01BAPS

Bc. Martin Kovanda
dle přednášek Ing. Václava Kůse, Ph.D.

7. března 2021

Obsah

Předmluva a poděkování

Materiál byl sestaven na základě online přednášek a prezentací Ing. Václava Kůse, Ph.D. 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.

Chtěl bych zde poděkovat svým spolužákům Bc. Filipu Mairingerovi za korektury, opravy a členění skripta, dále Bc. Janu Trödlerovi za opravu přepisu z přednášek a Bc. Filipu Bárovi, který mi pomohl s mnoha obrázky.

Doufám, že Vám skriptum usnadní studium tohoto předmětu, který je velice zajímavý a určitě není na škodu si o něm zjistit více informací, zvláště pokud se zajímáte o strojové učení a umělou inteligenci. Text poměrně kopíruje přednášky, nicméně neobsahuje některé souvislosti a zajímavosti, které na ní byly řečeny. Určitě tedy doporučuji navštěvovat přednášky, především po přečtení dané látky v těchto skriptech, abyste dané souvislosti lépe strávili.

Toto skriptum je určeno posluchačům 1. ročníku navazujícího magisterského studia navštěvujícím kurs 01BAPS *Bayesovské principy ve statistice*, který je zařazen mezi předměty oboru AMSM. Při sestavování textu se předpokládaly znalosti matematiky na úrovni kurzů 01MAB2-4, 01LAB1-2, 01MIP a 01MAS.

1 Úvodní motivační úlohy

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 takové požáry náhodné, tedy zda jde o deterministickou nebo stochastickou úlohu.

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. Vybraný logistický model pro tuto pravděpodobnost vypadá jako

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

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

d) Snažíme se nyní odhadnout parametry $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3$. 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 nejprve pouze r ryb, označíme je a pustíme zpět do vody. Dále počkáme, až budou tyto ryby rovnoměrně rozmístěny a vylovíme s ryb. Z nich bude x značit počet označených ryb. Rozdělení odpovídající náhodné veličiny X je potom $X \sim f_X(x) = \frac{\binom{r}{x}\binom{N-r}{s-x}}{\binom{N}{s}} \sim \mathrm{Hyp}(N, r, s)$. Nyní nám jde o to odhadnout, jak vypadá $\widehat{N}$ jako odhad počtu ryb v rybníce. Po krátké úvaze dospějeme k tomu, že $\widehat{N} = \frac{rs}{x}$.

Příklad 1.3 (Analýza dat o přežití). Mějme $X \geqslant 0$ jako čas do poruchy. Pokud tento problém řešíme stochasticky, máme rozdělení $X \sim \mathrm{F}_X, f_X$ a můžeme hledat hodnotu $\mathbb{E}[X] = \theta$, kterou značíme MTTF (*mean time to failure*). Zavedeme dále **intenzitu poruch** jako $r_X(x) = \frac{f_X(x)}{1-\mathrm{F}_X(x)}$. Úkolem je potom nalézt $\widehat{\theta}$, resp. $\widehat{r}_X(x)$.

Příklad 1.4 (MEX). Představme si počítačovou síť, kde máme za úkol kontrolovat síťový přístupový switch. Obecně o jednotlivých uživatelích sítě nevíme nic (nevíme, že zrovna dneska bude někdo stahovat filmy apod.) Označme X_i jako průtok dat od i-tého uživatele. Potom chceme mít pod kontrolou pravděpodobnost přetečení switche, tedy $\mathbb{P}\left(\sum\limits_{i=1}^{n} X_i > C_X\right) < \mathrm{e}^{-\gamma}$ pro relativně vysoká γ (např. $\gamma \geqslant 10$). Zabýváme se tedy jevy, které se dějí velice zřídka, ale které by mohli mít i vážný dopad, např. analogicky u povodní.

Příklad 1.5 (Ekonometrie). Víme, že změna mzdy souvisí se změnou ceny komodity, tedy $\Delta M \overset{?}{\leftrightarrow} \Delta C$ (např. $\Delta C = \alpha + \beta(\Delta M)^2 + \gamma \ln \Delta M + \varepsilon$). Abychom mohli takový regresní vztah odhadnout, potřebujeme data z minulosti $\left(\Delta M_i, \Delta C_i\right)_{i=1}^{n}$. Z toho pak lze odhadnout regresní koeficienty $\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}$. Pro zdárné použití modelu potřebujeme znát rozdělení ε, což můžeme aproximovat například pomocí CLT jako $\varepsilon = \sum\limits_{i=1}^{n} \varepsilon_i \sim \mathcal{AN}$.

2 Bayesovský princip

Věta 2.1 (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*

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

Důkaz.

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

□

Příklad 2.2 (Steve). https://www.youtube.com/watch?v=HZGCoVF3YvM

Věta 2.3 (Bayesova pro hustoty)**.** *Mějme* X, Y *jako náhodné veličiny s* $f_{X,Y}$. *Pak*

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{\mathcal{Y}} f_{X|Y}(x|y)f_Y(y)\mathrm{d}y}.$$

Postup

Mějme třídu hustot $\mathcal{F} = \left\{f(x,\theta) : \theta \in \Theta \subset \mathbb{R}^k\right\}$ (což je náš statistický model), kde $f_{\mathbf{X}} = \prod_{i=1}^n f_{X_i}$. Chceme určit $\widehat{\theta}$ na základě známých dat $\mathbf{x}$ (rozhodujeme o hodnotě θ).

- Znáhodníme parametr $\theta \sim \pi(\theta)$, kde $\pi(\theta)$ je hustota pravděpodobnosti pro θ.
- Pokud X_i *iid* $f(x|\theta)$, pak $\mathbf{X} = (X_1, ..., X_n)$ *iid* $f_{\mathbf{X}}(\mathbf{x}|\theta)$.
- Z toho vyplývá, že $\varphi(\mathbf{x},\theta) := f(\mathbf{x}|\theta)\pi(\theta)$ a $m(\mathbf{x}) := \int \varphi(\mathbf{x},\theta)\mathrm{d}\theta = \int f(\mathbf{x}|\theta)\pi(\theta)\mathrm{d}\theta$ jsou marginální rozdělení $\mathbf{X}$ a z Bayesovy věty získáme rovnost

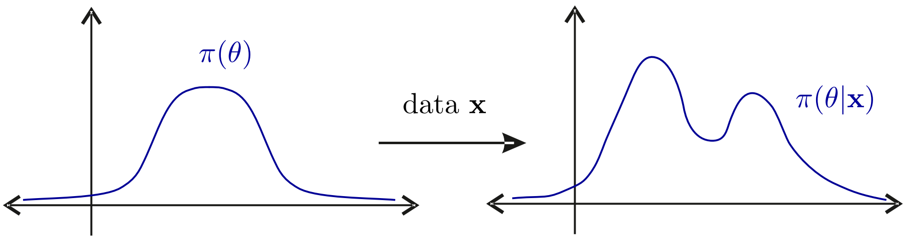
$$\pi(\theta|\mathbf{x}) \overset{B}{=} \frac{\varphi(\mathbf{x},\theta)}{m(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)\mathrm{d}\theta}.$$

Značení

- $\pi(\theta)$ nazýváme **apriorní hustota** (tohle jsme si mysleli o parametru θ dříve) a toto znáhodnění θ je tedy způsob, jak sumarizovat dosažitelnou informaci o θ.
- $\pi(\theta|\mathbf{x})$ nazýváme **aposteriorní hustota** (tohle si o parametru θ myslíme teď, potom co jsme naměřili nová data), viz schéma na obrázku 2.1.

$\pi(\theta)$ může představovat:

a) objektivní informace, tj. znalost z předchozích úloh (z minulosti),
b) subjektivní informace (znalost experta, naše znalost, ...),
c) jejich kombinace, tedy $\pi(\theta) = \alpha_1\pi_1(\theta) + \alpha_2\pi_2(\theta)$, $\alpha_1 + \alpha_2 = 1$,
d) neurčitost (neznalost), tzn. rovnoměrné rozdělení pro θ.



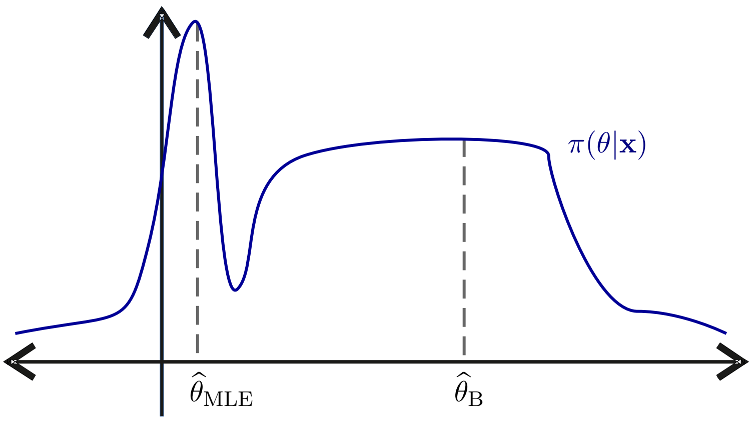
Obrázek 2.1: Naznačení přechodu od apriorní hustoty θ k aposteriorní hustotě pravděpodobnosti $\theta|\mathbf{x}$ při daných datech.

Poznámka 2.4. Tímto způsobem by se dalo pojmout 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 Bayesovsky víceméně funguje rozhodování, která v praxi děláme za účelem odhadu θ:

$$\pi(\theta) \xrightarrow{\text{data}} \pi(\theta|\mathbf{x}) \to \widehat{\theta}_B = \mathbb{E}\left[\pi(\theta|\mathbf{x})\right].$$

Chtěli bychom, aby byl náš odhad $\widehat{\theta}_B$ s rostoucím $n \to +\infty$ stále méně ovlivněn $\pi(\theta)$.



Obrázek 2.2: Možný tvar aposteriorní hustoty pravděpodobnosti. Demonstrace skutečnosti, že Bayesovský MLE odhad nemusí být vždy nejlepší.

Poznámka 2.5. Při určení odhadu parametru θ nemusí být vždy nejvhodnější volbou

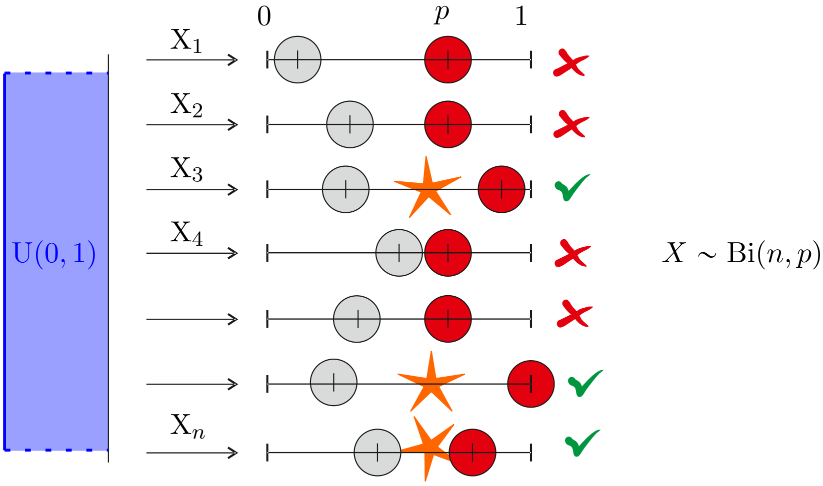
$$\widehat{\theta}_{\text{BMLE}} = \operatorname{argmax}[\pi(\theta|\mathbf{x})].$$

Můžou nastat případy, kdy je střední hodnota aposteriorní hustoty pravděpodobnosti π vhodnější.

Příklad 2.6. Představme si, že máme biliárový stůl délky 1 a dva hráče, kteří hrají mezi sebou. První hráč dostane kouli na pozici $p \in [0,1]$. Druhý hráč to neví a chce poté odhadnout místo, kde se koule prvního hráče nachází, čili chce dostat odhad $\widehat{p}$ na základě n náhodně rozložených šťouchů, o kterých víme, že se buďto dotkly koule, či nikoliv. Máme tedy n šťouchů s rovnoměrným rozdělením. Označme X jako počet neťuků, $X(\omega) = x$, což jsou data, která máme k dispozici a ptáme se na odhad $\widehat{p} =$?

a) Předpokládejme, že 1. hráč je uniformní. Potom $p \sim \pi(p) = \begin{cases} 1, & p \in (0,1), \\ 0, & \text{jinak} \end{cases}$ (podle principu neurčitosti). Potom

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



Obrázek 2.3: Kulečníkový stůl a zjišťování polohy p červené koule.

Dále pak z Bayesovy věty vyjádříme

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

Z toho vypočítáme Bayesovský odhad parametru θ jako

$$\widehat{\theta}_B = \widehat{p}_B = \mathbb{E}\left[\pi(p|x)\right] = \mathbb{E}\left[\mathrm{Beta}(x+1, n-x+1)\right] = \frac{x+1}{(x+1)+(n-x+1)} = \frac{x+1}{n+2}.$$

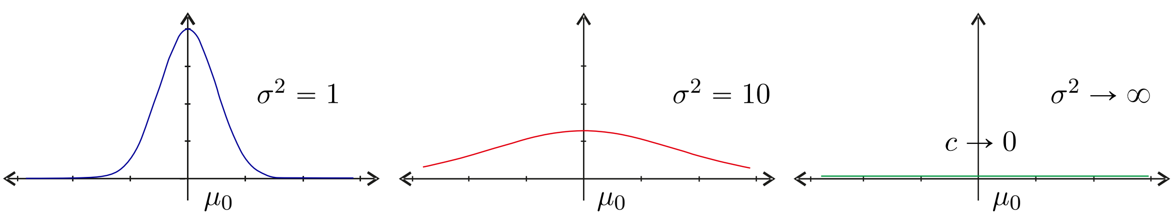
Klasický odhad by byl ve tvaru $\widehat{p}_{\mathrm{ML}} = \frac{x}{n}$.

b) $p \sim \pi(p) = \mathrm{Beta}(\alpha, \beta)$ Z toho opět získáme (c a c' budou značit normovací konstanty)

$$\pi(p|x) = \frac{f \cdot \pi}{\underbrace{\int f \pi \mathrm{d}p}_{c}} = \frac{1}{c}\binom{n}{x} p^x (1-p)^{n-x} \cdot \frac{1}{\mathrm{Beta}(x+1, n-x+1)} p^{\alpha-1}(1-p)^{\beta-1} =$$

$$= \frac{1}{c'} p^{x+\alpha-1}(1-p)^{n-x+\beta-1} = \mathrm{Beta}(x+\alpha, n-x+\beta).$$

Dále $\widehat{p}_B = \mathbb{E}\left[\mathrm{Beta}(x+\alpha, n-x+\beta)\right] = \frac{x+\alpha}{n+\alpha+\beta} \doteq \frac{x}{n} = \widehat{p}_{\mathrm{ML}}$ pro velká n, tedy i velká x.



