

01BAPS

Václav Kůs, Martin Kovanda
opravy Jan Trödler
ilustrace Filip Bár

1. října 2020

Obsah

1	Úvodní motivační úlohy ze spolehlivosti, medicíny, biologie, ekonometrie, experimentální fyziky.	1
2	Princip postačitelnosti, podmíněnosti, věrohodnosti, zastavovací pravidlo sekvenciho principu, vztahy.	5
3	Bayesovský princip, bayesovský úplný model, výhody jeho použití, vztah k ostatním principům.	6
4	Třídy optimálních strategií, užítková funkce a podmínky pro existenci užítkové funkce.	7
5	Bayesovské riziko, aposteriorní hustota, prediktivní bayesovská hustota.	8
6	Asymptotické vlastnosti aposteriorní hustoty a bayesovských bodových odhadů.	9
7	Formy apriorní informace, princip neurčitosti, Jeffreysova hustota, ukázky.	10
8	Konjugované systémy apriorních hustot, princip maximální entropie, limitní aposteriorní hustoty, příklady.	11
9	Nejméně příznivá apriorní rozdělení, souvislost s minimaxním principem rozhodování.	12
10	Přípustnost bayesovského řešení, Steinův efekt pro sféricky symetrická rozdělení, Bergerův jev.	13
11	Grupy transformací, ekvivariantní odhady a jejich bayesovská reprezentace.	14
12	Bayesovské odhady pro absolutní, multi-lineární, (váženou) kvadratickou a 0-1 ztrátovou funkci.	15
13	Hierarchický bayesovský model, empirický Bayes a jeho suboptimalita, aplikace.	16
14	Laplaceova asymptotická expanze, podmínky regularity stochastické aproximace, numerické kvadratury.	17
15	Vzorkování podle důležitosti, Metropolisův algoritmus, variační Bayes.	18
16	Bayesovská strategie testování hypotéz, Bayes faktor, vlastnosti, porovnání s klasickými testy.	19

Předmluva

Materiál byl sestaven na základě poznámek a prezentací Ing. Václava Kůse, Ph.D., kterému bych tímto chtěl poděkovat za rozsáhlou korekci vzniklého materiálu. Zmíněné přednášky proběhly v zimním semestru akademického roku 2020/2021 na Fakultě jaderné a fyzikálně inženýrské ČVUT v Praze. Přednášky nebyly uskutečněny prezenční formou vzhledem k probíhající pandemii Covid-19.

Tento učební text je určen posluchačům 1. ročníku navazujícího magisterského studia navštěvujícím kurs 01BAPS *Bayesovské principy ve statistice*, který je zařazen mezi předměty oborů AMSM. Při sestavování textu se předpokládaly znalosti základů matematiky na úrovni absolvování kurzů 01MAB2-4, 01LAB1-2, 01MIP a 01MAS.

Doporučená literatura:

(1) ...

1 Úvodní motivační úlohy ze spolehlivosti, medicíny, biologie, ekonometrie, experimentální fyziky.

Ve statistice provádíme experimenty, abychom získali nějaký užitek, případně abychom mohli předvídat budoucí děje.

PŘÍKLAD 1.1 (Požár lesa ve Španělsku).

- a) Abychom mohli předvídat budoucí požáry lesa, musíme si nejprve rozmyslet, jestli jsou požáry náhodné, tedy jestli je daná veličina deterministická, nebo stochastická úloha.
- b) V tom případě ale potřebujeme model, který by úlohu popisoval. V rámci toho modelu musíme zkoumat faktory ovlivňující požáry lesa. Těch je však mnoho, a proto musíme vybrat ty relevantní.
- c) Dále potom záleží na struktuře dat. Vezmeme p jako pravděpodobnost, že v lese vznikne požár. Potom logistický model pro tuto pravděpodobnost vypadá jako

$$p = \frac{e^{\alpha_1 h + \alpha_2 t + \alpha_3 x}}{1 + e^{\alpha_1 h + \alpha_2 t + \alpha_3 x}},$$

kde h je vlhost, t teplota a x stupeň péče o les.

- d) Snažíme se nyní odhadnout parametry $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ na základě dostupných dat.
- e) Dále musíme ověřit tento model na reálné situaci.

PŘÍKLAD 1.2 (Capture-Recapture). V rybníce je N ryb. Abychom tento počet zjistili, aniž bychom museli vylovit celý rybník, vylovíme pouze r ryb, označíme je a pustíme zpět do vody. Dále počkáme, až budou ryby rovnoměrně rozmístěny a vylovíme s ryb. Z nich bude x značit počet označených ryb.

