12. cvičení z PSt — 10.–13.5.2023

Bodové odhady

- Zkoumáme posloupnost n.n.v. se stejným rozdělením, např. $Geom(\theta)$, $U(0, \theta)$, kde θ je parametr.
- Zapisujeme $X_1, \dots, X_n \sim F_\theta$, tzv. náhodný výběr z F_θ (model s parametrem).
- Naměříme $X_1 = x_1, \dots,$ chceme odhadnout θ.
- $\hat\Theta \dots$ nějaká metoda jak odhadnout θ pomocí naměřených dat (hodnot $X_1, \dots, X_n$). Angl. *estimator* – jeden získaný odhad je *estimate*, ten značíme $\hat\theta$.

- $m_r(\theta) = \mathbb{E}(X^r)$ pro $X \sim F_\theta \dots$ *r-tý moment*, ideální vlastnost rozdělení
- $\widehat{m}_r(\theta) = \frac{1}{n}\sum_{i=1}^n X_i^r \dots$ *r-tý výběrový moment*, náhodná veličina, funkce našeho naměřeného vzorku (tj. statistika)
- *Odhad metodou momentů* vyřešíme rovnici $m_1(\theta) = \widehat{m}_1(\theta)$ pro neznámou θ.
- event. soustavu rovnice $m_r(\theta) = \widehat{m}_r(\theta)$ 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 θ.
- 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 θ, pro které je maximální $L(\theta; x_1, \dots, x_n)$, resp. $\ell(\dots)$. Obvykle pomocí derivací funkce L, resp. ℓ.

- bias (vychýlení): $\mathbb{E}(\hat\Theta - \theta) \dots \theta$ 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: $\hat\Theta$ konverguje k 0 v pravděpodobnosti: 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: $MSE = bias^2 + \mathrm{var}(\hat\Theta)$.

Pro praktickou ukázku, viz pythonový notebook na webu cvičení.

1. Máme náhodný výběr $X_1, \dots, X_n \sim U(0, \vartheta)$.
 (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.)
 (d) Pro každý z nich spočtěte střední kvadratickou odchylku (MSE). (Stačí experimentálně na počítači.)
 (e) Který odhad je lepší? Napadá vás nějaký ještě lepší?

2. Máme náhodný výběr $X_1, \dots, X_n \sim Geom(p)$.

(a) Navrhněte bodový odhad p momentovou metodou.
(b) Navrhněte bodový odhad p metodou maximální věrohodnosti.
(c) Pro každý z nich zjistěte, zda je nestranný a konzistentní. (Stačí experimentálně na počítači.)

3. Máme náhodný výběr $X_1, \dots, X_n \sim Exp(\lambda)$. Označme $\vartheta = 1/\lambda$.
(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.
(d) Spočtěte střední kvadratickou odchylku (MSE).

4. Máme náhodný výběr $X_1, \dots, X_n \sim Exp(\lambda)$. Zajímá nás pravděpodobnost p, že $X > 1$ pro $X \sim Exp(\lambda)$. (Připomeňme, že $p = e^{-\lambda \cdot 1}$.)
(a) Navrhněte bodový odhad p (libovolnou metodou), případně několik odhadů.
(b) Prozkoumejte jeho vlastnosti.

Intervalové odhady

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

6. Tentokrát vybíráme z rozdělení $N(\mu, \sigma^2)$: μ ani σ neznáme, parametr $\vartheta = (\mu, \vartheta)$. Naměřili jsme hodnoty 8.47, 10.91, 10.87, 9.46, 10.40.
(a) Spočtěte vyběrový průměr a výběrový rozptyl.
(b) Kdybychom věřili, že spočtený výběrový rozptyl je skutečná hodnota σ^2, najděte intervalový odhad pro μ.
(c) Najděte intervalový odhad pro μ použitím Studentova t-rozdělení.

7. Počet emailů za den modelujeme pomocí Poissonova rozdělení $Pois(\lambda)$. První týden v prosinci jsme dostali 34,35,29,31,30 emailů. Najděte pro λ intervalový odhad se spolehlivostí 95 %.

Použijte k tomu poslední metodu z přednášky – tu využívající Studentova rozdělení. (Poissonovo rozdělení sice není normální, ale pro dostatečně vysokou hodnotu λ je normálnímu dost podobné, metoda bude mít spolehlivost blízkou 95 %.)

K procvičení

8. Na webu `https://www.randomservices.org/random/data/Michelson.html` je soupis naměřených hodnot rychlosti světla při slavném Michelsonovu experimentu z roku 1879. Najděte 95%-ní intervalový odhad. (Můžete předpokládat, že chyba měření je normálně rozdělená se střední hodnotou nula.)

9. 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 $X_1, \dots, X_n \sim F_{X|B}$ 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.

10. Máme náhodný výběr $X_1, \dots, X_n \sim Pois(\lambda)$.
(a) Navrhněte bodový odhad λ momentovou metodou.
(b) Navrhněte bodový odhad λ metodou maximální věrohodnosti.
(c) Spočtěte střední kvadratickou odchylku (MSE).