Obrázek 2.4: Vykreslení změny normálního rozdělení při rostoucím rozptylu.

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

Definice 2.7 (Nevlastní apriorní hustota)**.** Definujeme apriorní nevlastní funkci jako hustotu $\pi(\theta)$ takovou, že

$$\int_\Theta \pi(\theta)\mathrm{d}\theta = +\infty.$$

Uvažujeme však pouze takové nevlastní apriorní hustoty, pro které je

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

stále hustotou pravděpodobnosti, tak jak ji známe. Nevlastní hustota $\pi(\theta)$ tedy stále nese jistou apriorní informaci o θ.

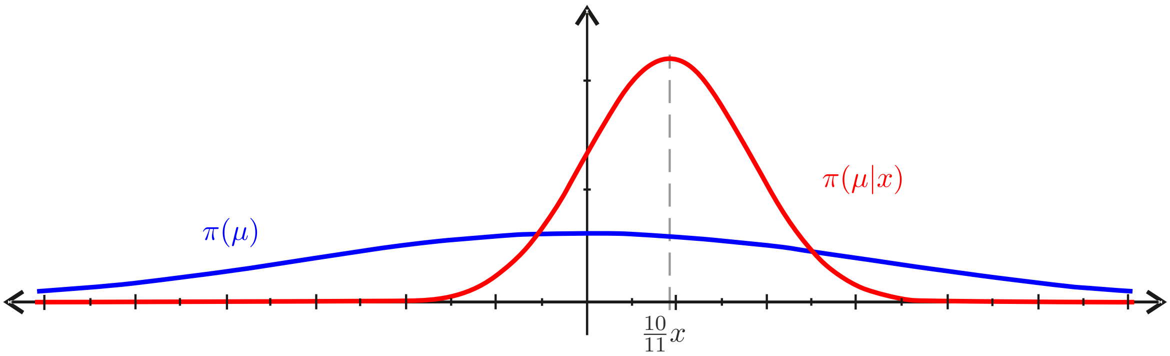
Příklad 2.8. a) Mějme $X \sim f(x|\mu) = \mathcal{N}(\mu, 1)$, kde $\mu \in \mathbb{R}$. Dále nechť $\pi(\mu) = c \neq 0$ na celém $\mathbb{R}$. Potom

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

Dále pak

$$\widehat{\mu}_B = \mathbb{E}\left[\mathcal{N}(x, 1)\right] = x = \widehat{\mu}_{\mathrm{ML}}.$$

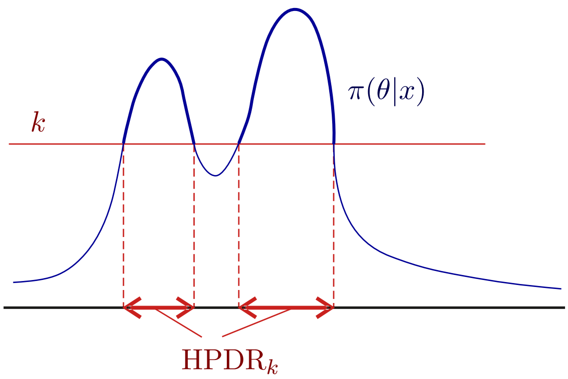
b) Nechť $f(x|\mu) = \mathcal{N}(\mu, 1)$ a $\pi(\mu) = \mathcal{N}(0, 10)$.



Obrázek 2.5: Vykreslení apriorní hustoty a aposteriorní hustoty pravděpodobnosti.

Pojďme nyní udělat apriorní odhad μ, tedy $\widehat{\mu}_{\text{apr.}} = \mathbb{E}\left[\mathcal{N}(0, 10)\right] = 0$. Nyní mějme navíc data x kde $X \sim f(x|\mu) = \mathcal{N}(\mu, 1)$, která je zadaná. Potom (s použitím konstanty c pro zpřehlednění) dostaneme, že

$$\pi(\mu|x) = \frac{1}{c} f(x|\mu) \cdot \pi(\mu) = \frac{1}{c} \mathrm{e}^{-\frac{1}{2}(x-\mu)^2} \mathrm{e}^{-\frac{1}{20}(\mu-0)^2} = \frac{1}{c'} \mathrm{e}^{-\frac{11}{20}(\mu - \frac{10}{11}x)^2} \sim \mathcal{N}\left(\frac{10}{11}x, \frac{10}{11}\right)$$



Obrázek 2.6: HPD Region.

Pro odhad μ pak platí, že

$$\widehat{\mu}_B = \mathbb{E}\left[\pi(\mu|x)\right] = \int \frac{\mu}{c} \mathrm{e}^{-\frac{11}{20}(\mu - \frac{10}{11}x)^2} \mathrm{d}\mu = \frac{10}{11}x.$$

Výhody Bayesovského přístupu

a) Výpočet HPDR (High Posterior Density Region), kde $\text{HPDR}(k) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) > k\}$ pro vhodnou konstantu $k > 0$.

b) Zkoumání hypotézy $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \notin \Theta_0$. Máme k dispozici $\mathbf{X}$ a k ní příslušnou statistiku $T(\mathbf{X})$, kde $T(\mathbf{X}) \sim \mathrm{F}_T$. V klasické statistice bychom nyní testovali hypotézy na základě kritické oblasti $W_\alpha = \left\{|T(\mathbf{X})| < K_\alpha\right\}$. Nyní nám ale Bayesova věta říká, že pokud je θ znáhodněný parametr, tak potom můžeme přímo vyjádřit pravděpodobnost (ať už apriorní, či aposteriorní), že platí H_0, a to následovně

$$\mathbb{P}^\pi(H_0) = \mathbb{P}^\pi(\theta \in \Theta_0), \quad \text{resp.} \quad \mathbb{P}^\pi(H_0|\mathbf{x}) = \mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}),$$

jelikož na parametr θ nahlížíme jako na náhodnou veličinu. Více se Bayesovskému testování bude věnovat poslední kapitola.

3 Další používané principy

Uvedené principy nejsou věty, ale něco, co chceme, aby platilo ve statistické procedůře (zatím ne Bayesovské).

SUFFICIENCY princip: Máme experiment $\mathcal{E}$ závislý na θ, pozorování x, y a nechť je v $\mathcal{E}$ k dispozici postačující statistika T (PS). Víme-li, že $T(x) = T(y)$, chceme, aby závěry o θ na základě x nebo y byly shodné. (T je PS, pokud rozdělení $X|T(X) = t$ nezávisí na θ).

LIKELIHOOD princip: Informace o θ nesená x je zcela obsažena ve věrohodnostní funkci $L(\theta) = f_{\mathbf{X}}(\mathbf{x}, \theta)$. Navíc, pokud máme pozorování x_1 v experimentu $\mathcal{E}_1$ a x_2 v experimentu $\mathcal{E}_2$ taková, že

$$L_1(\theta|x_1) = cL_2(\theta|x_2), \quad \forall\theta \in \Theta,$$

pak chceme, aby závěry o parametru θ v obou experimentech byly shodné. Označíme to $\mathrm{Ev}(\mathcal{E}_1, x_1) = \mathrm{Ev}(\mathcal{E}_2, x_2)$ (Ev jako Evidence).

CONDITIONALITY princip: Tento princip se od prvních dvou liší tím, že chceme, aby něco platilo pro všechna potenciálně naměřená data. Nechť máme dostupné $\mathcal{E}_1, \mathcal{E}_2$. Definujeme $\mathcal{E}^*$ experiment tak, že vybereme náhodně mezi $\mathcal{E}_1 \vee \mathcal{E}_2$ a v tom vybraném naměříme $x_j,\ j = 1 \vee 2$. Chceme, aby $\mathrm{Ev}\big(\mathcal{E}^*, (j, x_j)\big) = \mathrm{Ev}(\mathcal{E}_j, x_j),\ \forall j, \forall x_j$. V prvním (složeném) experimentu tedy zaznamenáváme $j \in \{0, 1\}$ jako index experimentu a x_j jako naměřenou hodnotu a chceme, aby závěr ze složeného experimentu byl stejný, jako kdyby nevybraný experiment vůbec neexistoval.

STOPPING rule: Máme posloupnost experimentů $\mathcal{E}_1, \mathcal{E}_2, ...$ a zastavovací pravidlo τ, které zastavuje posloupnost v bodě $\mathcal{E}_n$. Jsou naměřeny $\mathbf{x} = \big(x_1^{\mathcal{E}_1}, x_2^{\mathcal{E}_2}, ..., x_n^{\mathcal{E}_n}\big)$. Chceme, aby $\mathrm{Ev}\big(\{\mathcal{E}_1, ..., \mathcal{E}_n, \tau\}, \mathbf{x}\big)$ závisela na τ pouze prostřednictvím $\mathbf{x}$, tzn.

$$\mathrm{Ev}\big(\{\mathcal{E}_1, ..., \mathcal{E}_n, \tau_1\}, \mathbf{x}\big) = \mathrm{Ev}\big(\{\mathcal{E}_1, ..., \mathcal{E}_n, \tau_2\}, \mathbf{x}\big),\ \forall \mathbf{x}.$$

BAYESOVSKÝ princip: Taková procedura, která využívá k rozhodnutí o parametru θ aposteriorní hustotu $\pi(\theta|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x}|\theta)\cdot\pi(\theta)}{\int f_{\mathbf{X}}(\mathbf{x}|\theta)\cdot\pi(\theta)\mathrm{d}\theta}$, např. $\widehat{\theta}_{\mathrm{B}} ='' \mathbb{E}\left[\pi(\theta, \mathbf{x})\right]'' = \mathbb{E}^{\pi}(\theta)$.

PŘÍKLAD 3.1. Mějme $\mathcal{E}_j..x_j \in \mathrm{Ran}(X_j)$, kde $X_j \sim f(x_j, \theta),\ \forall j \in 1, 2, ...$ Teoreticky můžeme jít až do nekonečna, ale někdy chceme experiment zastavit, abychom mohli vyhodnotit data. Definujeme tedy τ jako zastavení v bodě n, pokud

$$\mathbf{x} := (x_1, ..., x_n) \in \mathrm{Ran}(X_1) \times \mathrm{Ran}(X_2) \times ... \times \mathrm{Ran}(X_n) \overset{ozn}{=\!=} \mathcal{A}_n.$$

Platí, že

$$L(\theta|\mathbf{x}) \overset{id}{=\!=} \Big(\prod f_{X_j}(x_j, \theta)\Big) \cdot I_{\mathcal{A}_n}(\mathbf{x}) \overset{non\ id}{=\!=\!=} f(x_1|\theta)f(x_2|x_1, \theta)...f(x_n|x_1, x_2, ..., x_{n-1}, \theta)I_{\mathcal{A}_n}(\mathbf{x}).$$

Vidíme, že $L(\theta|\mathbf{x})$ nezávisí na zastavovacím pravidlu τ přímo, ale pouze prostřednictvím $\mathbf{x}$. Tedy pokud platí L princip, potom platí i SR princip.

PŘÍKLAD 3.2. Máme TV seriál. Označme $\theta \in [0, 1]$ jako sledovanost daného dílu. Bylo zjištěno, že 9 diváků seriál sledují a 3 nikoliv (to jsou naše data). Problém je, že nevíme, jakým způsobem byla data naměřena.

$\mathcal{E}_{\mathrm{I}}$: vybráno $n = 12$ lidí, otestováno, získána data (9DIV,3NEDIV). Máme tedy náhodnou veličinu X jako počet diváků z $n = 12$ nezávislých opakování. Z toho plyne, že $X \sim \mathrm{Bi}(12, \theta)$. Máme napozorováno $X(\omega) = x = 9$.

$\mathcal{E}_{\mathrm{II}}$: vybíráme x osob a testujeme tak dlouho, dokud nezískáme 3 nediváky. Při tomto způsobu ale měření probíhá zcela odlišně. Tedy celkový počet otestovaných osob $N \sim \mathrm{NegBi}(3, 1-\theta)$. Napozorováno tedy bylo $N(\omega) = n = 12$.

Vidíme, že $\quad L_{\mathrm{I}}(\theta, \mathcal{E}_{\mathrm{I}}) = c_1\theta^9(1-\theta)^3 \quad \propto \quad L_{\mathrm{II}}(\theta, \mathcal{E}_{\mathrm{II}}) = c_2\theta^9(1-\theta)^3, \quad \forall\theta \in [0,1]$.

Pokud platí LIKELIHOOD princip, pak $\mathrm{Ev}\big(\mathcal{E}_{\mathrm{I}}, (9)\big) = \mathrm{Ev}\big(\mathcal{E}_{\mathrm{II}}, (12)\big)$ (takže vlastně na zastavovacím principu nezáleželo).

Příklad 3.3. Máme laboratoř a v ní dva přístroje:

- 1. přístroj přesný $X_1 \sim \mathcal{N}(\theta, 0.1)$ $[\mathcal{E}_1]$ (vytížený),
- 2. přístroj nepřesný $X_2 \sim \mathcal{N}(\theta, 10)$ $[\mathcal{E}_2]$ (volný).

Rozhodování o θ: 95%-interval spolehlivosti pro θ.

A) osobně, naměříme x_1 v $\mathcal{E}$, pak spočteme délku $(IS_{95\%})$= 0.62 (na 2. přístroji nezáviselo).

B) vyšleme laborantku. která nese data (ale neví, ze kterého přístroje jsou, prostě mohla použít i volný přístroj). Máme model $\Phi = \beta\mathcal{N}(\theta, 0.1) + (1-\beta)\mathcal{N}(\theta, 10)$, kde β vyjadřuje míru obsazenosti 1. přístroje. Pak zjistíme x jako délku$(IS_{95\%})$ s hodnotou 5.19.