Rozdělení x je potom $X \sim f_X(x) = \frac{\binom{r}{x} \binom{N-r}{s-x}}{\binom{N}{s}} \sim \text{Hyp}(N, r, s)$. Nyní nám jde o to odhadnout, jak vypadá \hat{N} jako odhad počtu ryb v rybníce. Po krátké úvaze dospějeme k tomu, že $\hat{N} = \frac{rs}{x}$.

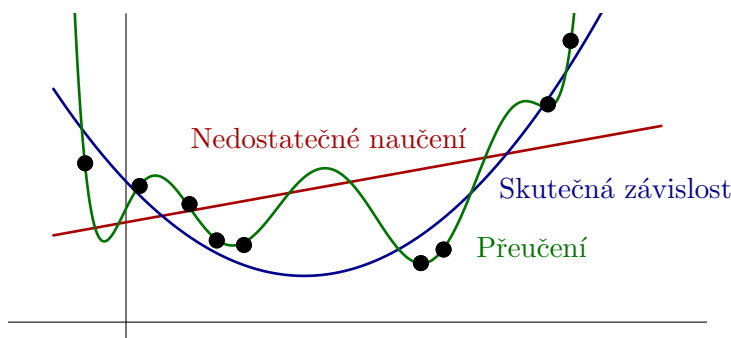
PŘÍKLAD 1.3 (Analýza dat o přežití). Mějme $X \geq 0$ jako čas do poruchy (ale i něčeho pozitivního). Pokud tento problém řešíme stochasticky, máme rozdělení $X \sim F_X, f_X$ a hledáme $\mathbb{E}[X] = \theta$, kterou značíme i jako MTTF (*mean time to failure*). Zavedeme dále **intenzitu poruch** jako $r_X(x) = \frac{f_X(x)}{1 - F_X(x)}$. Úkolem je potom nalézt $\hat{\theta}$.

PŘÍKLAD 1.4 (MEX). Představme si počítačovou síť, kde máme za úkol kontrolovat switch. Obecně o jednotlivých uživateli sítě nevíme nic (nevíme, že zrovna dneska bude někdo stahovat filmy apod.) Označme $X_i \sim \mathcal{N}$ jako průtok dat od i -tého uživatele. Potom chceme mít pod kontrolou pravděpodobnost přetečení switchu, tedy $\mathbb{P}\left(\sum_{i=1}^n X_i > C_s\right) < e^{-\gamma}$ (např.

1 Úvodní motivační úlohy ze spolehlivosti, medicíny, biologie, ekonometrie, experimentální fyziky.

10^{-10} , což je velice nepravděpodobné). Zabýváme se tedy jevy, které se dějí velice zřídka, ale které by mohli mít i vážný dopad, např. u povodní.

PŘÍKLAD 1.5 (Ekonometrie). Víme, že změna mzdy souvisí se změnou ceny komodity, tedy $\Delta M \stackrel{?}{\leftrightarrow} \Delta C$ (např. $\Delta C = \alpha + \beta(\Delta M)^2 + \gamma \ln \Delta M + \varepsilon$). Abychom mohli takový vztah odhadnout, potřebujeme data z minulosti $(\Delta M_i, \Delta C_i)_{i=1}^n$. Z toho pak lze odhadnout $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$. Deterministický přístup by byl použít např. lineární regresi. Pokud však bude součástí zadání předpovědět hodnotu v místě, které nikdy nebylo k dispozici, nemůžeme si být jisti výsledkem, viz obr. 1.1. Cílem je odhadnout ε , což můžeme například pomocí CLT jako $\varepsilon = \sum_{i=1}^n \varepsilon_i \sin \mathcal{AN}$.



Obrázek 1.1: Příklad přeučení a nedostatečného naučení u kvadratických dat.

$$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta \subset \mathbb{R}^k\}$$

$X_i \text{ iid } f$, kde $f_{\mathbf{X}} = \prod f_{X_i}$. $\hat{\theta}$ na základě \mathbf{x} (rozhodnutí o θ). Známujeme parametr θ , kde $\theta \sim \pi(\theta)$.

Věta 1.6 (Bayesova). Mějme $(H_k)_{k=1}^{n,+\infty}$ jako úplný rozklad Ω , $A \in \mathcal{A}$, $\mathbb{P}(A) > 0$. Potom $\forall k \in \mathbb{N}$ platí, že

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\sum_{j=1}^{n,+\infty} \mathbb{P}(A|H_j)\mathbb{P}(H_j)}$$

Důkaz.

