NMAI059 Pravděpodobnost a statistika 1
12. přednáška

Robert Šámal

Přehled

Bayesovská statistika

Srovnání dvou přístupů ke statistice

Frekventistický/klasický přístup

- Pravděpodobnost je dlouhodobá frekvence (z 6000 hodů kostkou padla šestka 1026-krát). Je to objektivní vlastnost reálného světa.
- Parametry jsou pevné, neznámé konstanty. Nelze o nich říkat smysluplné pravděpodobnostní výroky.
- Navrhujeme statistické procedury tak, aby měly žádané dlouhodobé vlastnosti. Např. 95 % z našich intervalových odhadů pokryje neznámý parametr.

Bayesovský přístup

- Pravděpodobnost popisuje, jak moc věříme nějakému jevu, jak moc jsme ochotní se vsadit. (Pravděpodobnost, že Thomas Bayes měl 18. prosince 1760 šálek čaje, je 90 %.)
- Můžeme vyslovovat pravděpodobnostní výroky i o parametrech (třebaže jsou to pevné konstanty).
- Spočítáme distribuci ϑ a z ní tvoříme bodové a intervalové odhady, atd.

Bayesovská metoda – základní popis

- neznámý parametr považujeme za náhodnou veličinu Θ
- zvolíme *apriorní distribuci (prior distribution)*, neboli hustotu pravděpodobnosti $f_\Theta(\vartheta)$ nezávislou na datech.
- zvolíme statistický model $f_{X|\Theta}(x|\vartheta)$, který popisuje, co naměříme (s jakou pravděpodobností), v závislosti na hodnotě parametru
- poté, co pozorujeme hodnotu $X = x$, spočítáme *posteriorní distribuci (posterior distribution)* $f_{\Theta|X}(\vartheta|x)$
- z té pak odvodíme, co potřebujeme např. najdeme a, b, aby $\int_a^b f_{\Theta|X}(\vartheta|x)d\vartheta \geq 1 - \alpha$

- $\vartheta = \theta$ malá théta, Θ je velká théta

Bayesova věta

Věta (Bayesova pro diskrétní náhodné veličiny)

X, Θ *jsou diskrétní n.v.*

$$p_{\Theta|X}(\vartheta|x) = \frac{p_{X|\Theta}(x|\vartheta)p_\Theta(\vartheta)}{\sum_{\vartheta' \in Im\Theta} p_{X|\Theta}(x|\vartheta')p_\Theta(\vartheta')}.$$

(sčítance s $p_\Theta(\vartheta') = 0$ *považujeme za 0).*

Věta (Bayesova pro spojité náhodné veličiny)

X, Θ *jsou spojité n.v., které mají hustotu* f_X, f_Θ *i sdruženou hustotu* $f_{X,\Theta}$

$$f_{\Theta|X}(\vartheta|x) = \frac{f_{X|\Theta}(x|\vartheta)f_\Theta(\vartheta)}{\int_{\vartheta' \in \mathbb{R}} f_{X|\Theta}(x|\vartheta')f_\Theta(\vartheta')d\vartheta'}.$$

(sčítance s $f_\Theta(\vartheta') = 0$ *považujeme za 0).*

Bayesovské bodové odhady – MAP a LMS

MAP – Maximum A-Posteriori

Volíme $\hat{\vartheta}$ tak, aby maximalizovalo

- $p_{\Theta|X}(\vartheta|x)$ v diskrétním případě
- $f_{\Theta|X}(\vartheta|x)$ ve spojitém případě

LMS – Least Mean Square

Též metoda podmíněné střední hodnoty.

- Volíme $\hat{\vartheta} = \mathbb{E}\big(\Theta \mid X = x\big)$

Příklad 1

Bayesovský klasifikátor spamů:

- vytvoříme seznam podezřelých slov (money, win, pharmacy, ...)
- N.v. X_i popisuje, zda email obsahuje podezřelé slovo w_i.
- N.v. Θ popisuje, zda email je spam $\Theta = 1$ nebo ne $\Theta = 0$.
- Z předchozích emailů získáme odhady $p_{X|\Theta}$ a p_Θ
- Použijeme Bayesovu větu na výpočet $p_{\Theta|X}$

Příklad 2

Romeo a Julie se mají sejít přesně v poledne. Julie ale přijde pozdě o dobu popsanou náhodnou veličinou $X \sim U(0, \vartheta)$. Parametr ϑ modelujeme náhodnou veličinou $\Theta \sim U(0, 1)$. Co z naměřené hodnoty $X = x$ usoudíme o ϑ?

Příklad 3

Pozorujeme náhodné veličiny $X = (X_1, \dots, X_n)$, předpokládáme $X_i \sim N(\vartheta, \sigma_i^2)$ a ϑ je hodnota náhodné veličiny $\Theta \sim N(x_0, \sigma_0)$. Co z naměřených hodnot $X = x = (x_1, \dots, x_n)$ usoudíme o ϑ?

Příklad 4

Házíme mincí, pravděpodobnost, že padne panna je ϑ. Z n hodů padla panna v $X = k$ případech. Pokud naše apriorní distribuce byla $U(0,1)$, jaká bude distribuce posteriorní?

Střední hodnota a součet čtverců

Věta

Pro libovolnou n.v. Θ *je hodnota* $\mathbb{E}\left(\Theta - \hat{\vartheta}\right)^2$ *nejmenší pro* $\hat{\vartheta} = \mathbb{E}\left(\Theta\right)$.

Podmíněná hodnota dává nejmenší součet čtverců

Věta

Bodový odhad $\hat{\vartheta} = \mathbb{E}(\Theta \mid X = x)$ *je nestranný a má nejmenší možnou hodnotu* $\mathbb{E}(\Theta - \hat{\vartheta})^2$.