Věta 3.4. *$S \wedge C \Leftrightarrow L \Rightarrow SR$. Implikace $L \Rightarrow C$ a $L \Rightarrow S$ jsou důležité, protože B " $\Rightarrow$ "L.*

Důkaz. B" $\Rightarrow$ "L: Mějme $\mathcal{E}_1, \mathcal{E}_2$, x_1, x_2 a předpokládejme, že $L_1(\theta) = cL_2(\theta)$. Ptáme se, jestli je $\mathrm{Ev}(\mathcal{E}_1, x_1) = \mathrm{Ev}(\mathcal{E}_2, x_2)$. Víme, že

$$\underbrace{\pi_1(\theta|x_1)}_{\sim\mathcal{E}_1} = \frac{f_{X_1}(x_1|\theta)\pi(\theta)}{\int f_{X_1}(x_1|\theta)\pi(\theta)\mathrm{d}\theta} = \frac{cf_{X_2}(x_2|\theta)\pi(\theta)}{\int cf_{X_2}(x_2|\theta)\pi(\theta)\mathrm{d}\theta} = \underbrace{\pi_2(\theta|x_2)}_{\sim\mathcal{E}_2}.$$

$L \Rightarrow S$: Nechť T je PS pro exponenciální $\mathcal{E}$ a nechť máme data x_1^0, x_2^0 (nula značí konkrétní výběr, nikoliv obecný), taková, že $T(x_1^0) = T(x_2^0)$. Máme k dispozici Neymannův faktorizační teorém, $X \sim f(x, \theta)$, pak $T(X)$ je PS právě tehdy, když $f(x|\theta) = h(x)g\big(T(x), \theta\big)$, $\forall\theta$. Potom tedy

$$L(\theta|x_1^0) = f(x_1^0, \theta) = h(x_1^0)g\big(T(x_1^0), \theta\big) = \frac{h(x_1^0)}{h(x_2^0)}\underbrace{h(x_2^0)g\big(T(x_2^0), \theta\big)}_{f(x_2^0,\theta)=L(\theta|x_2^0)}, \quad \forall\theta.$$

Z toho vyplývá (dle L principu), že $\mathrm{Ev}(\mathcal{E}, x_1^0) = \mathrm{Ev}(\mathcal{E}, x_2^0)$.

$L \Rightarrow C$: Uvažujeme experiment $\mathcal{E}^*$ tak, že $(\mathrm{I}, X_{\mathrm{I}}) = X^*$, $I = 1 \vee 2$, $X_I = x_1 \vee x_2$. Dále potom

$$\begin{aligned} L^*\big(\theta|(j, x_j)\big) &= \mathbb{P}\big(X^* = (j, x_j)\big) = \mathbb{P}(I = j \wedge X_{I=j} = x_j) = \mathbb{P}(I = j)\mathbb{P}(X_j = x_j) = \\ &= 0.5f(x_j, \theta) = 0.5L_j(\theta, x_j) \end{aligned}$$

pro $\forall\theta$, $\forall x_j$, tzn. $L^* = cL_j$. Potom tedy dle Likelihood principu máme

$$\mathrm{Ev}\big(\mathcal{E}^*, (j, x_j)\big) = \mathrm{Ev}\big(\mathcal{E}_j, x_j\big), \ \forall j, \ \forall x_j.$$

$\square$

4 Třídy optimálních strategií, užitková funkce

4.1 Statistical decision theory

Definice 4.1. Označíme $\mathscr{D}'$ jako **množinu možných rozhodnutí** o θ, případně $\tau(\theta)$. Dále potom $d \in \mathscr{D}'$ je rozhodnutí. $L : \Theta \times \mathscr{D}' \to [0, +\infty)$ nazýváme **loss function** (ztrátová funkce) a $L(\theta, d)$ je **míra ztráty** (shody, neshody), pokud pro θ použijeme rozhodnutí d.
Označme dále $\mathscr{R}$ jako **reward space**, který je spojen s $\mathscr{D}'$, tj. $\forall d \in \mathscr{D}'$ přiřazujeme $r \in \mathscr{R}$ ($d \leftrightarrow r$). Předpokládejme dále, že na $\mathscr{R}$ existuje úplné uspořádání ($\leqslant$) tak, že
A1) $\forall r_1, r_2 \in \mathscr{R},\ r_1 \leqslant r_2 \vee r_2 \leqslant r_1$,
A2) $\forall r_1, r_2, r_3 \in \mathscr{R},\ r_1 \leqslant r_2 \wedge r_2 \leqslant r_3 \ \Rightarrow\ r_1 \leqslant r_3$.
Z toho vyplývá, že nastává právě jedna varianta

$$r_1 < r_2, \quad r_2 < r_1, \quad r_1 = r_2\ (r_1 \sim r_2).$$

Poslední rovnost není myšlena jako číselná rovnost, ale spíše jako ekvivalence (peníze pro nás můžou mít stejnou cenu, jako získané znalosti).
Nad $\mathscr{R}$ definujeme prostor $\mathcal{P}$ pravděpodobnostních distribucí náhodné veličiny r ($r \sim \mathbb{P} \in \mathcal{P}$). Předpokládejme, že na $\mathcal{P}$ existuje úplné uspořádání ($\leqslant$) analogické s A1) a A2) (jinými slovy zvolíme způsob, jakým lze uspořádat pravděpodobnosti, třeba jak moc se odchylují od rovnoměrného rozdělení apod.).

Příklad 4.2 (Motivace). Označme d_i jako investice do i-té společnosti. Na konci roku pak očekáváme dividendy r_i. Máme tedy $(d_i)_1^k$, $(r_i)_1^k$ a celková dividenda $r = \sum_{i=1}^n r_i$ pak mají rozdělení $\mathbb{P} = \mathbb{P}_1 * \ldots * \mathbb{P}_k$ (konvoluce, protože sčítáme náhodné veličiny), které lze aproximovat z CLT.

Definice 4.3. Funkci U na $\mathscr{R}$ nazveme **užitkovou** (utility function), pokud $\forall P, Q \in \mathcal{P}$ platí, že

$$P \leqslant Q \quad \Leftrightarrow \quad \mathbb{E}^P\big[U(r)\big] \leqslant \mathbb{E}^Q\big[U(r)\big],$$

kde r je náhodná veličina. Pokud to jde, můžeme například volit $U(r) = r$.

Poznámka 4.4. Pokud má $\mathcal{P}$ vhodné vlastnosti, pak existuje užitková funkce $U(r)$, která zachovává dané uspořádání $\mathcal{P}$, a pro ztrátu $L(\theta, d) \geqslant 0$ můžeme volit $L(\theta, d) = U(\theta, d) = -\mathbb{E}[U(r)] + c$, kde konstantu c přičítáme proto, abychom se nedostali do záporných čísel. Tím zajistíme, že pokud budeme provádět minimalizaci L, pak provádíme i maximalizaci užitku.

Příklad 4.5 (volba L). Předpověď počasí v Kanadě. Předpovědi jsou ve tvaru "pravděpodobnost, že zítra bude pršet je p". Předpovědi od různých společností chceme nějak porovnat. Sledujeme tedy 365 dní všechny předpovědi a přiřadíme každému předpovídateli jeho použitá $p_1, \ldots, p_N$, kde N je počet různých procentuálních předpovědí v daném roce (tedy $N \leqslant 365$). Definujeme relativní četnost

$$\theta_j = \frac{\#\ \text{dnů, kdy pršelo a byla použita}\ p_j}{\#\ \text{dnů, kdy byla použita}\ p_j}, \quad j \in \widehat{N}.$$

Sestavíme ztrátovou funkci (použil De Groot v roce 1988)

$$L(\boldsymbol{\theta}, \mathbf{p}) = \sum_{j=1}^{N} q_j (p_j - \theta_j)^2 + \sum_{i=1}^{N} q_i \ln q_i,$$

kde $q_j = \frac{\#\text{ použití } p_j}{365}$ a $H(\mathbf{q}) = -\sum_1^N q_j \ln q_j \geqslant 0$ je entropie rozdělení $\{q_j\}_{j=1}^N$. Suma $\sum_{i=1}^N q_i \ln q_i$ pak penalizuje předpovědi, které jsou nevyvážené, protože pro systém neuspořádanosti (rovnoměrné rozdělení $q_j = \frac{1}{N},\ \forall j$) je $H(\mathbf{q})$ maximální.

Člen $\sum\limits_{j=1}^{n} q_j (p_j - \theta_j)^2$ nemusí být nutně na druhou, můžeme brát závorku i v absolutní hodnotě, případně ji umocnit na $\alpha \in (0,2)$ a regulovat tak robustnost $L(\theta, p)$. Tím vším vstupuje do našeho problému tzv. **Apriorno**, musíme totiž dopředu vědět některé informace, třeba $X \sim f(x, \theta)$, tvar $L(\theta, d)$.

5 Rozhodovací principy statistiky

Zatím tedy máme 3 prostory, výběrový prostor $\chi = \{\mathbf{x}\}$ pro data $\mathbf{x}$, parametrický prostor $\Theta = \{\theta\}$, rozhodovací prostor $\mathscr{D}' = \{d\}$ a definujeme na nich několik věcí.

Definice 5.1. Mějme $\chi, \Theta, \mathscr{D}'$ a prostor $\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\}$. Funkci $\delta : \chi \to \mathscr{D}'$ nazýváme **rozhodovací funkce** a platí, že $\delta(\mathbf{x}) = d \in \mathscr{D}'$, kde $\delta \in \mathscr{D}$ jako prvek prostoru rozhodovacích funkcí. Rozhodovací funkce nám tedy říká, jaké rozhodnutí uděláme na základě dat $\mathbf{x}$.

5.1 Metoda A) minimalizace L:

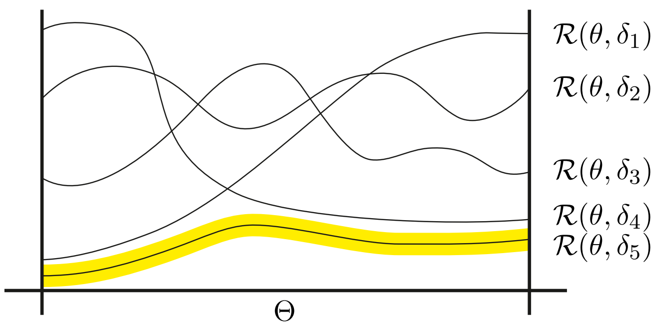
$\delta_L(\mathbf{x}) = \underset{\delta \in \mathscr{D}}{\operatorname{argmin}} L(\theta, \delta(\mathbf{x}))$, $\forall \theta \in \Theta$, $\forall \mathbf{x} \in \chi$. Avšak je obtížné, až nemožné, takto obecně stejnoměrně minimalizovat.

5.2 Metoda B) minimalizace $\mathbb{E}$L

Stejnoměrná minimalizace má spoustu problémů (ne vždy to jde). Pokud se jich chceme zbavit, můžeme například minimalizovat střední ztrátu. Použijeme k tomu rizikovou funkci (risk function) $\mathcal{R} : \Theta \times \mathscr{D}' \to \mathbb{R}^+$, definovanou jako

$$\mathcal{R}(\theta, \delta) = \mathbb{E}_\theta L(\theta, \delta(\mathbf{X})) = \int_\chi L(\theta, \delta(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}|\theta) \mathrm{d}\mathbf{x}.$$

Rizikovou funkci také nazýváme jako *střední ztrátu*. Definujeme $\delta_U = \underset{\delta \in \mathscr{D}}{\operatorname{argmin}} \mathcal{R}(\theta, \delta)$, $\forall \theta \in \Theta$.



Obrázek 5.1: HPD Region (Highest Posterior Density). Pokud by nějaká funkce $\mathcal{R}(\theta, \delta_6)$ náhodou na nějakém intervalu byla menší, než $\mathcal{R}(\theta, \delta_5)$, tak bychom ji jednoduše zahodili.

Ne vždy lze najít stejnoměrně nejlepší řešení. V takovém případě pak přejdeme na prostor $\mathscr{D}_0 \subset \mathscr{D}$, kde již lze najít stejnoměrně nejlepší řešení (typicky vypustíme některé rizikové funkce), viz 5.1. Můžeme tedy optimalizovat na různých prostorech rozhodovacích funkcí, např.

a) Prostor rozhodovacích funkcí, které jsou nestranné – pak vede na UMVUE.

b) $\mathscr{D}_0'$ takový, který je prostorem ekvivariantních rozhodovacích funkcí na modelech invariantních na určitou transformaci (např. posunutí, přeškálování, tzn. tzv. **lokálně-měřítkové modely**).

Problémy této stejnoměrně optimální strategie:

a) δ_1, δ_2 a k nim $\mathcal{R}(\theta, \delta_1), \mathcal{R}(\theta, \delta_2)$ takové, že se kříží – nejsme schopni rozhodnout, která strategie je lepší.

b) Minimalizujeme $\mathbb{E}L$, ale δ nezávisí na $\mathbf{x}$, tedy výběr nejlepšího δ nezávisí na naměřených datech.

c) Na rozmyšlenou je příklad 5.2.

Příklad 5.2. Mějme $X = \begin{cases} \theta - 1, & \mathbb{P} = \frac{1}{2}, \\ \theta + 1, & \mathbb{P} = \frac{1}{2}, \end{cases}$ $\theta \in \mathbb{R}$, $\mathscr{D}' = \mathbb{R}$. Volme $L(\theta, d) = 1 - I_\theta(d)$, kde $I_\theta(d)$ je 1, pokud $d = \theta$ a nula, pokud ne (nazýváme ji "0-1"ztrátovou funkci).

1) Máme $\mathbf{x} = (x_1, x_2)$ data a sestrojíme $\delta_0(\mathbf{x}) = \frac{x_1 + x_2}{2} = \begin{cases} \theta \\ \theta - 1 \\ \theta + 1 \end{cases}$ symetricky kolem θ.

Potom

$$\mathcal{R}(\theta, \delta_0) = \mathbb{E}L(\theta, \delta_0(\mathbf{X})) = 1 - \mathbb{E}I_\theta\big(\delta_0(\mathbf{X})\big) = 1 - \mathbb{P}\big(\delta_0(\mathbf{X}) = \theta\big) = 1 - \mathbb{P}(X_1 \neq X_2) = \frac{1}{2}.$$

To nám říká, že střední ztráta je rovna $\frac{1}{2}$, tzn. že v průměru 50% případů bude příznivých (trefíme se do parametru).

2) Máme data $\mathbf{x} = (x_1, x_2)$ a sestrojíme $\delta_1(\mathbf{x}) = x_1 + 1 = \begin{cases} \theta \\ \theta + 2 \end{cases}$ nesymetricky rozdělené kolem θ. Potom

$$\mathcal{R}(\theta, \delta_1) = \dots = 1 - \mathbb{P}\big(\delta_1(\mathbf{X}) = \theta\big) = 1 - \mathbb{P}(X_1 = \theta - 1) = \frac{1}{2}.$$

Srovnání δ_0 a δ_1 nelze rozhodnout na základě $\mathcal{R}$-strategie.

