11. cvičení

Pravděpodobnost a statistika 1, 30. 4. 2026 `https://iuuk.mff.cuni.cz/~chmel/2526/past/`

Úloha 1 (MLE společně)
Máme náhodný výběr $X_1, \dots, X_n \sim \text{Geom}(p)$, jako parametr nás zajímá $\vartheta = p$. Navrhněte bodový odhad ϑ metodou maximální věrohodnosti.

Úloha 2 (Opět mince)
Hodíme 100krát spravedlivou mincí. Kolikrát nám v průměru padne orel? Pomocí CLV dále odhadněte pravděpodobnost, že padne více než 60 orlů.

Úloha 3
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) Spočtěte $\mathbb{E}(M) = \frac{k(N+1)}{k+1}$. Může se hodit hockey-stick identity: $\sum_{m=k}^{N} \binom{m}{k} = \binom{N+1}{k+1}$. Pak si připomeňte, 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.)**

Úloha 4 (MLE once more)
Máme náhodný výběr $X_1, \dots, X_n \sim N(\mu, 1)$ – data na vstupu $x_1, ..., x_n$ tedy pochází z normální distribuce s neznámým středem μ, ale známou směrodatnou odchylkou 1.

a) Napište věrohodnostní (likelihood) funkci $p_\theta(x)$; později se může hodit pracovat s funkcí $\log p_\theta(x)$, které se říká log-likelihood.

b) Derivací spočtěte $\hat{\mu}_{MLE}$, mělo by vyjít $\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

c) Ověřte, že $\hat{\mu}_{MLE}$ je nestranný.

d) Přesvědčte se, že kdyby směrodatná odchylka nebyla 1, ale byl to jakýkoliv (nám známý) parametr σ^2, $\hat{\mu}_{MLE}$ by vyšlo úplně stejně; volba $\sigma^2 = 1$ jen zjednodušuje výpočet.

Úloha 5 (Nestranný $\neq$ dobrý)
Najděte příklad nějakého nestranného estimátoru (třeba pro problémy z předchozích úloh), který je zjevně hodně špatný.

Tahák

- **Centrální limitní věta:** Označme $Y_n = ((X_1 + \cdots + X_n) - n\mu)/(\sqrt{n} \cdot \sigma)$. Pak $Y_n \xrightarrow{d} N(0, 1)$. Neboli
$$\lim_{n \to \infty} F_{Y_n}(x) = \Phi(x) \quad \text{pro každé } x \in \mathbb{R}.$$
- Na počítání $\Phi(x)$: `https://t.ly/JRQ2`
- 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 $X_1 = x_1$, ..., 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 $\hat{\theta}$.
- $L(\theta; x_1, \dots, x_n) = P[X_1 = x_1 \wedge \cdots \wedge X_n = x_n]$ ... pravd. pozorovaných dat závislá na parametru θ.
- nebo $L(\dots) = f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ ... hustota pravděpodobnosti ...

- $\ell(\theta; x_1, \ldots, x_n) = \log L(\ldots)$ … 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: $\hat{\Theta}$ 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$(\hat{\Theta})$.