11. cvičení z PSt — 28.4.–2.5.2025

Konvoluce

- **Konvoluční vzorec:** Nechť X, Y jsou spojité n.n.v. Pak $S = X + Y$ má hustotu

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x)dx.$$

1. Buďte $X, Y, Z \sim U(0,1)$ nezávislé náhodné veličiny.

(a) Jaké je rozdělení $X + Y$? Určete hustotu (dvěma způsoby) – podle konvolučního vzorce i „podle obrázku“.

(b) Jaké je rozdělení $X + Y + Z$? Pro jednoduchost určete hustotní funkci jen na intervalu $[0, 1]$.

(c) Jak výsledek ověřit samplováním?

Aplikace nerovností a Centrální Limitní Věty

- **Markovova nerovnost:** Nechť n.v. X splňuje $P(X < 0) = 0$ a $E(X) = \mu$. Pak $\forall a \geq 0$ platí

$$P(X \geq a \cdot \mu) \leq \frac{1}{a}.$$

- **Čebyševova nerovnost:** Nechť n.v. X má konečnou střední hodnotu μ a rozptyl σ^2. Pak $\forall a \geq 0$ platí

$$P(|X - \mu| \geq a \cdot \sigma) \leq \frac{1}{a^2}.$$

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

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

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

(a) …pomocí Markovovy nerovnosti.

(b) …pomocí Čebyševovy nerovnosti.

(c) …pomocí CLV, konkrétně pro $n = 32$.

3. 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 $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Odhaduje, že směrodatná odchylka jednoho měření je nejvýše 1 metr.

(a) Jak velké n má volit, aby směrodatná odchylka $\bar{X}_n$ byla nejvýše 1 cm?

(b) Pro jaké n zajistí Čebyševova nerovnost, že $\bar{X}_n$ 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?

4. Odhadněte $\binom{100}{30}$ pomocí CLV. Nápověda: použijte CLV pro odhad $P(29.5 < X < 30.5)$ pro vhodnou n.v. X. Na druhou stranu pro $P(X = 30)$ máme vzorec $\binom{100}{30}/2^{100}$ z binomického rozdělení.

Bodové odhady

- Zkoumáme posloupnost n.n.v. se stejným rozdělením, např. $Geom(\theta)$, $U(0,\theta)$, kde θ je parametr.
- Zapisujeme $X_1, \ldots, X_n \sim F_\theta$, tzv. **náhodný výběr** z F_θ (model s parametrem).
- Naměříme $X_1 = x_1$, ..., chceme odhadnout θ.
- $\hat\theta$ ... nějaká metoda jak odhadnout θ pomocí naměřených dat (hodnot $X_1, \ldots, X_n$), angl. *estimator*
- $m_r(\theta) = \mathbb{E}(X^r)$ pro $X \sim F_\theta$ ... r**-tý moment**, ideální vlastnost rozdělení
- $\widehat{m}_r(\theta) = \frac{1}{n}\sum_{i=1}^n X_i^r$ ... r**-tý výběrový moment**, náhodná veličina, funkce našeho naměřeného vzorku (tj. statistika)
- **Odhad metodou momentů** vyřešíme rovnici $m_1(\theta) = \widehat{m}_1(\theta)$ pro neznámou θ.
- event. soustavu rovnice $m_r(\theta) = \widehat{m}_r(\theta)$ pro $r = 1, 2, \ldots$ podle potřeby.
- $L(\theta; x_1, \ldots, x_n) = P(X_1 = x_1 \,\&\, \ldots \,\&\, X_n = x_n)$ ... pravd. pozorovaných dat závislá na parametru θ.
- nebo $L(\ldots) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$ ... hustota pravděpodobnosti ...
- $\ell(\theta; x_1, \ldots, x_n) = \log L(\ldots)$ ... pro snazší výpočty.
- **Odhad metodou maximální věrohodnosti (Maximal Likelihood)** hledáme θ, pro které je maximální $L(\theta; x_1, \ldots, x_n)$, resp. $\ell(\ldots)$. Obvykle pomocí derivací funkce L, resp. ℓ.

- **bias (vychýlení):** $\mathbb{E}(\hat\theta - \theta)$ ... θ skutečný parametr, $\hat\theta$ 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 $\mathbb{E}(\hat\theta) \to \theta$
- odhad je **konzistentní:** $\hat\theta \xrightarrow{P} 0$: pro všechna $\varepsilon > 0 \;\; P(|\hat\Theta - \theta| > \varepsilon) \to 0$
- **MSE (mean square error, střední kvadratická odchylka):** $\mathbb{E}((\hat\theta - \theta)^2)$
- Věta: $MSE = bias^2 + \mathrm{var}(\hat\Theta)$.

5. Máme náhodný výběr $X_1, \ldots, X_n \sim U(0, \vartheta)$.
(a) Navrhněte bodový odhad ϑ momentovou metodou.
(b) Navrhněte bodový odhad ϑ metodou maximální věrohodnosti.
(c) Pro každý z nich zjistěte, zda je nestranný a konzistentní. (Stačí experimentálně na počítači.)

6. Pro náhodný výběr $X_1, \ldots, X_n \sim Geom(p)$ řešte části (a)–(c) jako výše.

7. Pro náhodný výběr $X_1, \ldots, X_n \sim Exp(1/\theta)$ řešte části (a)–(c) jako výše.

K procvičení

8. 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 $X_1, \ldots, X_n$ jsou nezávislé. Je tedy $X \sim Bin(100, 1/2)$.

(a) Vyjádřete S pomocí distribuční funkce F_X.

(b) Použijte CLV na odhad této pravděpodobnosti.

(c) Případně vyčíslete S vhodným softwarem a srovnejte.

9. 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?

10. Nechť $X \sim Exp(\lambda)$ popisuje dráhu, kterou uletí radioaktivní částice, nechť se rozpadne. Náš přístroj jeji rozpad (a polohu rozpadu, tj. hodnotu X) zachytí, ale jen pokud $1 \le X \le 2$. Formálně, budeme zkoumat náhodný výběr $X_1, \ldots, X_n \sim F_{X|B}$ pro jev $B = 1 \le X \le 2$.

(a) Navrhněte bodový odhad λ momentovou metodou.

(b) Navrhněte bodový odhad λ metodou maximální věrohodnosti.

(c) Pro každý z nich zjistěte, zda je nestranný a konzistentní.