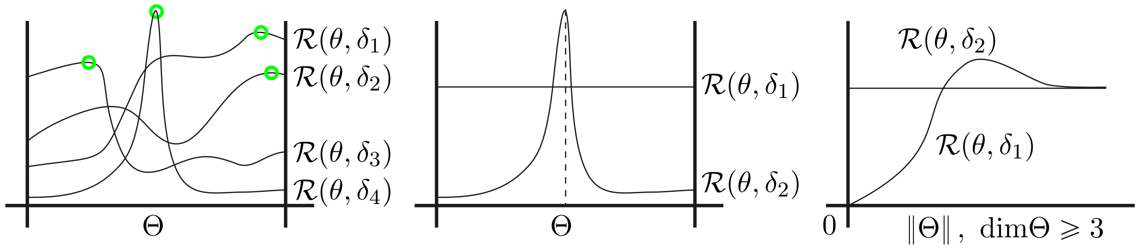
5.3 Metoda C) Strategie MINIMAXní

Definice 5.3. Mějme $\delta \in \mathscr{D}$. Potom definujeme $\mathcal{R}_{\sup}(\theta, \delta) := \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta) \in \mathbb{R}^+$ jako "*maxní*" (*supremální*) rizikovou funkci. Definujeme dále

$$\delta_0 = \operatorname*{argmin}_{\delta \in \mathscr{D}} \mathcal{R}_{\sup}(\theta, \delta) = \operatorname*{argmin}_{\delta \in \mathscr{D}} \sup_{\theta \in \Theta} \underbrace{\mathbb{E}_\theta L\big(\theta, \delta(\mathbf{X})\big)}_{\mathcal{R}(\theta, \delta)}.$$

Definice 5.4. Definujeme **minimaxní riziko** jako $\overline{\mathcal{R}} = \inf_{\delta \in \mathscr{D}} \sup_{\theta \in \Theta} \mathcal{R}(\theta, \delta)$.

Poznámka 5.5. Analogie z teorie her. Máme 2 hráče, kteří mají antagonický vztah (snaží se navzájem poškodit a nezáleží jim na ztrátě). První hráč udělá nějaký tah. Druhý hráč pak udělá tah, který co nejvíc poškodí 1. hráče. První hráč to předvídá a snaží se tedy minimalizovat nejhorší možnou ztrátu.



Obrázek 5.2: 1: zeleně suprema, nejmenší supremum je pro δ_2, 2: lepší je δ_2, protože ač je riziko vysoké, jeho šance je velice malá - toto je nevýhoda minimaxní strategie 3: δ_2 jen lehce překmitne δ_1 a pak se k němu blíží asymptoticky

5.4 Metoda D) Bayesovská riziková funkce

Definice 5.6. Nechť $\Pi = \{\pi\}$ je množina všech apriorních rozdělení pro parametr θ. Definujeme **Bayesovskou rizikovou funkci** $r : \Pi \times \mathscr{D} \to \mathbb{R}_0^+$ vztahem

$$r(\pi, \delta) := \mathbb{E}^\pi\big[\mathcal{R}(\theta, \delta)\big] = \int\limits_\Theta \mathcal{R}(\theta, \delta)\pi(\theta)\mathrm{d}\theta = \int\limits_\Theta \Big(\int\limits_\chi L\big(\theta, \delta(\mathbf{X})\big) f_\mathbf{X}(\mathbf{x}|\theta)\mathrm{d}\mathbf{x}\Big)\pi(\theta)\mathrm{d}\theta.$$

Definujeme dále **Bayesovskou rozhodovací funkci**, za předpokladu existence, jako

$$\delta^\pi := \operatorname*{argmin}_{\delta \in \mathscr{D}} r(\pi, \delta).$$

Číslo $r(\pi) := r(\pi, \delta^\pi)$ nazýváme **Bayesovské (apriorní) riziko**.

Poznámka 5.7. Mějme π fixní. Pak $\delta_1 \leqslant \delta_2$ právě tehdy, když $r(\pi, \delta_1) \leqslant r(\pi, \delta_2)$ jsou uspořádané v $\mathbb{R}^1$. To lze rozšířit i na nevlastní apriorní hustoty (informace), ale pouze pokud $r(\pi) < +\infty$.

Příklad 5.8. Mějme $X \sim \underbrace{\mathcal{N}(\theta, 1)}_{f_X(x|\theta)}$, $\theta \in \mathbb{R}$, $\pi(\theta) = 1$ (konstantní) na $\mathbb{R}$ (nevlastní hustota, tedy nepreferujeme dopředu žádnou hodnotu). Pak

$$\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{\int_{-\infty}^{+\infty} f_X(x|\theta)\pi(\theta)\mathrm{d}\theta} = \frac{1}{c}\mathrm{e}^{-\frac{1}{2}(\theta - x)^2} = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(\theta - x)^2} = \mathcal{N}(x, 1).$$

$$\delta^\pi(x) = \mathbb{E}\left[\pi(\theta|x)\right] = \mathbb{E}\mathcal{N}(x, 1) = x,$$

$$r(\pi) = r(\pi, \delta^\pi) = \mathbb{E}^\pi\mathcal{R}(\theta, \delta^\pi) = \mathbb{E}^\pi\big[\mathbb{E}^f L_2\big(\theta, \delta^\pi(X)\big)\big] \stackrel{def}{=\!=} \mathbb{E}^\pi\big[\mathbb{E}^f\big(\theta - \delta^\pi(X)\big)^2\big] = \mathbb{E}^\pi\mathbb{E}^f(\theta - X)^2 =$$
$$= \mathbb{E}^\pi\big[\mathbb{E}^f(X - \mathbb{E}^f X)^2\big] = \mathbb{E}^\pi(\sigma^2) = \int_{-\infty}^{+\infty} \sigma^2 \cdot \underbrace{\pi(\theta)}_{1}\,\mathrm{d}\theta = +\infty.$$

Definice 5.9. Máme L, $f_\mathbf{X}$, $\pi(\theta)$. Definujeme **aposteriorní Bayesovskou rizikovou funkci** vztahem

$$\varrho(\pi, \delta|\mathbf{x}) := \mathbb{E}^\pi\big[L\big(\theta, \delta(\mathbf{x})\big)|\mathbf{X} = \mathbf{x}\big] = \int_\Theta L\big(\theta, \delta(\mathbf{x})\big)\pi(\theta|\mathbf{x})\mathrm{d}\theta.$$

Definice 5.10 (Metoda E) nic nového)**.** Definujeme-li $\delta^\pi_\varrho(\mathbf{x}) := \operatorname*{argmin}_\delta \varrho(\pi, \delta|\mathbf{x})$ za předpokladu existence pro skoro všechna $\mathbf{x}$, pak tato riziková funkce splývá s Bayesovskou rozhodovací funkcí δ^π.

Poznámka 5.11. Výhoda této definice je, že δ_ϱ^π závisí přímo na datech $\mathbf{x}$. Nevýhoda pak je, že musíme minimalizaci dělat opakovaně pro každá data $\mathbf{x}$.

Věta 5.12. *Máme-li δ_ϱ^π pro s.v. $\boldsymbol{x}$, pak $\delta_\varrho^\pi = \delta^\pi$, tj. je rovna s.v. Bayesovské rizikové funkci $\delta^\pi = \operatorname*{argmin}_\delta r(\pi,\delta)$ podle D strategie.*

Důkaz.

$$\varrho\big(\pi, \delta_\varrho^\pi(\mathbf{x})|\mathbf{x}\big) \leqslant \varrho(\pi, \delta(\mathbf{x})|\mathbf{x}) \text{ pro s.v. } \mathbf{x}, \text{ což víme z definice. Potom ale}$$
$$\mathbb{E}^m \varrho\big(\pi, \delta_\varrho^\pi(\mathbf{x})|\mathbf{x}\big) \leqslant \mathbb{E}^m \varrho(\pi, \delta(\mathbf{x})|\mathbf{x}), \text{ kde } m(\mathbf{x}) = \int_\Theta f(\mathbf{x}|\theta)\cdot\pi(\theta)\mathrm{d}\theta.$$

Nyní rozepíšeme pravou stranu (potom to uděláme i s levou)

$$\mathbb{E}^m \varrho(\pi, \delta(\mathbf{x})|\mathbf{x}) = \int_\chi \Big(\int_\Theta L(\theta,\delta) \underbrace{\pi(\theta|\mathbf{x})}_{\frac{f\pi(\theta)}{\int f\pi(\theta)\mathrm{d}\theta} = \frac{f\pi(\theta)}{m(\mathbf{x})}} \mathrm{d}\theta\Big) m(\mathbf{x})\mathrm{d}\mathbf{x} \overset{Fubini}{=\!=\!=} \int_\Theta \Big(\underbrace{\int_\chi L\big(\theta, \delta(\mathbf{x})\big) f(\mathbf{x}|\theta)\mathrm{d}\mathbf{x}}_{\mathbb{E}^f L = \mathcal{R}(\theta,\delta)}\Big)\pi(\theta)\mathrm{d}\theta =$$
$$= \mathbb{E}^\pi \mathbb{E}^f L = r(\pi,\delta).$$
$$\mathbb{E}^m \varrho(\pi, \delta_\varrho^\pi(\mathbf{x})|\mathbf{x}) = \{\text{Analogický postup}\} = r(\pi, \delta_\varrho^\pi).$$

A tedy
$$r(\pi, \delta_\varrho^\pi) \leqslant r(\pi,\delta),\ \forall\delta \quad \Rightarrow \quad \delta_\varrho^\pi = \delta^\pi.$$

Závěr: D strategie Bayes (δ^π) je rovna E strategii Bayes ($\delta_\varrho^\pi(\mathbf{x})$ pro s.v. $\mathbf{x}$). □

Rozšíření: Máme-li $\delta_\varrho^\pi(\mathbf{x})$ pro s.v. $\mathbf{x}$ a stane se, že $r(\pi) = +\infty$, pak toto řešení δ^π nazveme **Zobecněnou Bayesovskou rozhodovací funkcí**.

Postup při aplikaci Bayesovské strategie

Můžeme postupně iterovat přes nová data. Na začátku máme znalosti o parametru θ z minulosti v podobě $\pi(\theta)$ a nově naměříme data $\mathbf{X} \sim f(\mathbf{x}|\theta)$. Za pomoci úplného Bayesovského modelu můžeme updatovat naše znalosti, přičemž získáme $\pi(\theta|\mathbf{x})$. Z toho můžeme například odhadnout $\widehat{\theta}_\mathrm{B}$, případně interval spolehlivosti apod. Poté nová data označíme za aktuální (aposteriorní hustotu přeznačíme na apriorní) a pokračujeme dále, tedy $\widetilde{\pi}(\theta) = \pi(\theta|\mathbf{x})$. Následně získáme nová data $\widetilde{\mathbf{X}} \sim \widetilde{f}(\widetilde{\mathbf{x}}|\theta)$ a získáme $\widetilde{\pi}(\theta|\widetilde{\mathbf{x}})$...

5.5 Úloha predikce

Zkusme porovnat klasickou a Bayesovskou statistiku v úloze predikce.

Klasická statistika

Předpokládáme $X \sim f(x,\theta)$, aniž bychom měli apriorní informaci, a naměříme data $\mathbf{x} = (x_1, ..., x_n)$ jako realizaci $\mathbf{X} = (X_1, ..., X_n)$. Hledáme predikci (realizaci) X_{n+1}. Pokud jsou $X_1, ..., X_n$ *iid*, pak $X_{n+1} \sim f(x,\theta)$ a stačí nám najít příslušné rozdělení. Proto zkusíme najít odhad závislý na $\mathbf{x}$ ve tvaru $\widehat{\theta} = \widehat{\theta}(\mathbf{x}) = \widehat{\theta}_\mathrm{ML}$. Z toho pak získáme $\widehat{f}(x) = f(x, \widehat{\theta}_\mathrm{ML})$ a z toho $X_{n+1} \sim \widehat{f}(x)$.

Bayesovská statistika

Opět máme $X \sim f(x|\theta)$, data $\mathbf{x}$ a navíc apriorní informaci $\pi(\theta)$. Máme několik možností jak pokračovat.

a) Z toho vypočítáme $\pi(\theta|\mathbf{x})$ a z toho získáme odhad $\widehat{\theta}_{\mathrm{B}} = \mathbb{E}\left[\pi(\theta|\mathbf{x})\right]$. Následně určíme hustotu $\widehat{f}(x) = f(x|\widehat{\theta}_{\mathrm{B}})$ a pokračujeme klasicky s odhadem $X_{n+1} \sim \widehat{f}$. Bayes obecně není invariantní na tranformace, takže nemůžeme říct, že jsme získali Bayesovský odhad hustoty.

b) Můžeme také zkusit odhadnout přímo celou hustotu $\tau(\theta) = f(x|\theta)$, kde za použití apriorní hustoty $\pi(\theta)$ získáme $\widehat{f}_{\mathrm{B}}(x) = \widehat{\tau_{\mathrm{B}}(\theta)}$. Následní můžeme predikovat X_{n+1} jako $X_{n+1} \sim \widehat{f}_{\mathrm{B}}$.

c)

Definice 5.13. Máme ÚBM (Úplný Bayesovský model) pro $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}|\theta)$, apriorní hustotu $\pi(\theta)$ a aposteriorní hustotu $\pi(\theta|\mathbf{x})$. Pak definujeme **Bayesovskou prediktivní hustotu** f_{B}^{PR} vztahem

$$f_{\mathrm{B}}^{\mathrm{PR}}(t|\mathbf{x}) = \int\limits_{\Theta} f_X(t|\theta)\pi(\theta|\mathbf{x})\mathrm{d}\theta = \mathbb{E}^{\pi}[f_X(t|\theta)|\mathbf{x}].$$

Potom $X_{n+1} \sim f_{\mathrm{B}}^{\mathrm{PR}}(x_{n+1}|\mathbf{x})$, z čehož můžeme predikovat

$$\mathbb{P}(X_{n+1} > b) = \int\limits_{b}^{+\infty} f_{\mathrm{B}}^{\mathrm{PR}}\mathrm{d}x_{n+1}.$$

Výhodou c) je to, že obsahuje "plnou" (ve smyslu integrálu) informaci aposteriorní hustoty $\pi(\theta|\mathbf{x})$.

6 Asymptotické vlastnosti aposteriorní hustoty a Bayesovských bodových odhadů

Definice 6.1. Mějme $P, Q \in \mathcal{P}$ a definujeme **totální variaci** $\mathrm{TV}(P,Q) = \|P-Q\|_{\mathrm{TV}} = \sup\limits_{|g|\leqslant 1} \left|\int g(\mathbf{X})\mathrm{d}P - \int g(\mathbf{Y})\mathrm{d}Q\right| \overset{ASR}{=\!=\!=} \left| \begin{array}{ll} p \overset{ozn}{=\!=} \frac{\mathrm{d}P}{\mathrm{d}\mu} & P, Q \ll \mu \\ q = \frac{\mathrm{d}Q}{\mathrm{d}\mu} & \end{array} \right| = \int |p-q|\mathrm{d}\mu = \|p-q\|_{L_1}$.

Definice 6.2. Mějme posloupnost měr $(P_n)_1^{+\infty}$ a limitní míru Q. Řekneme, že P_n konverguje silně ke Q, značíme $P_n \xrightarrow{s} Q$, pokud $\mathrm{TV}(P_n, Q) \to 0$.

