11. cvičení

Aplikace nerovností a CLV

Úloha 1 (Pošta)
Na poště v Jindřišské v průměru za den zpracují 10 000 dopisů denně. (Tj. střední hodnota je 10 000.)

a) Co nám Markovova nerovnost říká o pravděpodobnosti, že zítra bude poště muset zpracovat aspoň 15 tisíc dopisů?

b) Předpokládejme navíc, že rozptyl $\sigma^2 = 2000$. Co nám říká Čebyševova nerovnost o pravděpodobnosti toho, že pošta bude muset zítra zvládnout mezi 8000 a 12000 dopisy?

c) Můžeme nějak použít Čebyševovu nerovnost k tomu, abychom určili pravděpodobnost toho, že pošta bude muset zítra zpracovat aspoň 15 tisíc dopisů?

Řešení a) (Pozor na předpoklady!) $P[X \geq 15000] \leq \frac{\mathbb{E}[X]}{15000} = 2/3$

b) $P[8000 < X < 12000] = P[-2000 < X - 10000 < 2000] = P[|X - 10000| < 2000]$. Z Čebyševa, $P[|X - 10000| \geq 2000] \leq \frac{\sigma^2}{2000^2} = \frac{1}{2000}$, tedy hledaná pravděpodobnost je $1 - \frac{1}{2000}$.

c) Ano: $P[X \geq 15000] = P[X - 10000 \geq 5000] \leq P[X - 10000 \geq 5000$ nebo $X - 10000 \leq -5000] = P[|X - 10000| \geq 5000] \leq \frac{2000}{5000^2} = \frac{1}{12500}$.

Úloha 2 (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 %?

Řešení
(Všechno počítáme v cm.)

a) X_i nezávislé, tedy rozptyl součtu je součet rozptylů, vynásobení konstantou znamená, že je to druhá mocnina konstanty rozptylu, tedy $\mathrm{var}(S_n) = \frac{1}{n^2}\mathrm{var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) = \frac{n}{n^2}100^2 = \frac{10000}{n}$, tedy chceme $n \geq 10000$.

b) Máme tedy dle Čebyševa $P(|S_n - h| \geq \frac{100a}{\sqrt{n}}) \leq \frac{1}{a^2}$. Položíme-li $100a/\sqrt{n} = 5$, máme $a = \frac{\sqrt{n}}{20}$, a tedy $1/a^2 = 400/n$, což chceme rovno 0.01, a tedy $n = 40000$.

Úloha 3 (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 ± 1 s pravděpodobností 1/2 a veličiny $X_1, \ldots, X_n$ jsou nezávislé.

a) Spočítejte $P(X = k)$ pro každé k.

b) Vyjádřete S pomocí pravděpodobnosti vhodného výroku o X.

c) Použijte Čebyševovu nerovnost a CLV na odhad této pravděpodobnosti.

d) Vyčíslete S vhodným softwarem a srovnejte.

Řešení a) Pro $k \in 2\mathbb{Z} \cap [-100, 100]$: (tj. $k = 2\ell, -50 \leq \ell \leq 50$) $P[X = 2k - 50] = \binom{100}{k}\frac{1}{2^{100}}$.

b) $S = 2^{100} \cdot P[X \leq -40]$

c) Protože evidentně $\mathbb{E}[X] = 0$, $\text{var}(X) = 4 \cdot 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 100$ (bo $S \sim 2\text{Bin}(100, 1/2) - 100$), $P[X \leq -40] \leq P[|X| \geq 40] \leq \frac{1}{16}$, podle CLV, máme $\mathbb{E}[X_i] = 0, \text{var}(X_i) = 4p(1-p) = 1$, když budeme mít jednotlivé samply $X_1, \ldots, X_n$, zajímá nás $P[\sum X_i \leq -40] = P[\frac{\sum X_i}{10} \leq -4] = \Phi(-4)$ (řekněme, že 100 je dost velké)

d) TODO

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 $\mathbb{E}(X_i)$, $\text{var}(X_i)$.

b) Položte $S_n = (X_1 + \cdots + X_n)/n$. Určete $\mathbb{E}(S_n)$ a $\text{var}(S_n)$.

c) Všimněte si, že lze počítat S_n z S_{n-1}, X_n a n (nižší nároky na paměť).