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_k)\mathbb{P}(H_k)}{\sum_{j=1}^{n,+\infty} \mathbb{P}(A|H_j)\mathbb{P}(H_j)}$$

□

Věta 1.7. Mějme X, Y jako náhodné veličiny s $f_{X,Y}$. Pak

$$f(y|x) = \frac{f(x|y)g(y)}{\int_{\mathcal{Y}} f(x|y)g(y)dy}.$$

Důkaz.

$$f(y|x) = \frac{f_{X,Y}(x, y)}{f(x)} = \frac{f(x|y)g(y)}{\int_{\mathcal{Y}} f(x|y)g(y)dy}$$

□

1 Úvodní motivační úlohy ze spolehlivosti, medicíny, biologie, ekonometrie, experimentální fyziky.

POZNÁMKA 1.8. Bayesova věta... znáhodnění θ . Máme třídu hustot $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta \subset \mathbb{R}^k\}$, $\theta \sim \pi(\theta)$. Dále $X \sim f(x|\theta) \Rightarrow \mathbf{X} = (X_1, \dots, X_n) \text{ iid } f_{\mathbf{X}}(\mathbf{x}|\theta)$.

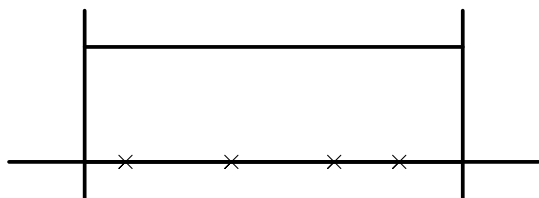
Z toho vyplývá, že $\varphi(\mathbf{x}, \theta) = f(\mathbf{x}|\theta)\pi(\theta)$, $m(\mathbf{x}) = \int \varphi(\mathbf{x}, \theta) d\theta = \int f(\mathbf{x}|\theta)\pi(\theta) d\theta$ jako marginální rozdělení \mathbf{X} .

$$\pi(\theta|\mathbf{x}) \stackrel{B}{=} \frac{\varphi(\mathbf{x}, \theta)}{m(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta) \dots \dots \dots}$$

...tady jsem to nestihl ...

$\pi(\theta) \dots$

- a) objektivní informace, tj. znalost z předchozích úloh (z minulosti),
- b) subjektivní informace (znalost experta, naše znalost, ...),
- c) kombinace a) a b) ($C = \alpha_1 \pi_1(\theta) + \alpha_2 \pi_2(\theta)$), případně více a) nebo více b)
- d) neurčitost (neznalost)



POZNÁMKA 1.9. Tímto způsobem by se dalo pojmut i strojové učení. To bere nějaká trénovací a testovací data, kde na trénovacích datech dochází k učení modelu a na testovacích datech (která nebyla použita při trénování) pak vyhodnocuje, jak moc daný model funguje.

Takto víceméně funguje rozhodování, která děláme.

$$\pi(\theta) \xrightarrow{\text{data}} \pi_n(\theta|\mathbf{x}) \rightarrow \hat{\theta}_B = \mathbb{E}[\pi(\theta|\mathbf{x})]$$

Chtěli bychom, aby byl náš odhad $\hat{\theta}_B(\pi_s(\theta|\mathbf{x}))$ s rostoucím $n \rightarrow +\infty$ stále méně ovlivněn $\pi(\theta)$.

PŘÍKLAD 1.10. –zde by to chtělo asi lepší popis– Představme si, že máme biliárový stůl a ... statistika: máme n šňouchů s rovnoměrným rozdělením. Označme X jako počet neťuků. $X(\omega) = \mathbf{x}$, což jsou data, která máme k dispozici a ptáme se na odhad $\hat{p} = ?$.

—————PIC03—————

- a) Předpokládejme, že 1. hráč je uniformní. Potom $p \sim \pi(p) = 1$ na $(0, 1)$ (podle principu neurčitosti). Potom

$$X \sim \text{Bi}(n, p) = f(x|p).$$

Dále pak

$$\pi(\theta|x) = \frac{f \cdot \pi}{\int_0^1 f \pi dp} = \frac{\binom{n}{x} p^x (1-p)^{n-x} \cdot 1}{\int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \cdot 1 dp} = \text{Beta}(x+1, n-x+1).$$

Z toho vyplývá, že

$$\hat{\theta}_B = \hat{p}_B = \mathbb{E}[p|x] = \int \dots = \frac{x+1}{n+2}.$$

Potom se můžeme ptát, jaké je p , pokud známe \mathbf{x} . Klasický odhad by byl ve tvaru $\hat{p}_{\text{ML}} = \frac{x}{n}$.