Poznámka 6.3. Připomenutí slabé konvergence

$$P_n \xrightarrow{w} Q \;\Leftrightarrow\; \int g(\mathbf{X})\mathrm{d}P_n \to \int g(\mathbf{Y})\mathrm{d}Q, \qquad \forall g \in C_{\mathrm{B}}^{(0)} \text{ spojité a omezené.}$$

Věta 6.4. *$P_n \xrightarrow{s} Q \Rightarrow P_n \xrightarrow{w} Q$ (AN).*

Důkaz. Nechť funkce g je spojitá a omezená, $|g| \leqslant M$, $g^* = \frac{g}{M}$, $|g^*| \leqslant 1$. Pak

$$\left|\int g(\mathbf{X})\mathrm{d}P_n - \int g(\mathbf{Y})\mathrm{d}Q\right| = M\left|\int g^*(\mathbf{X})\mathrm{d}P_n - \int g^*(\mathbf{Y})\mathrm{d}Q\right| \leqslant M \cdot \|P_n - Q\|_{\mathrm{TV}} \overset{\mathrm{LDCT}}{\longrightarrow} 0.$$

□

Věta 6.5 (Scheffé theorem)**.** *Mějme P_n, Q s ASR_μ, $p_n = \frac{\mathrm{d}P_n}{\mathrm{d}\mu}$, $q = \frac{\mathrm{d}Q}{\mathrm{d}\mu}$. Nechť dále hustota pravděpodobnosti $p_n \to q$ s.j. μ. Pak $P_n \overset{s}{\longrightarrow} Q$.*

Důkaz.

$$\left|\int_A (p_n - q)\mathrm{d}\mu\right| \leqslant \int_A |p_n - q|\mathrm{d}\mu \leqslant \int_{\mathbb{R}} |p_n - q|\mathrm{d}\mu = 2\int (p_n - q)^+\mathrm{d}\mu \to 0,$$

kde poslední rovnost platí, protože $\int(p_n - q)\mathrm{d}\mu = 0$, a tedy $\int(p_n-q)^+\mathrm{d}\mu - \int(p_n-q)^-\mathrm{d}\mu = 0$. □

Poznámka 6.6. Z 01MAS víme, že pro $\mathcal{F}_{reg}^{\mathrm{MLE}}$ platí, že
1) $\operatorname{supp} f$ nezávisí na θ,
2) $f \in \mathcal{C}^{(3)}$ vzhledem k θ pro s.v. $x \in \operatorname{supp} f$,
3) $\int f'_\theta = 0$ a $\int f''_{\theta\theta} = 0$,
4) $(\mathbb{I}(\theta))_{i,j} = \mathbb{E}\left[\frac{\partial \log f}{\partial \theta_i} \cdot \frac{\partial \log f}{\partial \theta_j}\right] = -\mathbb{E}\left[\frac{\partial^2 \log f}{\partial \theta_i \partial \theta_j}\right]$, $\mathbb{I}(\theta)$ je PD a konečná pro $\forall \theta \in \Theta$,
5) $\forall \theta_0\ \exists H_{\theta_0}\ \exists M(\mathbf{X}) \in \mathscr{L}_1(P_{\theta_0})$ tak, že $\left\|\frac{\partial^3 \log}{\partial \theta^3}\right\| \leqslant M(x)$ na H_{θ_0}, kde $\mathbb{E}_{\theta_0}[M(X)] < +\infty$.

Věta 6.7 (Bernstein - von Mises)**.** *Předpokládejme ÚBM, $X \sim f(x|\theta)$ vzhledem k míře λ, která je σ-finitní, $\mathcal{F} = \left\{f(x|\theta) : \theta \in \Theta \subset \mathbb{R}^k\right\}$, Θ otevřená, $\theta \sim \pi(\theta)$ vlastní hustota, spojitá a $\pi(\theta) > 0$ (Tedy máme $\varphi(x,\theta) = f \cdot \pi$, $m(\boldsymbol{x}) = \int f \cdot \pi \mathrm{d}\theta$, $\pi(\theta|\boldsymbol{x}) = \varphi/m = \frac{f\cdot\pi}{\int f\cdot\pi\mathrm{d}\theta}$). Nechť dále $\mathcal{F} = \mathcal{F}_{reg}^{\mathrm{MLE}}$. Pak*

$$\left\|\pi(\theta|\boldsymbol{x}) - f_{\mathcal{N}\left(\widehat{\theta}_n, \frac{1}{n}\mathbb{I}^{-1}(\theta_0)\right)}\right\|_{TV} \to 0 \quad \text{pro } \forall \theta_0,$$

kde $\widehat{\theta}_n$ je konzistentní řešení LE_Q, tzn. $\widehat{\theta}_n$ řeší $\frac{\partial \log f}{\partial \theta_j} = 0$, $\forall j \in \widehat{k}$.

Důsledek 6.8. *Mějme $\mathcal{F}_{reg}^{\mathrm{MLE}}$, $\pi(\theta) > 0$ vlastní hustota a spojitá, θ_0 skutečná hodnota parametru θ. Pak $(\theta|\boldsymbol{x}) \sim \mathcal{AN}\big(\widehat{\theta}_n(\boldsymbol{x}), \frac{1}{n}\mathbb{I}^{-1}(\theta_0)\big)$ pro $\forall\theta_0$.*

Závěr: $\sqrt{n}\big((\theta|\boldsymbol{x}) - \widehat{\theta}_n(\boldsymbol{x})\big) \xrightarrow{\mathscr{D}} \mathcal{N}\big(0, \mathbb{I}^{-1}(\theta_0)\big)$, $\forall\theta_0$. *Bayesovská strategie je asymptoticky ekvivalentní MLE.*

Důkaz. $k = 1$: Máme f_X, $(X_j)_1^n$, $f_{X_j}(x_j|\theta)$, $f_n \overset{id}{=} \prod_1^n f_{X_j} = f_n(\mathbf{x}|\theta)$, $\pi(\theta|\mathbf{x}) = \frac{f_n\pi}{\int f_n\pi \mathrm{d}\theta}$. Transformujeme pomocí $g : t = \sqrt{n}\big(\theta|\mathbf{x} - \widehat{\theta}_n\big)$, kde $\widehat{\theta}_n$ je ELE (konzistentní řešení LE_Q), $g^{-1} : \theta|\mathbf{x} = \widehat{\theta}_n + \frac{t}{\sqrt{n}}$, $\mathbb{J}_{g^{-1}} = \frac{1}{\sqrt{n}} > 0$. Potom transformujeme

$$\theta|\mathbf{x} \sim \pi(\theta|\mathbf{x}),$$

$$\sqrt{n}\big(\theta|\mathbf{x} - \widehat{\theta}_n\big) \sim \frac{f_n\big(\mathbf{x}|\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big)\pi\big(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big)\big(\frac{1}{\sqrt{n}}\big)}{\int f_n(\mathbf{x}|\theta)\pi(\theta)\mathrm{d}\theta} = \left| \begin{array}{l} \theta = \widehat{\theta}_n + \frac{s}{\sqrt{n}} \\ \mathrm{d}\theta = \frac{\mathrm{d}s}{\sqrt{n}} \end{array} \right| =$$

$$= \frac{f_n\big(\mathbf{x}|\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big)\pi\big(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big)}{\int f_n\big(\mathbf{x}|\widehat{\theta}_n + \frac{s}{\sqrt{n}}\big)\pi\big(\widehat{\theta}_n + \frac{s}{\sqrt{n}}\big)\mathrm{d}s}.$$

Máme spojité ∂_θ^3, a proto rozvineme pomocí Taylora $\log f_n$ do 2. řádu:

$$\log f_n(\mathbf{x}|\widehat{\theta}_n + \frac{t}{\sqrt{n}}) = \log f_n(\mathbf{x}|\widehat{\theta}_n) + 0^{*} + \frac{1}{2}\frac{t^2}{n}\frac{\partial^2 \log f_n}{\partial\theta^2}\Big|_{\theta=\widehat{\theta}_n} + R_n =$$

$$= \log f_n(\mathbf{x}|\widehat{\theta}_n) + \frac{t^2}{2}\underbrace{\Big(\frac{1}{n}\sum_{j=1}^{n}\frac{\partial^2 \log f_{X_j}(x_j|\widehat{\theta}_n)}{\partial\theta^2}\Big)}_{\xrightarrow[\text{ZVČ}]{\mathbb{P}_{\theta_0}}\mathbb{E}\left[\frac{\partial^2 \log f_X}{\partial\theta^2}\right] = -\mathbb{I}(\widehat{\theta}_n) \wedge \widehat{\theta}_n \xrightarrow{\mathbb{P}_{\theta_0}} \theta_0} + \underbrace{R_n}_{\xrightarrow[(s.j.)]{\mathbb{P}_{\theta_0}} 0}.$$

Z toho potom plyne, že $f_n\big(\mathbf{x}|\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big) = f_n(\mathbf{x}|\widehat{\theta}_n)\mathrm{e}^{-\frac{t^2}{2}\mathbb{I}(\widehat{\theta}_n)} \cdot \widetilde{R}_n$ a tedy

$$\pi(t|\mathbf{x}) = \frac{f_n \mathrm{e}^{-\frac{t^2}{2}\mathbb{I}(\widehat{\theta}_n)} \cdot \widetilde{R}_n \pi\big(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\big)}{f_n \int \mathrm{e}^{-\frac{s^2}{2}\mathbb{I}(\widehat{\theta}_n)}\widetilde{R}_n\pi\big(\widehat{\theta}_n + \frac{s}{\sqrt{n}}\big)\mathrm{d}s} \xrightarrow[(n\to\infty)]{\mathbb{P}_{\theta_0}{}^{\ddagger}} \frac{\mathrm{e}^{-\frac{t^2}{2}\mathbb{I}(\theta_0)} \cdot 1 \cdot \pi(\theta_0)}{\int \mathrm{e}^{-\frac{s^2}{2}\mathbb{I}(\theta_0)} \cdot 1 \cdot \pi(\theta_0)\mathrm{d}s} = \left| \begin{array}{l} u = s\sqrt{\mathbb{I}(\theta_0)} \\ \mathrm{d}s = \frac{\mathrm{d}u}{\sqrt{\mathbb{I}(\theta_0)}} \end{array} \right| =$$

$$= \frac{\sqrt{\mathbb{I}(\theta_0)}}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}\mathbb{I}(\theta_0)} \sim \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\theta_0)}\Big).$$

□

Věta 6.9. *Mějme ÚBM, f_X, $\pi(\theta) > 0$ omezenou a vlastní. Označme $\pi(\theta|\boldsymbol{x}) = \frac{f\cdot\pi}{\int f\cdot\pi\mathrm{d}\theta}$, $\pi_0(\theta|\boldsymbol{x}) = \frac{f\cdot 1}{\int f\cdot 1\mathrm{d}\theta}$, kde $\int f \cdot \pi\mathrm{d}\theta < +\infty$ a $\int f \cdot 1\mathrm{d}\theta < +\infty$. Pak*

$$\|\pi(\theta|\boldsymbol{x}) - \pi_0(\theta|\boldsymbol{x})\|_{\mathrm{TV}} \leqslant \max\Big(\frac{a+b}{1-a}, \frac{a+b+ac}{1+a+b+ac}\Big) + \frac{a(2-a+c)}{1-a} = \varepsilon_{a,b,c},$$

kde $a \in [0,1)$, $b > 0$, $c > 0$, jsou definovány následovně:

1) $\exists A \subset \Theta$ *tak, že* $\int_A \pi_0(\theta|\boldsymbol{x})\mathrm{d}\theta \geqslant 1 - a$,

2) $\inf_A \pi(\theta) = m > 0$,

3) $\sup_A \pi(\theta) \leqslant (1+b)m$,

* $\frac{\partial \log f_n}{\partial\theta}\big|_{\theta=\widehat{\theta}_n} = 0$, protože $\widehat{\theta}_n$ je řešením soustavy věrohodnostních rovnic

‡Záměna limity a integrálu: z MIPu známe větu, která říká, že v případě hustot a integrací dle pravděpodobnostní míry nepotřebujeme Lebesgueovskou majorantu.

4) $\sup\limits_{\Theta\setminus A} \pi(\theta) \leqslant (1+c)m$.

Důkaz. Bez důkazu. □

Důsledek 6.10. *Pokud dokážeme stlačit horní hranici* $\varepsilon_{a,b,c}$ *k nule, tak v* TV *jsou* $\pi(\theta|\boldsymbol{x})$ *a* $\pi_0(\theta|\boldsymbol{x})$ *velmi blízko. Potom tedy* $\widehat{\theta}_{\mathrm{B}}^{\pi} = \mathbb{E}\left[\pi(\theta|\boldsymbol{x})\right]$ *a* $\widehat{\theta}_{\mathrm{B}}^{1} = \mathbb{E}\left[\pi_0(\theta|\boldsymbol{x})\right]$ *jsou blízko. To znamená, že vliv* $\pi(\theta)$ *se ztrácí.*

Poznámka 6.11. $\varepsilon_{a,b,c}$ chceme malé. Potřebujeme a, b malé, c ne příliš velké.

7 Princip neurčitosti

Problém principu neurčitosti spočívá v tom, že máme možnost volit nevlastní apriorní hustotu ve tvaru $\pi(\theta) = c$ na Θ, což znamená, že původně o parametru θ vůbec nic nevíme. Z fyzikální podstaty můžeme třeba vědět jen to, že se parametr nachází na $\mathbb{R}^+$. Uveďme příklad tohoto problému.

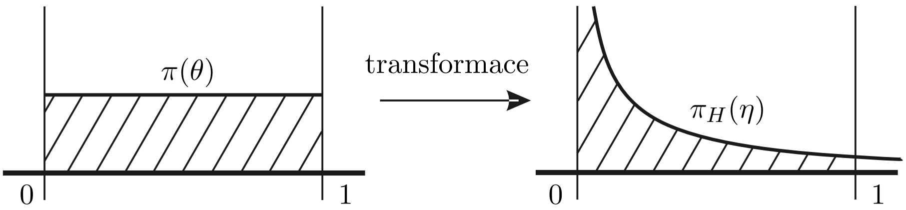
PŘÍKLAD 7.1. Mějme ÚBM, $X \sim f(x|\theta),\ \theta \in (0,1),\ \pi(\theta) = 1$ na $(0,1)$. Podle Bayese získáme aposteriorní hustotu $\pi(\theta|\mathbf{x})$ a z ní můžeme provést odhad $\widehat{\theta}_{\mathrm{B}}(\mathbf{x})$. Představme si nyní, že budeme chtít místo parametru odhadnout např. odmocninu z tohoto parametru. Označme tedy $\eta = \theta^2$, $\theta = \sqrt{\eta}$. Po dosazení do původní hustoty získáme reparametrizaci

$$f(x \mid \theta = \sqrt{\eta}) = f(x|\eta).$$

θ považujeme za znáhodněné (apriorně předpokládáme, že o θ nic nevíme). Transformace hustoty, θ na η probíhá jako

$$\pi_H(\eta) = \pi(\underbrace{\sqrt{\eta}}_{\theta}) \cdot \big|\mathbb{J}_H(\eta)\big| = 1 \cdot \frac{1}{2\sqrt{\eta}} = \frac{1}{2\sqrt{\eta}},$$

kde π_H značíme obecně transformovanou informaci. Je paradoxní, že o θ nemáme konkrétní apriorní informaci a přitom o $\eta = \theta^2$ víme, že je nevlastní (viz obr. 7.1).



