12. cvičení

Bodové odhady

Úloha 1 (Bodový odhad 1)
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.)

d) Pro každý z nich spočtěte střední kvadratickou odchylku (MSE). (Stačí experimentálně na počítači.)

e) Který odhad je lepší? Napadá vás nějaký ještě lepší?

Úloha 2 (Bodový odhad 2)
Máme náhodný výběr $X_1, \ldots, X_n \sim Geom(p)$, jako parametr nás zajímá $\vartheta = p$.

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.)

Úloha 3 (Bodový odhad 3)
Máme náhodný výběr $X_1, \ldots, X_n \sim Exp(\lambda)$. 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. (Stačí experimentálně na počítači.)

d) Spočtěte střední kvadratickou odchylku (MSE).

Intervalové odhady

Úloha 4 (Intervalový odhad 1)
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 $x_1, \ldots, x_9 = 1.82, 1.00, 2.50, 3.00, 0.50, 2.97, 1.76, 1.35, 3.41$.

c) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?

Úloha 5 (Intervalový odhad 2)
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 výběrový 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.

Úloha 6 (Intervalový odhad 3)
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 %.)

Testování hypotéz

Úloha 7 (Testujeme)
Máme jedno měření $X \sim N(\mu, 1)$. Chceme ověřit hypotézu H_0: $\mu = 5$ s hladinou významnosti $\alpha = 5\,\%$.

a) Jaký zvolíme kritický obor – množinu měření, ve které hypotézu zamítneme?

b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude kritický obor pro $\overline{X}_n$?

c) Pokud je ve skutečnosti $\mu = 4$ a máme $n = 10$ měření, jaká je pravděpodobnost, že hypotézu nezamítneme?

d) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?

Úloha 8 (Chybovost)
Podle slibu výrobce bude jeho stroj dělat chyby nejvýše ve $3\,\%$ případů. Z 600 pokusů došlo k chybě v 28 případech. Posuďte slib výrobce (coby nulovou hypotézu) na hladině významnosti $5\,\%$.

a) Počet chyb modelujte přesně, tj. pomocí binomického rozdělení.

b) Počet chyb modelujte přibližně pomocí normálního rozdělení (s vhodným μ, σ^2).

Tahák

- 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.
- $\hat{\Theta}$ …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 $X \sim F_\theta$ … *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 $r = 1, 2, \dots$ podle potřeby.
- $L(\theta; x_1, \dots, x_n) = P[X_1 = x_1 \wedge \dots \wedge X_n = x_n]$ …pravd. pozorovaných dat závislá na parametru θ.
- nebo $L(\dots) = f_{X_1,\dots,X_n}(x_1, \dots, x_n)$ …hustota pravděpodobnosti …
- $\ell(\theta; x_1, \dots, x_n) = \log L(\dots)$ …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): $\mathbb{E}(\hat{\Theta} - \theta)$ …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: $\hat{\Theta}$ 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 $+$ var$(\hat{\Theta})$.

K procvičení

Úloha 9 (Bodový odhad 3')
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.

Úloha 10 (Bodový odhad 4)
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.

Úloha 11 (Bodový odhad 5)
Máme náhodný výběr $X_1, \dots, X_n \sim U(\vartheta, \vartheta + 1)$.

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?

Úloha 12 (Michelsonův experiment)
Na webu `https://www.randomservices.org/random/data/Michelson.html` je soupis naměřených hodnot rychlosti světla při slavném Michelsonovu experimentu z roku 1879. Najděte intervalový odhad se spolehlivostí 95 %. (Můžete předpokládat, že chyba měření je normálně rozdělená se střední hodnotou nula.)

Domácí úkol 12

`https://iuuk.mff.cuni.cz/~chmel/2223/past/`
Pravděpodobnost a statistika 1

Zadáno 17. 5. 2023
Odevzdejte do ∞ (ale realisticky konec zkouškového) přes Poštovní sovu

---

Úloha 1 (Odhad Poissona)
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) – stačí experimentálně na počítači.