := (X_1 + \cdots + X_n)/n$.
  - **Silný zákon velkých čísel** $\overline{X}_n \xrightarrow{s.j.} \mu$
  - **Slabý zákon velkých čísel** $\overline{X}_n \xrightarrow{P} \mu$, neboli $(\forall \varepsilon > 0)\ \Pr(|\overline{X}_n - \mu| > \varepsilon) \to 0$.
    Ukazovali jsme si, že dokonce $\Pr(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$.
  - **Centrální limitní věta:** Označme $Y_n = ((X_1 + \cdots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$. Pak $Y_n \xrightarrow{d} N(0,1)$. Neboli

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

x	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	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