Obrázek 7.1: Znázornění transformace $\pi(\theta)$. O θ nemáme konkrétní apriorní informaci, ale o $\eta = \theta^2$ víme, že je nevlastní (není ani integrovatelná na 1). To ukazuje problém principu neurčitosti.

7.1 Jeffreysova hustota

Mějme $X \sim f(x|\theta),\ \theta \in \Theta,\ \theta \sim \pi(\theta),\ \mathcal{F} = \{f(x|\theta)\}$. Zavedeme reparametrizaci $\eta = H^{-1}(\theta)$, tzn. $\theta = H(\eta)$.

$$f_H(x|\eta) = f\big(x|\theta = H(\eta)\big) = f\big(x|H(\eta)\big), \quad \text{takto získáme } \mathcal{F}_H.$$

Z $\pi(\theta)$ vytvoříme $\pi_H(\eta) = \pi\big(H(\eta)\big) \cdot \big|\mathbb{J}_H(\eta)\big|$, kde $\mathbb{J}_H(\eta)$ je Jacobián. Z ÚBM $\{\mathcal{F}, \pi\}$ potom plyne aposteriorní hustota $\pi(\theta|\mathbf{x}) \ \rightarrow\ \widehat{\theta}_{\mathrm{B}}(\mathbf{x}) = \mathbb{E}\,[\pi(\theta|\mathbf{x})]$. Dále z ÚBM $\{\mathcal{F}, \pi\}$ plyne ÚBM $\{\mathcal{F}_H, \pi_H\}$ a z něho aposteriorní hustota $\pi(\eta|\mathbf{x})$ a následně $\widehat{\eta}_{\mathrm{B}}(\mathbf{x}) = \mathbb{E}\,[\pi(\eta|\mathbf{x})]$.
Pozor na $\widehat{\theta}_{\mathrm{B}} \neq H(\widehat{\eta}_{\mathrm{B}})$! To může být problém.

Věta 7.2 (Jeffreys). *Máme ÚBM $\mathcal{F}, \pi(\theta)$, $\eta = H^{-1}(\theta)$, $\theta = H(\eta)$ a předpokládejme, že $\theta \in \Theta \subset \mathbb{R}^k$ a že H je regulární transformace. Nechť $\mathcal{F}$ je regulární systém hustot ($\mathcal{F}^+_{reg}$). Po reparametrizaci máme $\mathcal{F}_H$ a π_H. Volme $\pi(\theta) = \sqrt{\det \mathbb{I}(\theta)}$, kde $\mathbb{I}(\theta)$ je Fishera informační matice pro θ v $\mathcal{F}$. Pak $\pi_H(\eta) = \sqrt{\det \mathbb{I}_H(\eta)}$ v $\mathcal{F}_H$ a platí, že*

$$\int_B \pi(\theta|\boldsymbol{x})\mathrm{d}\theta = \int_{H^{-1}(B)} \pi_H(\eta|\boldsymbol{x})\mathrm{d}\eta, \qquad \forall B \in \mathcal{B}_\Theta \text{ (B je borelovská množina).}$$

Důkaz. Pro $k = 1$: (zbytek ponechán čtenáři ;)) Nejprve tedy spočítáme

$$\int_B \pi(\theta|\mathbf{x})\mathrm{d}\theta = \frac{1}{c}\int_B f(\mathbf{x}|\theta)\pi(\theta)\mathrm{d}\theta = \left| \begin{matrix} \theta = H(\eta) \\ \mathbb{J}_H(\eta) \end{matrix} \right| = \frac{1}{c}\int_{H^{-1}(B)} \underbrace{f\big(\mathbf{x}|H(\eta)\big)}_{f_H(\mathbf{x}|\eta)} \underbrace{\pi\big(H(\eta)\big)\cdot\big|\mathbb{J}_H(\eta)\big|}_{\pi_H(\eta)}\mathrm{d}\eta =$$

$$= \frac{1}{c}\int_{H^{-1}(B)} f_H(\mathbf{x}|\eta)\pi_H(\eta)\mathrm{d}\eta = \int_{H^{-1}(B)} \pi_H(\eta|\mathbf{x})\mathrm{d}\eta,$$

neboť c je v obou případech stejné. Protože platí, že

$$\mathbb{I}_H(\eta) = \mathbb{E}\left[\frac{\partial \log f_H(\eta)}{\partial \eta}\right]^2 = \mathbb{E}\left[\frac{\partial \log f(X|H(\eta))}{\partial \theta}\frac{\partial H(\eta)}{\partial \eta}\right]^2 = \mathbb{I}\big(H(\eta)\big)\cdot \mathbb{J}^2_H(\eta),$$

pak

$$\pi_H(\eta) = \pi\big(H(\eta)\big)\cdot\big|\mathbb{J}_H(\eta)\big| = \sqrt{\det \mathbb{I}\big(H(\eta)\big)}\cdot\big|\mathbb{J}_H(\eta)\big| = \sqrt{\det \mathbb{I}_H(\eta)\mathbb{J}_H^{-2}(\eta)}\cdot\big|\mathbb{J}_H(\eta)\big| =$$
$$= \sqrt{\det \mathbb{I}_H(\eta)}.$$

Získali jsme tedy $\pi(\theta) = \sqrt{\det \mathbb{I}(\theta)}$, což je **Jeffraysova apriorní hustota** (nebo také Jeffreysův princip neurčitosti). □

Tedy pokud obecně o parametru θ nevíme nic, tak máme 2 možnosti. Můžeme na celém oboru zvolit $\pi(\theta) = 1$, ale nebudeme mít zaručenou invarianci na reparametrizaci. Druhou možností je právě zvolit Jeffreysovu apriorní hustotu, která často není vlastní, ale můžeme pak použít Jeffreysovu větu. Pojďme si to ukázat na nějakém příkladu.

Příklad 7.3. Mějme $X \sim f_X(x|\lambda) = \mathrm{Po}(\lambda)$, tzn. $f_{X_j} = \frac{\lambda^{x_j}}{x_j!}\mathrm{e}^{-\lambda}$, $f_{\mathbf{X}}(\mathbf{x}|\lambda) = \frac{\lambda^{\sum x_j}}{\prod x_j!}\mathrm{e}^{-n\lambda}$, $\lambda > 0$.

a) Pokud je λ neznámé, pak $\pi(\lambda) = 1$ na $\mathbb{R}^+$ (nevlastní). (Je jedno, jestli vezmeme rovno 1 nebo nějaké jiné konstantě.)

$$\pi(\lambda|\mathbf{x}) = \frac{f_{\mathbf{X}}\pi}{\underbrace{\int_0^{+\infty} f_{\mathbf{X}}\pi\mathrm{d}\lambda}_{<+\infty}} = \underbrace{\frac{1}{c}\frac{1}{\prod x_j!}}_{\frac{1}{c'}}\lambda^{\sum_{j=1}^n x_j}\mathrm{e}^{-n\lambda} \sim \mathrm{Gamma}\Big(\sum_{j=1}^n x_j + 1, n\Big).$$

Pro připomenutí: $\frac{1}{c'}x^{\alpha-1}e^{-\beta x} = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \sim \mathrm{Gamma}(\alpha,\beta)$
Bayesovský odhad pak bude ve tvaru

$$\widehat{\lambda}_{\mathrm{B}} = \mathbb{E}\Big[\mathrm{Gamma}\Big(\underbrace{\sum x_j + 1}_{\alpha}, \underbrace{n}_{\beta}\Big)\Big] = \frac{\alpha}{\beta} = \frac{1}{n}\sum_{j=1}^n x_j + \frac{1}{n} = \overline{X_n} + \frac{1}{n} \sim \overline{X_n} \text{ (MLE pro velká } n).$$

b) Jeffreys: Zkonstruujeme apriorní informaci jako

$$\pi(\lambda) = \sqrt{\det \mathbb{I}(\lambda)} = \sqrt{\det \mathbb{E}\left[-\frac{\partial^2 \log f_{\mathbf{X}}}{\partial \lambda^2}\right]} = \sqrt{n\cdot\frac{1}{\lambda}} = \sqrt{n}\frac{1}{\sqrt{\lambda}}.$$

Volíme $\pi(\lambda) = \frac{1}{\sqrt{\lambda}}$. V apriorní informaci nehraje konstanta roli, protože se vždy v Bayesově vzorci pokrátí, tedy můžeme vypustit $\sqrt{n}$.

$$\pi(\lambda|\mathbf{x}) = \frac{f(\mathbf{x}|\lambda)\pi(\lambda)}{\underbrace{\int f(\mathbf{x}|\lambda)\pi(\lambda)\mathrm{d}\lambda}_{<+\infty}} = \frac{1}{c'}\lambda^{\sum_{j=1}^{n} x_j - \frac{1}{2}}\mathrm{e}^{-n\lambda} \sim \mathrm{Gamma}\Big(\underbrace{\sum_{j=1}^{n} x_j + \frac{1}{2}}_{\widetilde{\alpha}}, \underbrace{n}_{\widetilde{\beta}}\Big).$$

Bayesovský odhad $\widehat{\lambda}_{\mathrm{B}}^{J} = \mathbb{E}\Big[\mathrm{Gamma}(\widetilde{\alpha}, \widetilde{\beta})\Big] = \frac{\widetilde{\alpha}}{\widetilde{\beta}} = \overline{X_n} + \frac{1}{2n} \sim \overline{X_n}$ (MLE pro velká n).
Pro malá n se odhady v a) a b) liší.

8 Konjugované systémy apriorních hustot, princip maximální entropie

8.1 Limitní aposteriorní hustoty ($\widehat{\theta}_{\mathrm{B}}$)

Definice 8.1. Mějme ÚBM $\{\mathcal{F}, \pi(\theta)\}$ a předpokládejme, že $\pi \in \mathcal{H} := \{\pi(\theta, \boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$. Volme $\lambda_0 \in \partial\Lambda$ z hranice a definujme **limitní apriorní hustotu** $\pi_{\lambda_0}(\theta) = \lim\limits_{\substack{\lambda \to \lambda_0 \\ \lambda \in \Lambda}} c \cdot \pi(\theta, \lambda)$, která bývá často nevlastní, a proto tak píšeme i konstantu c, která umožňuje vyhodit nějaké konstanty z původní hustoty, které by se případně nezlimitovaly na vlastní hodnotu.

Definice 8.2. Dále pak definujeme **limitní aposteriorní hustotu** (za předpokladu záměny integrálu a limity v parametrickém integrálu)

$$\pi_{\lambda_0}(\theta|\mathbf{x}) \overset{ozn}{=\!=} \frac{f\pi_{\lambda_0}(\theta)}{\int_\Theta f\pi_{\lambda_0}(\theta)\mathrm{d}\theta} = \frac{f \lim\limits_{\lambda\to\lambda_0} c\pi(\theta,\lambda)}{\int_\Theta f \lim\limits_{\lambda\to\lambda_0} c\pi(\theta,\lambda)\mathrm{d}\theta} = \lim_{\substack{\lambda\to\lambda_0\in\partial\Lambda \\ \lambda\in\Lambda}} \frac{f\pi(\theta,\lambda)}{\int f\pi(\theta,\lambda)\mathrm{d}\theta} = \lim_{\lambda\to\lambda_0} \pi(\theta,\lambda|\mathbf{x}).$$

Z toho vyplývá, že $\widehat{\theta}_{\mathrm{B}}^{\lim} = \mathbb{E}\left[\pi_{\lambda_0}(\theta|\mathbf{x})\right]$, což často vede na nějakou klasickou statistickou proceduru.

Příklad 8.3. Mějme $X \sim f_X(x|\theta) = \theta\mathrm{e}^{-\theta x} = \mathrm{Exp}(\theta)$, $\theta > 0$. Mějme

$$\mathcal{H} = \{\theta \sim \mathrm{Gamma}(\alpha,\beta),\ \text{tzn. } \pi\big(\theta, \underbrace{(\alpha,\beta)}_{\boldsymbol{\lambda}}\big) = \theta^{\alpha-1}\mathrm{e}^{-\beta\theta}\},\ \Lambda = \mathbb{R}^{2+}.$$

Víme, že $\boldsymbol{\lambda}_0 = (0,0) \in \partial\Lambda$, a proto $\pi_{0,0}(\theta) = \lim\limits_{(\alpha,\beta)\to(0,0)^+} \theta^{\alpha-1}\mathrm{e}^{-\beta\theta} = \frac{1}{\theta}$ (nevlastní).

$$\pi_{(0,0)}(\theta|\mathbf{x}) = \frac{f_X(x|\theta)\cdot\frac{1}{\theta}}{\underbrace{\int_0^{+\infty} f_X(x|\theta)\cdot\frac{1}{\theta}\mathrm{d}\theta}_{<+\infty}} \overset{iid}{=\!=} \frac{\theta^n\mathrm{e}^{-\theta\sum_{j=1}^n x_j}\cdot\frac{1}{\theta}}{c} = \frac{1}{c}\theta^{n-1}\mathrm{e}^{-\theta\sum_{j=1}^n x_j} \sim \mathrm{Gamma}\Big(n, \sum_{j=1}^n x_j\Big),$$

$$\widehat{\theta}_{\mathrm{B}}^{\lim} = \mathbb{E}\Big[\mathrm{Gamma}\Big(n, \sum_{j=1}^n x_j\Big)\Big] = \frac{n}{\sum_{j=1}^n x_j} = (\overline{X_n})^{-1} = \widehat{\theta}_{\mathrm{ML}}.$$

Pokud bychom chtěli dostat pouze $\overline{X_n}$, tak bychom v $\mathcal{H}$ museli namísto Gamma rozdělení použít inverzní Gamma rozdělení.

Příklad 8.4. Volba $\pi(\theta)$ je subjektivní. Příkladem je studie New England Journal of Medicine. Zkoumali, kolik lidí by bylo ochotných jít na operaci kvůli rakovině plic, ale pacienty rozdělili na dvě skupiny, přičemž každé podali informaci o přežití jiným způsobem.

