

10. cvičení z PSt — 23. 4. 2026

Věta 1. Necht X_1, \dots, X_n jsou stejně rozdělené n.n.v. se střední hodnotou μ a rozptylem σ^2 . Necht $Y_n = \frac{(X_1 + \dots + X_n) - n \cdot \mu}{\sigma \cdot \sqrt{n}}$. Pak platí

$$\begin{aligned} \bar{X}_n &\xrightarrow{s.j.} \mu && \text{(Silný zákon velkých čísel)} \\ (\forall \varepsilon > 0) \quad P(|\bar{X}_n - \mu| > \varepsilon) &\leq \frac{\sigma^2}{n\varepsilon^2} && \text{(Slabý zákon velkých čísel)} \\ (\forall x \in \mathbb{R}) \quad \lim_{n \rightarrow \infty} F_{Y_n}(x) &= \Phi(x) && \text{(Centrální limitní věta)} \end{aligned}$$

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

Zákony velkých čísel

1. Počítání obsahu kruhu náhodným samplováním. Vygenerujeme náhodný bod ve čtverci (obě souřadnice budou mít rozdělení $U(0, 1)$). Označíme X_i indikátor jevu „ i -tý bod leží v kruhu s poloměrem 1 a středem v počátku“.

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

Řešení: (Předpokládáme, že jsme si zadefinovali že π je prostě obsah kruhu s poloměrem jedna, a teď chceme toto π odhadnout.)

Víme, že $\mathbb{E}(X_i)$ je prostě poměr kolikrát se trefíme do (čtvrt)kruhu, tj $\mathbb{E}(X_i) = \frac{\pi}{4}$.

Rozptyl můžeme odhadnout buď z definice, a nebo si prostě uvědomíme že $X_i \sim \text{Ber}(\frac{\pi}{4})$ a tedy $\text{var}(X_i) = \frac{\pi}{4} (1 - \frac{\pi}{4})$. □

(b) Položte $\bar{X}_n = (X_1 + \dots + X_n)/n$. Určete $\mathbb{E}(\bar{X}_n)$, $\text{var}(\bar{X}_n)$.

Řešení: Z linearit střední hodnoty rovnou získáváme

$$\mathbb{E}(\bar{X}_n) = \frac{\sum_{i=1}^n \mathbb{E}(X_i)}{n} = \frac{n \cdot \frac{\pi}{4}}{n} = \frac{\pi}{4}.$$

Rozptyl je zajímavější: nejprve použijeme $\text{var}(aX) = a^2 \text{var} X$, a pak použijeme linearitu rozptylu pro nezávislé n.v., dostáváme

$$\text{var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n} \cdot \frac{\pi}{4} \left(1 - \frac{\pi}{4}\right).$$

□

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

Řešení: Silný nám ukazuje že s pravděpodobností 1 konverguje průměr k $\frac{\pi}{4}$, a tedy pro $n \rightarrow \infty$ dobře aproximuje $\frac{\pi}{4}$. Slabý ve standardním znění říká něco slabšího, ale my jsme na přednášce dokonce odvodili vzorec, viz věta 1, tedy máme dokonce pro konkrétní n a konkrétní ε odhad na pravděpodobnost, že budeme vzdáleni od $\frac{\pi}{4}$ více než o ε . □

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

Řešení: Abychom získali přesnost na k desetinných míst, musíme omezit $\varepsilon = 10^{-k}$ (cca). Teď otázka zní, co přesně znamená *čekáte*. Tím se může myslet spousta věcí, ale nejpřirozenější je, že si můžeme stanovit nějakou konstantní pravděpodobnost p a říct si, že se chceme mýlit (tj, v prvních k des. místech bude chyba) s pravděpodobností nejvýše p .

Chceme tedy, aby

$$P\left(\left|X - \frac{\pi}{4}\right| > \varepsilon\right) \leq p.$$

Nyní můžeme použít slabý zákon o velkých číslech a dostaneme

$$P\left(\left|X - \frac{\pi}{4}\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \leq p,$$

a z toho

$$n \geq \frac{\sigma^2}{p\varepsilon^2} = \frac{\sigma^2}{p \cdot 10^{-2k}} = \frac{\sigma^2}{p} \cdot 10^{2k}. \quad (1.1)$$

První část tohoto výrazu je nějaká konstanta závislá na rozdělení X_i a na námi určené pravděpodobnosti chyby (třeba $p = \frac{1}{2}$, nebo $p = 10^{-5}$). Druhá část je ta která určuje závislost na k , a vidíme že $n \sim 10^{2k}$. \square