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

Řešení a) $X_i \sim \text{Bern}(\pi/4) \rightsquigarrow \mathbb{E}[X_i] = \pi/4, \text{var}(X_i) = (\pi/4)(1 - \pi/4)$

b) $\mathbb{E}[S_n] = \pi/4$ z linearity, z nezávislost $\text{var}(S_n) = \frac{1}{n^2} \cdot (\sum \text{var}(X_i)) = \frac{(\pi/4)(1-\pi/4)}{n}$

c) $S_n = ((n-1)S_{n-1} + X_n)/n$

d) TODO

Spojité vektory

Úloha 5 (Rozlomit do trojúhelníka)
Metrový klacek rozlomíme na tři kusy jedním z níže popsaných způsobů. Pro každý z nich spočítejte, jaká je pravděpodobnost, že ze získaných tří kusů jde sestavit trojúhelník. (Nápověda: napřed si rozmyslete, kdy jsou tři kladná čísla se součtem 1 stranami nějakého trojúhelníku.)

a) Vybereme uniformně náhodně dva body zlomu.

b) Vybereme uniformně náhodně první bod zlomu. Pak totéž uděláme s kusem klacku v pravé ruce.

c) Vybereme uniformně náhodně první bod zlomu. Pak totéž uděláme s větším kusem klacku.

Řešení
Máme $X + Y + Z = 1$, potřebujeme trojúhelníkovou nerovnost: $X + Y \geq Z, X + Z \geq Y, Y + Z \geq X$, speciálně tedy musí platit $1 - X \geq X$, a tedy $X \leq 1/2$, pro oba další analogicky.

a) Problém je, když se oba řezy trefí do stejné poloviny, nebo když vzdálenost mezi řezy je aspoň $1/2$: tedy $P[\triangle] = 1 - (1/4 + 1/4 + 1/4) = 1/4$ – alternativně se dá nakreslit prostor daný X, Y, a případy odpovídají částem prostoru.

b) $\int_0^{1/2} \int_{1/2-x}^{1/2} \frac{1}{1-x} dy dx = \log(2) - 1/2 \approx 0.193$ (tj. x musí být maximálně $1/2$, a y musí být aspoň $1/2 - x$, aby z nebylo velké, a zároveň musí být maximálně $1/2$).

c) $\int_0^{1/2} \int_{1/2-x}^{1/2} \frac{2}{1-x} dy dx = 2(\log(2) - 1/2)$ (hustota je dvojnásobná, protože $x, 1-x$ odpovídají témuž; alternativně jde říct, že nejdříve vezmeme uniformně hodnotu z intervalu $[0, 1/2]$ jako tu menší, a potom pracujeme se zbytkem "jakoby v pravé ruce").

Tahák

- *Markovova nerovnost:* $P(X \geq a) \leq \mathbb{E}(X)/a$ pro $X \geq 0$, $a > 0$.
- *Čebyševova nerovnost:* $P(|X - \mu| \geq a\sigma) \leq 1/a^2$ pro $\mu = \mathbb{E}(X)$, $\sigma^2 = \text{var}(X)$.
- *Černovova nerovnost:* $P(X \geq t) = P(X \leq -t) \leq e^{\frac{-t^2}{2\sigma^2}}$ pro $X = X_1, \ldots, X_n$, kde všechna $X_i = \pm 1$ jsou n.n.v. a $\sigma^2 = \text{var}(X)$.

- *Centrální limitní věta:* Pro $X_1, X_2, \ldots$ stejně rozdělené n.v. s konečným μ a σ, $Y_n = \frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$ platí $Y_n \xrightarrow{d} \Phi$.
- Na počítání $\Phi(x)$: `https://t.ly/JRQ2`

Domácí úkol 11

`https://iuuk.mff.cuni.cz/~chmel/2223/past/` Zadáno 26. 4. 2023

Pravděpodobnost a statistika 1 Odevzdejte do 3. 5. 2023 9:00 přes Poštovní sovu

---

Úloha 1 (Mince)
Máme minci, na které padne orel s pravděpodobností 1/10. Hodíme s ní 200x. Použijte Markovovu nerovnost abyste shora odhadli pravděpodobnost, že na minci padne aspoň 120x orel.
Pak tento odhad vylepšete pomocí Čebyševa.