První skupině řekli, že šance na přežití je 68%. V takovém případě by šlo na operaci 44% lidí. Druhé skupině bylo sděleno, že pravděpodobnost úmrtí je 32%. V tom případě by šlo na operaci jen 18% lidí.

V obou případech tak byla pacientům sdělena stejná apriorní informace, ale v obou případech se počet lidí, kteří by šli na operaci, lišil. Vyhodnocení apriorní informace je tedy hodně subjektivní, můžeme získat jiné odhady na základě stejných dat.

Empirical Bayes

Máme $\mathcal{H} = \{\pi(\theta, \lambda) : \lambda \in \Lambda\}$. Spočítá se marginální rozdělení $m_\lambda(\mathbf{x}) = \int_\Theta \underbrace{f(\mathbf{x}|\theta)\pi(\theta, \lambda)}_{\varphi(\mathbf{x},\theta,\lambda)} \mathrm{d}\theta$.

Vezmeme data $\mathbf{x}$ a na základě dat odhadneme klasickou statistickou procedurou (např. pomocí MLE) parametr $\widehat{\lambda}(\mathbf{x})$ v $m_\lambda(\mathbf{x})$ a provedeme Bayesovský princip s apriorní hustotou $\pi(\theta, \widehat{\lambda}_{\mathrm{ML}})$ klasickou statistickou procedurou $\widehat{\lambda}_{\mathrm{ML}}$.
Druhou možností volby je použít princip maximální entropie ve tvaru $H(\pi) = -\int_\Theta \pi(\theta)\log\pi(\theta)\mathrm{d}\theta$. Parametr π potom volíme tak, aby $H(\pi)$ bylo maximální.
Pokud máme omezení ve formě $\mathbb{E}^\pi[g_i(\theta)] = \omega_i,\ \forall i \in K$ (což je počet vazeb), pak maximální $H(\pi)$ dostaneme za použití

$$\pi^*(\theta_i) = \frac{\mathrm{e}^{\sum_{j=1}^n \lambda_j g_j(\theta_i)}}{\sum_i \mathrm{e}^{\sum_{j=1}^n \lambda_j g_j(\theta_i)}},\ \forall i,$$

kde λ_j jsou Lagrangeovy multiplikátory.
Pro spojitý model musíme zavést tzv. *Kulback-Leiblerovu vzdálenost* vztahem

$$\mathrm{KL}(\pi, \pi_0) = \mathbb{E}^\pi\Big[\log\frac{\pi_0}{\pi}\Big] = \int \pi\log\frac{\pi_0}{\pi}\mathrm{d}\theta.$$

Pokud $\pi_0 = 1$, pak $\mathrm{KL}(\pi, 1) = -\int \pi\log\pi = H(\pi)$. Naše volba π tedy závisí na π_0 (za nejméně informativní rozdělení), což je opět velice subjektivní volba.

8.2 Konjugované rodiny

Definice 8.5 (Konjugované systémy apriorních hustot)**.** Mějme systém hustot $\mathcal{F} = \{f(\mathbf{x}|\theta) : \theta \in \Theta \subset \mathbb{R}^k\}$ a označme $\mathcal{H}$ systém apriorních hustot pro θ. Pak $\mathcal{H}$ nazveme **konjugovaný systém apriorních hustot** (*Conjugated Family* = **CF**) pro $\mathcal{F}$, pokud $\pi(\theta|\mathbf{x}) \in \mathcal{H}$.

Věta 8.6. *Mějme $\mathcal{F}_n = \{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = \prod_{j=1}^n f_{X_j}(x_j|\theta)\}$ id a předpokládejme, že $T_n(\boldsymbol{X})$ je postačující statistika pro $\mathcal{F}_{\boldsymbol{X}}$, tzn. dle Neymannova faktorizačního kritéria*

$$f_{\boldsymbol{X}}(\boldsymbol{y}|\theta) = h_n(\boldsymbol{y})g_n\big(\theta, T_n(\boldsymbol{y})\big),\ \forall\theta, \forall\boldsymbol{y}, \forall n.$$

Pak platí, že

$$\mathcal{H} = \left\{\frac{g_m(\boldsymbol{\theta}, \boldsymbol{t})}{\int_\Theta g_m(\boldsymbol{\theta}, \boldsymbol{t})\mathrm{d}\theta} : \forall m \in \mathbb{N}, \forall\boldsymbol{t} \in S_m,\ \textit{kde}\ S_m = \{\boldsymbol{t} : \exists\boldsymbol{y},\ \textit{že}\ T_m(\boldsymbol{y}) = \boldsymbol{t}\}\right\}$$

je konjugovaný systém apriorních hustot.

Důkaz. Volíme apriorní hustotu $\pi(\theta) \in \mathcal{H}$, tzn. $\pi(\theta) \sim g_m(\theta, \mathbf{t})$. Z toho vyplývá, že $\exists\mathbf{y}$ tak, že $T_m(\mathbf{y}) = \mathbf{t}$. Dále platí, že

$$\pi(\theta|\underbrace{\mathbf{x}}_{\dim n}) \sim f_n(\mathbf{x}|\theta)\cdot g_m(\theta, \mathbf{t}) \sim \underbrace{g_m\big(\theta, T_m(\mathbf{y})\big)\cdot h_m(\mathbf{y})}_{\mathrm{NFK}\Rightarrow f_m(\mathbf{y}|\theta)}\cdot f_n(\mathbf{x}|\theta) = f_m(\mathbf{y}|\theta)\cdot f_n(\mathbf{x}|\theta) =$$

$$= \prod_{j=1}^m f(y_j|\theta)\prod_{j=1}^n f(x_j|\theta) = \prod_{j=1}^{m+n}\underbrace{f_{Z_j}(z_j|\theta)}_{\mathbf{z}=(\mathbf{y},\mathbf{x})} = f_{m+n}(\mathbf{z}|\theta) \overset{\mathrm{NFK}}{=\!=} \underbrace{h_{n+m}(\mathbf{z})}_{\mathrm{konst}}\cdot g_{n+m}\big(\theta, \underbrace{T_{n+m}(\mathbf{y})}_{\mathbf{t}\in S_{n+m}}\big) \sim g_{n+m}(\theta, \mathbf{t})$$

Z tohoto důkazu jsme dostali přirozenou CF. Vhodným rozšířením na $m \in \mathbb{R}$ vznikne obvyklá CF. $\square$

Příklad 8.7. Mějme $X \sim f_X(x|\theta) = \lambda e^{-\lambda x} = \mathrm{Exp}(\lambda)$, $\lambda > 0$. Potom

$$\mathcal{F}_n:\ f_n(\mathbf{x}|\underbrace{\theta}_{\lambda}) = \lambda^n e^{-\lambda\sum_{j=1}^n x_j} = \left| \begin{array}{c} ozn.\ T_n(\mathbf{x}) = \sum_{j=1}^n x_j \\ \text{je PS pro } \mathcal{F}_n \end{array} \right| = 1\cdot g_n\big(\lambda, T_n(\mathbf{x})\big).$$

Označme $\sum_{j=1}^n x_j = t$. Potom

$$\mathcal{H} = \Big\{\frac{g_m(\lambda,t)}{c} : m\in\mathbb{N},\ t\in S_m, \ldots\Big\} = \Big\{ \underbrace{\lambda^m \frac{e^{-\lambda t}}{c}}_{\mathrm{Gamma}(m+1,t)} : m\in\mathbb{N},\ t\in S_m = \{\mathbf{t} = \sum_{j=1}^m x_j\} = \mathbb{R}^+\Big\},$$

čímž jsme získali přirozenou CF. To lze opět rozšířit na obvyklou CF: $\mathrm{Gamma}(\alpha,\beta) = \mathcal{H}_{obv}$, $\alpha,\beta > 0$ (rozšíření v α, $\mathbb{N}\to\mathbb{R}^+$).
Jelikož $X\sim\mathrm{Exp}(\lambda)$, $\pi(\lambda) = \mathrm{Gamma}(\alpha,\beta)$, pak

$$\pi(\lambda|\mathbf{x}) = \frac{\lambda^n e^{-\lambda\sum x_j}\frac{1}{c}\lambda^{\alpha-1}e^{-\lambda\beta}}{\int_0^{+\infty}\lambda^n e^{-\lambda\sum x_j}\frac{1}{c}\lambda^{\alpha-1}e^{-\lambda\beta}\mathrm{d}\lambda} = \frac{\lambda^{n+\alpha-1}e^{-\lambda(\sum x_j+\beta)}}{c} \sim \mathrm{Gamma}\Big(n+\alpha, \sum_{j=1}^n x_j+\beta\Big) \in \mathcal{H}_{\mathrm{obv}}.$$

Příklad 8.8 (Příklady CF). a) $X\sim\mathrm{Bi}(n,p) \Rightarrow CF\ \mathcal{H} = \{\pi(p)\sim\mathrm{Beta}(\alpha,\beta)\}$,
b) $X\sim\mathrm{Poisson}(\lambda) \Rightarrow$ CF $\mathcal{H} = \{\pi(\lambda)\sim\mathrm{Gamma}(\alpha,\beta)\}$,
c) $X\sim U(0,\theta)$, $\theta\in\mathbb{R}^+ \Rightarrow$ CF $\mathcal{H} = \{\pi(\theta)\sim\mathrm{Pareto}(a,b)\}$, kde Paretovo rozdělení je mocninného typu $\theta\sim\frac{a}{b}\big(\frac{b}{\theta}\big)^{a+1}$, $\theta\geqslant b$, které má těžké chvosty.

d) 1) $X\sim\mathcal{N}(\mu,\sigma_0^2)$ a známe $\sigma_0>0$. Pak postačující statistika je ve tvaru $T_n(\mathbf{X}) = \sum_{j=1}^n X_j \in \mathbb{R}$. Potom

$$\mathcal{H}_{\mathrm{CF}_{obv}} = \big\{\pi(\mu)\sim\mathcal{N}(a,b^2),\ a\in\mathbb{R}, b>0\big\}$$

a pomocí parametrů a,b ladíme tvar apriorní informace.

2) $X\sim\mathcal{N}(\mu_0,\sigma^2)$ a známe μ_0. Pak postačující statistika je ve tvaru $T_n(\mathbf{X}) = \sum_{j=1}^n (X_j-\mu_0)^2$, protože

$$f_{\mathbf{X}} = \underbrace{\frac{1}{c}e^{-\frac{1}{2\sigma^2}\sum_{j=1}^n (X_j-\mu_0)^2}}_{g_n(t,\sigma^2)},$$

a proto $\mathcal{H}_{\mathrm{obv}}$ pro parametr $\underbrace{\frac{1}{\sigma^2}}_{\theta} = \{\pi(\theta)\sim\mathrm{Gamma}(\alpha,\beta)\}$.

3) $X\sim\mathcal{N}(\mu,\sigma^2)$ a neznáme nic. Potom $T_n(\mathbf{X}) = \Big(\sum_{j=1}^n X_j, \sum_{j=1}^n X_j^2\Big) \in \mathbb{R}\times\mathbb{R}_0^+$ je postačující statistika, a proto $\mathcal{H}_{\mathrm{obv}}$ pro $(\mu,\frac{1}{\sigma^2})$ definujeme tak, že $\pi\big(\mu|\frac{1}{\sigma^2}\big)\sim\mathcal{N}$ a $\pi\big(\frac{1}{\sigma^2}\big)\sim\mathrm{Gamma}$. Z toho pak

$$\Big\{\pi\big(\mu,\frac{1}{\sigma^2}\big) = \pi\big(\mu|\frac{1}{\sigma^2}\big)\pi\big(\frac{1}{\sigma^2}\big) = \underbrace{\mathcal{N}\cdot\mathrm{Gamma}}_{\text{normální-gamma rozdělení}}\Big\}.$$

Poznámka 8.9 (Směs z CF). Pokud CF nevyhovuje naší představě, můžeme zkusit vybrat směs (váženou kombinaci) jako $\pi(\theta) = \sum_{i=1}^l w_i\pi_i(\theta)$, kde $\pi_i\in\mathcal{H}_{\mathrm{CF}}$ a $w_i>0$ jsou váhy splňující podmínku $\sum_{i=1}^l w_i = 1$. Pak můžeme vyjádřit aposteriorní hustotu jako

$$\pi(\theta|\mathbf{x}) = \frac{f\pi}{\int f\pi} = \frac{1}{c}\sum_{i=1}^l w_i f\pi_i(\theta) = \frac{1}{c}\sum_{i=1}^l w_i c_i\frac{f\pi_i}{c_i} = \sum_{i=1}^l \underbrace{\frac{w_i c_i}{c}}_{w_i'}\pi_i(\theta|\mathbf{x}) = \sum_{i=1}^l w_i'\underbrace{\pi_i(\theta|\mathbf{x})}_{\mathcal{H}_{\mathrm{CF}}},$$

kde $c_i = \int f\pi_i\mathrm{d}\theta$. Můžeme tedy spojit několik rozdělení do jednoho složitějšího a získat tím daleko hodnotnější informaci.

9 Nejméně příznivá apriorní rozdělení

9.1 Propojování Bayesovské statistiky s klasickou statistikou (nejméně příznivé rozdělení)

V předchozích kapitolách jsme měli **parametrický prostor** $\theta \in \Theta$, případně $\tau(\theta) \in \mathcal{T}$, dále **rozhodovací prostor** $\mathscr{D}' \ni d$, **ztrátovou funkci** $L(\theta, d)$, kterou se snažila minimalizovat rozhodovací funkce δ_L, $\forall\theta, \forall\mathbf{x}$. Měli jsme také rizikovou funkci $\mathcal{R}(\theta,\delta) = \mathbb{E}^f L(\theta, \delta(\mathbf{X}))$ a rozhodovací funkci δ_U, $\forall\theta$, která ji minimalizovala. Dále pak **supremální rizikovou funkci** $\sup_\theta \mathcal{R}(\theta,\delta)$ a rozhodovací funkci δ_0, která ho minimalizovala.
Definovali jsme dále **apriorní Bayesovské riziko** vztahem $r(\pi,\delta) = \mathbb{E}^\pi \mathcal{R}(\theta,\delta)$, které se snažilo minimalizovat $\delta^\pi = \underset{\delta\in\mathscr{D}}{\operatorname{argmin}}\, r(\pi,\delta)$.

Definice 9.1. Definujeme **maximinimální riziko** vztahem

$$\underline{\mathcal{R}} = \sup_{\pi\in\Pi} r(\pi,\delta^\pi) = \sup_{\pi\in\Pi}\inf_{\delta} r(\pi,\delta) = \sup_{\pi\in\Pi} r(\pi),$$

kde Π je rodina apriorních hustot.