(e) Jiný výpočet: $Y_i = \sqrt{1 - U_i^2}$, kde $U_i \sim U(0, 1)$. Jaké je $\text{var}(Y_i)$ a $\text{var}(\bar{Y}_n)$? (*Hint dole*)¹

Řešení: Spočítáme $\text{var}(Y_i) = \mathbb{E}(Y_i^2) - (\mathbb{E}(Y_i))^2$. Druhou část známe z hintu: to je $(\frac{\pi}{4})^2$, a první část je z definice $\mathbb{E}(g(U))$ pro $U \sim U(0, 1)$, tedy

$$\int_0^1 (\sqrt{1 - x^2})^2 \cdot 1 \, dx = \int_0^1 1 - x^2 \, dx = 1 - \int_0^1 x^2 \, dx = 1 - \frac{1}{3} = \frac{2}{3}.$$

Rozptyl je tedy

$$\text{var}(Y_i) = \frac{2}{3} - \left(\frac{\pi}{4}\right)^2.$$

\square

(f) Která metoda je přesnější?

Řešení: Vyčíslením těchto dvou rozptylů dostáváme že rozptyl Y_i je menší než X_i . Protože rozptyly ve vztahu 1.1 vystupují tak že zvětšují n , znamená to že Y_i potřebuje méně samplů na dosažení stejné přesnosti. Poznámka: na to, abychom toto zjistili, potřebujeme odhad na π ... \square

Centrální limitní věta

2. Připomeňme, že standardizací n.v. X myslíme $\text{stand}(X) = (X - \mathbb{E}(X))/\sigma_X$.

(a) Ukažte, že $\text{stand}(X)$ má střední hodnotu 0 a rozptyl 1.

(b) Ukažte, že Y_n v CLV je rovna $\text{stand}(\bar{X}_n)$ a také $\text{stand}(X_1 + \dots + X_n)$.

3. 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). Stahujeme celkem 50 souborů.

(a) Jaká je přibližná pravděpodobnost, že celková doba stahování přesáhne 270 minut?

¹Uvědomte si, že $\mathbb{E}(Y_i)$ je obsah čtvrtkruhu, tedy $\pi/4$.

Řešení: Označme X_1, \dots, X_n jednotlivé časy stahování, kde tedy $n = 50$ (budu odvozovat obecný vzorec, a dosadím do něj za n, μ, σ až na konci). Označme také $X = \sum_{i=1}^n X_i$ a $Y = (X - n\mu) / (\sigma\sqrt{n})$. Centrální limitní věta nám říká, že Y je blízko $N(0, 1)$, tj. pravděpodobnostní funkci Y můžeme odhadnout Φ .

Příklad chce $P(X > 270)$. To co vlastně teď chceme udělat je převést to do stavu $P(Y > \dots)$, kde levá strana pak je prostě něco co najdeme v tabulce pro Φ .

Asi nejvíce mechanický postup je následující. Protože $Y = (X - n\mu) / (\sigma\sqrt{n})$, víme, že

$$P(X > 270) = P\left(Y > \frac{270 - n\mu}{\sigma\sqrt{n}}\right).$$

Více méně jsme do definice Y dosadili 270 za X . Toto platí protože ty dvě nerovnice lze na sebe navzájem převést. Teď už stačí dosadit, tedy

$$\begin{aligned} P\left(Y > \frac{270 - n\mu}{\sigma\sqrt{n}}\right) &= P\left(Y > \frac{270 - 50 \cdot 5}{2\sqrt{50}}\right) = P\left(Y > \frac{20}{2\sqrt{50}}\right) = P\left(Y > \frac{20}{2 \cdot 5 \cdot \sqrt{2}}\right) \\ &= P\left(Y > \frac{2}{\sqrt{2}}\right) = P\left(Y > \sqrt{2}\right) = 1 - P\left(Y \leq \sqrt{2}\right) \doteq 1 - \Phi\left(\sqrt{2}\right). \end{aligned}$$

Trochu méně mechanický způsob je si říct následující: víme, že libovolná $Z \sim N(\mu, \sigma^2)$ je prostě $\sigma Z' + \mu$ pro $Z' \sim N(0, 1)$. Tedy $\Phi(x) = P(Z' \leq x)$ udává pravděpodobnost, že Z bude nejvýše o x -násobek své odchylky nad svou střední hodnotou. Pro naši $X = X_1 + \dots + X_n$ je střední hodnota $n \cdot \mu = 250$ a odchylka $\sqrt{n}\sigma = \sqrt{50} \cdot 2 = 10\sqrt{2}$. Příklad po nás chce vzdálenost od 270, což je o 20 od střední hodnoty, tj. $\frac{20}{10\sqrt{2}} = \sqrt{2}$ -násobek odchylky.