1 Úvodní motivační úlohy ze spolehlivosti, medicíny, biologie, ekonometrie, experimentální fyziky.

b) $\pi(\theta) = \pi(p) = \text{Beta}(\alpha, \beta)$ Z toho pak

$$\begin{aligned}\pi(p|x) &= \frac{f \cdot \pi}{\underbrace{\int f \pi}_c} = \frac{1}{c} \binom{n}{x} p^x (1-p)^{n-x} \cdot \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} = \frac{1}{c'} p^{x+\alpha-1} (1-p)^{n-x+\beta-1} = \\ &= \text{Beta}(x+\alpha, n-x+\beta).\end{aligned}$$

Dále

$$\hat{p}_B = \mathbb{E}[\text{Beta}(x+\alpha, n-x+\beta)] = \frac{x+\alpha}{n+\alpha+\beta} \doteq \frac{x}{n} \dots ??? \dots$$

—————PIC04—————

$\pi(\theta) = c \neq 0$ konstantní, takže $\pi(\theta)$ můžeme volit tak, aby $\int_{\Theta} \pi(\theta) = +\infty$.

Definice 1.11 (Nevlastní hustota (apriorní)). Definujeme regulární hustotu (apost.??) vztahem

$$\pi(\theta|x) = \frac{f \cdot \pi}{\int f \pi d\theta}.$$

PŘÍKLAD 1.12. a) Mějme $X \sim f(x|\mu) = \mathcal{N}(\mu, 1)$, kde $\mu \in \mathbb{R}$. Dále nechť $\pi(\mu) = c \neq 0$. Potom

$$\pi(\mu|x) = \frac{f \cdot \pi}{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} c d\mu} = \frac{e^{-\frac{1}{2}(\mu-x)^2}}{\sqrt{2\pi}} \sim \mathcal{N}(x, 1).$$

Dále pak

$$\hat{\mu}_B = \mathbb{E}[\mathcal{N}(x, 1)] = x = \hat{\mu}_{\text{ML}}.$$

b) Nechť $f = \mathcal{N}(0, 1)$ a $\pi(\mu) = \mathcal{N}(0, 10)$.

—————PIC05—————

Pojďme nyní udělat apriorní odhad μ , tedy $\hat{\mu}_{\text{apr.}} = \mathbb{E}[\mathcal{N}(0, 10)] = 0$. $X = x \dots f_X$. Dále

$$\pi(\mu|x) = \frac{1}{c} f \cdot \pi = \frac{1}{c} e^{-\frac{1}{2}(x-\mu)^2} e^{-\frac{1}{20}(\mu-0)^2} = \frac{1}{c} e^{-\frac{11}{20}(\mu-\frac{10}{11}x)^2}.$$

Pro odhad μ pak platí, že

$$\hat{\mu}_B = \mathbb{E}[\pi(\mu|x)] = \int \frac{1}{c} e^{????} d\mu = \frac{1}{c'}$$

s rozdělením $\mathcal{N}(\frac{10}{11}x, \frac{10}{11})$.

HPD (High Posterior Density) region

—————PIC06—————(i s tou rovnicí, tu bych dal do obrázku)—————

Zkoumáme tedy hypotézu $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$. Máme k dispozici $\mathbf{X} \dots T(\mathbf{X})$, kde $T(\mathbf{X}) \sim F_T$. Testujeme tedy $W_\alpha = \{|T(\mathbf{X})| < K_\alpha\}$. Bayesova věta nám říká, že pokud je θ znáhodněný parametr, potom pravděpodobnost, že platí H_0 je rovna ????? (tohle jsem nějak nestihnul).

**2 Princip postačitelnosti, podmíněnosti,
věrohodnosti, zastavovací pravidlo
sekvenčního principu, vztahy.**

3 Bayesovský princip, bayesovský úplný model, výhody jeho použití, vztah k ostatním principům.

4 Třídy optimálních strategií, užitková funkce a podmínky pro existenci užitkové funkce.

5 Bayesovské riziko, posteriorní hustota, prediktivní bayesovská hustota.

6 Asymptotické vlastnosti posteriorní hustoty a bayesovských bodových odhadů.

7 Formy apriorní informace, princip neurčitosti, Jeffreysova hustota, ukázky.

8 Konjugované systémy apriorních hustot, princip maximální entropie, limitní aposteriorní hustoty, příklady.

9 Nejméně příznivá apriorní rozdělení, souvislost s minimaxním principem rozhodování.

10 Přípustnost bayesovského řešení, Steinův efekt pro sféricky symetrická rozdělení, Bergerův jev.

11 Grupy transformací, ekvivariantní odhady a jejich bayesovská reprezentace.

12 Bayesovské odhady pro absolutní, multi-lineární, (váženou) kvadratickou a 0-1 ztrátovou funkci.

13 Hierarchický bayesovský model, empirický Bayes a jeho suboptimalita, aplikace.

14 Laplaceova asymptotická expanze, podmínky regularity stochastické aproximace, numerické kvadratury.

15 Vzorkování podle důležitosti, Metropolisův algoritmus, variační Bayes.

16 Bayesovská strategie testování hypotéz, Bayes faktor, vlastnosti, porovnání s klasickými testy.