

## 12. cvičení z PSt — 6.–12.5.2025

### Aplikace Centrální Limitní Věty

- **Centrální limitní věta:** Necht'  $X_1, \dots, X_n$  jsou stejně rozdělené n.n.v. se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ . Označme  $Y_n = ((X_1 + \dots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$ . Pak  $Y_n \xrightarrow{d} N(0, 1)$ . Neboli

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

1. 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). Stáhneme 50 souborů.

(a) 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  $\Phi$ .

### Bodové odhady

- Zkoumáme posloupnost n.n.v. se stejným rozdělením, např.  $Geom(\theta)$ ,  $U(0, \theta)$ , kde  $\theta$  je parametr.
- Zapisujeme  $X_1, \dots, X_n \sim F_\theta$ , tzv. **náhodný výběr** z  $F_\theta$  (model s parametrem).
- Naměříme  $X_1 = x_1, \dots$ , chceme odhadnout  $\theta$ .
- $\hat{\theta}_n \dots$  nějaká metoda jak odhadnout  $\theta$  pomocí naměřených dat (hodnot  $X_1, \dots, X_n$ ), angl. *estimator*
- $m_r(\theta) = \mathbb{E}(X^r)$  pro  $X \sim F_\theta \dots$  **r-tý moment**, ideální vlastnost rozdělení
- $\hat{m}_r(\theta) = \frac{1}{n} \sum_{i=1}^n X_i^r \dots$  **r-tý výběrový moment**, náhodná veličina, funkce našeho naměřeného vzorku (tj. statistika)
- **Odhad metodou momentů** vyřešíme rovnici  $m_1(\theta) = \hat{m}_1(\theta)$  pro neznámou  $\theta$ .
- event. soustavu rovnic  $m_r(\theta) = \hat{m}_r(\theta)$  pro  $r = 1, 2, \dots$  podle potřeby.
- $L(\theta; x_1, \dots, x_n) = P(X_1 = x_1 \& \dots \& X_n = x_n) \dots$  pravd. pozorovaných dat závislá na parametru  $\theta$ .
- nebo  $L(\dots) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) \dots$  hustota pravděpodobnosti  $\dots$
- $\ell(\theta; x_1, \dots, x_n) = \log L(\dots) \dots$  pro snazší výpočty.
- **Odhad metodou maximální věrohodnosti (Maximal Likelihood)** hledáme  $\theta$ , pro které je maximální  $L(\theta; x_1, \dots, x_n)$ , resp.  $\ell(\dots)$ . Obvykle pomocí derivací funkce  $L$ , resp.  $\ell$ .
- **bias (vychýlení):**  $\mathbb{E}(\hat{\theta}_n - \theta) \dots \theta$  skutečný parametr,  $\hat{\theta}_n$  náš odhad (náhodná veličina, protože závisí na naměřených datech)
- odhad je **nevychýlený/nestranný/unbiased**:  $bias = 0$
- odhad je **asymptoticky nevychýlený**: bias konverguje k 0, neboli  $\mathbb{E}(\hat{\theta}_n) \rightarrow \theta$
- odhad je **konzistentní**:  $\hat{\theta}_n \xrightarrow{P} \theta$ : pro všechna  $\varepsilon > 0$   $P(|\hat{\theta}_n - \theta| > \varepsilon) \rightarrow 0$

- **MSE (mean square error, střední kvadratická odchylka):**  $\mathbb{E}((\hat{\theta}_n - \theta)^2)$
- Věta:  $MSE = bias^2 + \text{var}(\hat{\theta}_n)$ .

Pro praktickou ukázkou, viz pythonový notebook na webu přednášky <https://iuuk.mff.cuni.cz/~samal/vyuka/PSt1/>.

- Máme náhodný výběr  $X_1, \dots, X_n \sim U(0, \theta)$ .
  - Navrhňte bodový odhad  $\theta$  momentovou metodou.
  - Navrhňte bodový odhad  $\theta$  metodou maximální věrohodnosti.
  - Pro každý z nich zjistěte, zda je nestranný a konzistentní.
  - Pro každý z nich spočítejte střední kvadratickou odchylku (MSE). (Stačí experimentálně na počítači.)
  - Který odhad je lepší? Napadá vás nějaký ještě lepší?
- Máme náhodný výběr  $X_1, \dots, X_n \sim \text{Geom}(p)$ .
  - Navrhňte bodový odhad  $p$  momentovou metodou.
  - Navrhňte bodový odhad  $p$  metodou maximální věrohodnosti.
  - Pro každý z nich zjistěte, zda je nestranný.
- Pro náhodný výběr  $X_1, \dots, X_n \sim \text{Exp}(1/\theta)$  řešte části (a)–(c) jako výše.
  - Navrhňte bodový odhad  $\theta$  momentovou metodou.
  - Navrhňte bodový odhad  $\theta$  metodou maximální věrohodnosti.
  - Pro každý z nich zjistěte, zda je nestranný a konzistentní.
- Máme náhodný výběr  $X_1, \dots, X_n \sim \text{Exp}(\lambda)$ . Zajímá nás pravděpodobnost  $p$ , že  $X > 1$  pro  $X \sim \text{Exp}(\lambda)$ . (Připomeňme, že  $p = e^{-\lambda \cdot 1}$ .)
  - Navrhňte bodový odhad  $p$  (libovolnou metodou), případně několik odhadů.
  - Prozkoumejte jeho vlastnosti.

## Intervalové odhady

- Máme jedno měření  $X \sim N(\mu, 1)$ . (Tj. parametr  $\theta = \mu$ .)
  - Najděte intervalový odhad pro  $\mu$  se spolehlivostí 95 %. (Pro konkrétnost: naměřili jsme  $x = 2.9$ .)
  - Místo jednoho měření jich provedeme  $n$  (pochopitelně nezávislých). Jaký bude teď intervalový odhad pro  $\mu$ ? Pro konkrétnost: naměřili jsme  $x_1, \dots, x_9 = 1.82, 1.00, 2.50, 3.00, 0.50, 2.97, 1.76, 1.35, 3.41$ .
  - Nechť  $X$  má stále střední hodnotu  $\mu$  a rozptyl 1, ale není už nutně normální. Co se změní?

## K procvičení

- Nechť  $X \sim \text{Exp}(\lambda)$  popisuje dráhu, kterou uletí radioaktivní částice, než se rozpadne. Náš přístroj její rozpad (a polohu rozpadu, tj. hodnotu  $X$ ) zachytí, ale jen pokud  $1 \leq X \leq 2$ . Formálně, budeme zkoumat náhodný výběr  $X_1, \dots, X_n \sim F_{X|B}$  pro jev  $B = 1 \leq X \leq 2$ .
  - Navrhňte bodový odhad  $\lambda$  momentovou metodou.
  - Navrhňte bodový odhad  $\lambda$  metodou maximální věrohodnosti.
  - Pro každý z nich zjistěte, zda je nestranný a konzistentní.
- Máme náhodný výběr  $X_1, \dots, X_n \sim \text{Pois}(\lambda)$ .
  - Navrhňte bodový odhad  $\lambda$  momentovou metodou.
  - Navrhňte bodový odhad  $\lambda$  metodou maximální věrohodnosti.
  - Spočítejte střední kvadratickou odchylku (MSE).