Mimochodem, $\sqrt{2}$ je cca 1.5, tedy výsledek je z tabulky něco pod 0.93. □

(b) Jaká je přibližná pravděpodobnost, že průměrná doba stahování na soubor je kratší než 4,5 minuty?

Řešení: Stejně jako předtím, jen tentokrát se bude hodit druhý způsob vyjádření Y (viz příklad 2b), a to jako

$$Y = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Znovu jako předtím funguje mechanický postup

$$\begin{aligned} P(\bar{X}_n < 4,5) &= P\left(Y < \frac{4,5 - \mu}{\sigma/\sqrt{n}}\right) = P\left(Y < \frac{4,5 - 5}{2/\sqrt{50}}\right) = P\left(Y < \frac{-0,5}{2/(\sqrt{2} \cdot 5)}\right) = P\left(Y < \frac{-5\sqrt{2}}{4}\right) \\ &\doteq \Phi\left(\frac{-5\sqrt{2}}{4}\right) \doteq \Phi(-1,77) \in [0,02, 0,07]. \end{aligned}$$

Mimochodem i tady funguje ten méně mechanický postup, jen pozor že teď ho používáme pro průměr místo součtu, tj. bude střední hodnota μ a rozptyl μ/\sqrt{n} . □

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

4. 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 vyčístele S vhodným softwarem a srovnejte.

Kvantilová funkce

5. Kvantilovou funkci jsme definovali předpisem $Q_X(p) = F_X^{-1}(p)$.

- Jaký je obor hodnot Q_X ? Kdy dává definice smysl?
- Rozmyslete si, jak je rozumné definovat Q_X pro diskrétní n.v. (a v čem je vlastně problém).

6*. Ukažte pro libovolnou n.v. X a $U \sim U(0, 1)$ že $Q_X(U)$ má stejné rozdělení jako X . (*Hint dole*)²

²Ukažte, že $Q_X(U)$ i X mají stejnou distribuční funkci.

MLE a nestrannost

7. Připomeňte si, co je to věrohodnost, estimátor, bias estimátoru.

8. Němci vyrábějí tanky s pořadovými čísly $1, \dots, N$ pro neznámé N . Ukořistíme k z nich a vidíme pořadová čísla X_1, \dots, X_k , tj. rovnoměrně náhodnou k -prvkovou podmnožinu $\{1, \dots, N\}$. Nechť $M = \max(X_1, \dots, X_k)$.

(a) Ukažte, že $P(M = m) = \binom{m-1}{k-1} / \binom{N}{k}$ pro $m \in \{k, \dots, N\}$.

(b) Připomeňte si, že M je MLE pro N (ukazovali jsme na přednášce).

(c) Spočtete $\mathbb{E}(M) = \frac{k(N+1)}{k+1}$. Může se hodit hockey-stick identity: $\sum_{m=k}^N \binom{m}{k} = \binom{N+1}{k+1}$.

(d) Jak z toho plyne nestranný odhad $\hat{N}_{unbiased} = \frac{k+1}{k} M - 1$?

(Pozor: na přednášce bylo místo -1 napsáno $-\frac{k+1}{k}$ — to byla chyba.)

K procvičení

9. Víme, že průměrný počet bodů z písemky byl 40 (ze 100). Odhadněte odsud podíl studentů s alespoň 80 body. Vylepšete odhad, pokud víte, že směrodatná odchylka počtu bodů je 10.

10. 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?

11. Vzpomeňte si na větu z přednášky. Nechť $U \sim U(0, 1)$. Jak vyrobíte náhodnou veličinu

(a) s rozdělením $U(a, b)$?

(b) s rozdělením $N(0, 1)$? (Využijte funkce Φ jako “black box”).

(c) s uniformním rozdělením na množině $\{1, 2, \dots, 6\}$?

12. Odhadněte $\binom{100}{30}$ pomocí CLV. (*Hint dole*)³

³Použijte CLV pro odhad $P(29.5 < X < 30.5)$ pro vhodnou n.v. X .