Věta 9.2. *$\underline{\mathcal{R}} \leqslant \overline{\mathcal{R}}$, kde $\overline{\mathcal{R}} = \inf_{\delta}\sup_{\theta} \mathcal{R}(\theta,\delta)$ je minimaxní riziko.*

Důkaz. Volme apriorní rozdělení $\pi(\theta)$. Potom

$$r(\pi,\delta) = \int_\Theta \underbrace{\mathcal{R}(\theta,\delta)}_{\geqslant 0 \wedge \leqslant \sup_\theta \mathcal{R}(\theta,\delta)} \underbrace{\pi(\theta)}_{\geqslant 0} \,\mathrm{d}\theta \leqslant \int_\Theta \sup_\theta \mathcal{R}(\theta,\delta)\pi(\theta)\mathrm{d}\theta = \sup_\theta \mathcal{R}(\theta,\delta)\cdot \underbrace{\int_\Theta \pi(\theta)\mathrm{d}\theta}_{1}.$$

Dále platí nerovnice

$$\begin{aligned} r(\pi,\delta) &\leqslant \sup_\theta \mathcal{R}(\theta,\delta), \quad \forall\delta,\forall\pi \\ \inf_\delta r(\pi,\delta) &\leqslant \inf_\delta \sup_\theta \mathcal{R} = \overline{\mathcal{R}}, \quad \forall\pi \\ \sup_\pi \inf_\delta r = \underline{\mathcal{R}} &\leqslant \sup_\pi \overline{\mathcal{R}} = \overline{\mathcal{R}} \end{aligned}$$

□

Poznámka 9.3. Pokud δ^π dosahuje na $\underline{\mathcal{R}}$ a $\underline{\mathcal{R}} = \overline{\mathcal{R}}$ ve větě 9.2, pak δ^π je současně **minimaxní strategie**.

Definice 9.4. Pokud existuje $\pi^*(\theta)$ apriorní rozdělení takové, že $r(\pi^*) = \underline{\mathcal{R}}$, tzn. $\forall\pi \in \Pi$, $r(\pi) \leqslant r(\pi^*)$, pak se nazývá **nejméně příznivé** (*least favourable* = **LF**).

Věta 9.5. *Nechť π je vlastní apriorní rozdělení na Θ takové, že $r(\pi) = \sup_\theta \mathcal{R}(\theta,\delta^\pi)$. Pak*

1. *δ^π je minimaxní strategií,*
2. *pokud δ^π jednoznačné Bayesovské řešení, pak δ^π je jednoznačná minimaxní strategie,*

3. π je nejméně příznivé rozdělení (LF).

Důkaz. 1. Vezmeme δ libovolné fixní. Víme, že

$$\sup_\theta \mathcal{R}(\theta,\delta)\cdot 1 = \sup_\theta \mathcal{R}(\theta,\delta)\int \pi(\theta)\mathrm{d}\theta \ \wedge\ \sup_\theta \mathcal{R}(\theta,\delta) \geqslant \mathcal{R}.$$

Potom

$$\sup_\theta \mathcal{R}(\theta,\delta) \geqslant \int \mathcal{R}\pi(\theta)\mathrm{d}\theta = \mathbb{E}^\pi\mathcal{R} = r(\pi,\delta) \geqslant r(\pi,\delta^\pi) = r(\pi) = \sup_\theta \mathcal{R}(\theta,\delta^\pi).$$

2. Analogicky jako předchozí bod, akorát $r(\pi,\delta) > r(\pi,\delta^\pi)$
3. Volíme π' libovolné fixní. Pak $r(\pi') = r(\pi',\delta^{\pi'}) \leqslant r(\pi',\delta^\pi) \leqslant \sup_\theta \mathcal{R}(\theta,\delta^\pi) = r(\pi)$.

   Potom $\pi = \pi^*$ je LF.

□

Poznámka 9.6. Věta 9.5 je jednou z prvních vět o tom, jak lze za daných podmínek z Bayesovské strategie získat klasickou minimaxní strategii.

Důsledek 9.7. *Věta platí například tehdy, když $\mathcal{R}(\theta,\delta^\pi)$ je konstantní v θ na Θ s.j. vzhledem k apriornímu rozdělení π, protože pak platí, že $r(\pi) = \sup \mathcal{R}$.*

Definice 9.8. Nejméně příznivá posloupnost $(\pi_n)_{n=1}^{+\infty}$ je taková, pro kterou existuje $\lim_{n\to+\infty} r(\pi_n) = \underline{\mathcal{R}}$.

Věta 9.9. *Nechť $(\pi_n)_1^{+\infty}$ je posloupnost taková, že $\lim_{n\to+\infty} r(\pi_n) = r$. Nechť dále δ_0 je libovolná rozhodovací funkce taková, že $\sup_\theta \mathcal{R}(\theta,\delta_0) = r$. Pak v analogii s větou 9.5*

1. *δ_0 je minimaxní,*
2. *NELZE VYSLOVIT a*
3. *$(\pi_n)_1^{+\infty}$ je nejméně příznivá posloupnost apriorních rozdělení (LF).*

Důkaz. Analogicky jako ve větě 9.5 (tedy za domácí cvičení). □

Příklad 9.10. Mějme $X \sim \mathrm{Bi}(n,p)$, kde se snažíme nalézt parametr $\widehat{p}$. $\pi(p) \sim \mathrm{Beta}(a,b) \sim \frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}$, kde $a,b>0$.
Příkladem může být teorie spolehlivosti, kde máme datové pole v RAIDu, tj. n-komponentní systém. Označme X počet poruch za čas 1. Pokud nastane jev $\{X \geqslant 5\}$, tak nastane výpadek. Tuto pravděpodobnost se snažíme najít, tedy $\mathbb{P}(X \geqslant 5) = \widehat{p}$. Střední počet poruch pak můžeme odhadnout jako $\widehat{\mathbb{E}X} = n\widehat{p}$ za čas 1.

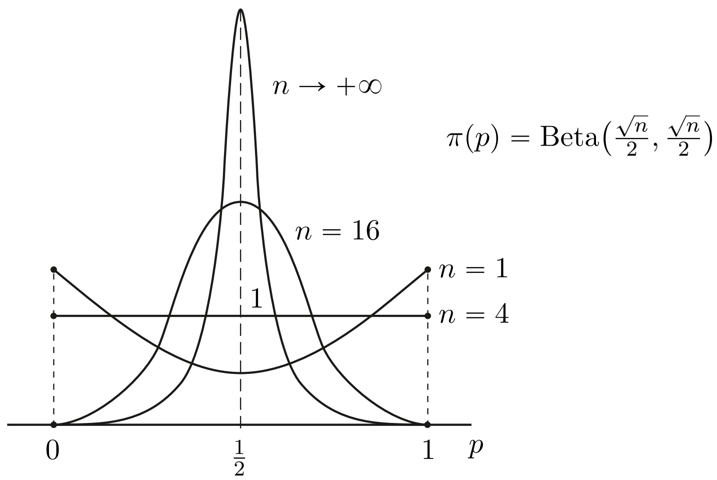
1) Můžeme například udělat apriorní odhad $\widehat{p}_a = \mathbb{E}[\mathrm{Beta}(a,b)] = \frac{a}{a+b}$.
2) Můžeme také naměřit data x a vypočíst klasický ML odhad jako $\widehat{p}_{\mathrm{ML}} = \frac{x}{n} = \overline{x_n}$. Pokud $L = L_2$, tak je to dokonce $\widehat{p}_{\mathrm{UMVUE}}$.
3) Můžeme také udělat Bayesovský odhad jako

$$\pi(p|x) = \frac{\binom{n}{x}p^x(1-p)^{n-x}\cdot\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}}{c} = \frac{1}{c}p^{x+a-1}(1-p)^{n-x+b-1} \sim \mathrm{Beta}(x+a,n-x+b).$$

Z toho potom $\widehat{p}_{\mathrm{B}} = \mathbb{E}[\mathrm{Beta}(x+a,n-x+b)] = \frac{x+a}{n+a+b} = \frac{\overline{x_n}+\frac{a}{n}}{1+\frac{a+b}{n}}$. Zároveň však lze $\widehat{p}_{\mathrm{B}}$ zapsat jako konvexní kombinaci z bodu 2 ve tvaru

$$\widehat{p}_{\mathrm{B}} = \underbrace{\frac{a+b}{a+b+n}}_{\alpha_n}\frac{a}{a+b} + \underbrace{\frac{n}{n+a+b}}_{1-\alpha_n}\frac{x}{n} = \alpha_n\widehat{p}_a + (1-\alpha_n)\widehat{p}_{\mathrm{ML}},$$

kde α_n se nazývá porce neurčitosti apriorní znalosti. Ukažme si nyní, co vlastně dělá to $\widehat{p}_a$:



Obrázek 9.1: Rozdělení $\pi(p) = \mathrm{Beta}(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ pro vybraná n.

a) Pokud je n pevné, a, b velké a $b/a = konst$. Potom $\alpha_n \to 1$, a tedy $\widehat{p}_\mathrm{B} = \widehat{p}_a$.
b) Pokud jsou a, b pevné a n velké, pak $\alpha_n \to 0$ ($\alpha_n \doteq 0$), a tedy $\widehat{p}_\mathrm{B} = \widehat{p}_\mathrm{ML}$.
c) Pokud $a = b = 1$, tedy $\pi(p) = U(0,1)$ podle principu neurčitosti. Potom $\widehat{p}_\mathrm{B} = \frac{x+1}{n+2} \doteq \widehat{p}_\mathrm{ML}$ pro vysoká n.
d) Pokud $a = b = 0$, $\pi(p) = p^{-1}(1-p)^{-1}$ je apriorní nevlastní limitní rozdělení, pak $\widehat{p}_B = \widehat{p}_\mathrm{ML}$.
e) Pokud $a = a_n = \frac{\sqrt{n}}{2}$, $b = b_n = \frac{\sqrt{n}}{2}$, tedy $\pi(p) = \mathrm{Beta}(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ (závisí na n). Potom

$$\widehat{p}_B = \frac{x + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} \quad \Rightarrow \quad \mathcal{R}(p, \widehat{p}_B) \overset{L=L_2}{=\!=} \mathbb{E}\,\underbrace{(p - \widehat{p}_\mathrm{B})^2}_{L_2(p,\widehat{p}_B)} = \mathbb{E}\left(p - \frac{X + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}\right)^2 =$$

$$= c \underbrace{\mathbb{E}X^2}_{np-np^2+n^2p^2} + c' \underbrace{\mathbb{E}X}_{np} + e \overset{\cdots}{=\!=} \frac{n}{4(n+\sqrt{n})^2} = konst.$$

To ilustruje obrázek 9.1. Dle věty 9.5 a důsledku 9.7 víme, že $r(\pi) = \sup_\theta \mathcal{R}(\theta, \delta^\pi)$ implikuje to, že $\widehat{p}_B = \widehat{p}_\mathrm{minimax}$ a $\pi^* = \pi(p)$ je LF.

Příklad 9.11. Mějme $X \sim \mathcal{N}(\mu, 1) = \mathcal{N}(\theta, 1)$ a zkoumáme $\widehat{\mu}$ na základě dat x. Můžeme tedy vzít $\widehat{\mu}_\mathrm{ML} = \overline{x_n}$ (pokud naměříme více dat).

A) Zvolíme nevlastní neurčitou informaci $\pi(\mu) = konst = 1$ na $\mathbb{R}$ ($\widehat{\mu}_a$ není nevlastní).

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

$$\widehat{\mu}_\mathrm{B} = \mathbb{E}\left[\mathcal{N}(x,1)\right] = \overline{x_n},$$

$$\mathcal{R}(\mu, \widehat{\mu}_\mathrm{B}) \overset{L=L_2}{=\!=} \mathbb{E}(\mu - X)^2 = \mathrm{D}X = 1,$$

tzn. $\mathcal{R}(\mu, \widehat{\mu}_\mathrm{B}) = konst$. Nemůžeme ale aplikovat větu 9.5, protože π není vlastní.

B) $\pi_n(\mu) = \mathcal{N}(0, n)$ je vlastní hustota $\forall n$ (CF) a zkusíme použít větu 9.9. (Měli bychom vědět, že $A \propto B$ značí, že existuje konstanta k taková, že $A = kB$.)

$$\pi_n(\mu|x) \propto \mathrm{e}^{-\frac{1}{2}(x-\mu)^2}\mathrm{e}^{-\frac{1}{2n}(\mu-0)^2} \propto \mathrm{e}^{-\frac{1}{2}\left[\left(1+\frac{1}{n}\right)\mu^2 - 2\mu x + x^2\right]} \propto \mathrm{e}^{-\frac{1}{2}\left(1+\frac{1}{n}\right)\left[\mu - \frac{n}{n+1}x\right]^2} \sim \mathcal{N}\left(\frac{n}{n+1}x, \frac{n}{n+1}\right)$$

$$\widehat{\mu}_{\mathrm{B}} = \mathbb{E}\mathcal{N}\Big(\frac{n}{n+1}x, \frac{n}{n+1}\Big) = \frac{n}{n+1}x = \delta^{\pi_n} \sim \widehat{\mu}_{B,n}.$$

$$r(\pi_n) \overset{L=L_2}{=\!=\!=} \mathrm{D}\pi(\mu|x) = \frac{n}{n+1} \to 1 \text{ (vzorec ověříme v následující větě).}$$

$\lim\limits_{n\to+\infty} r(\pi_n) = 1 = r$ a víme, že máme δ_0, pro které je $r = \mathcal{R}(\mu, \overbrace{\underbrace{\widehat{\mu}_{\mathrm{B}}}_{=\delta_0=x}}^{\text{z A)}})$. Potom dle věty 9.9 je $\delta_0 = x$ minimaxní a posloupnost $\mathcal{N}(0, n)$ je LF.

Věta 9.12 (!!!Vyhazovací!!!)**.** *Mějme ÚBM, $X \sim f(x|\theta)$, $\pi(\theta)$, $L(\theta, d) = L_2(\theta, d) = (\theta - d)^2$. Nechť pro aposteriorní střední hodnotu platí, že $0 < \mathbb{E}^\pi(\theta^2|x) < +\infty$ ($\neq 0$ aby nebyla degenerovaná). Pak*

$$\delta^\pi(x) = \mathbb{E}^\pi(\theta|x) = \int_\Theta \theta\pi(\theta|x)\mathrm{d}\theta = \frac{\int \theta f(x|\theta)\pi(\theta)\mathrm{d}\theta}{\int f(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

a Bayesovské riziko

$$r(\pi) = \mathbb{E}^m[\mathrm{D}^\pi(\theta|x)],$$

kde $m(x) = \int f\pi\mathrm{d}\theta$.

Důkaz. Vyjdeme z definice Bayesovského odhadu $\delta^\pi = \operatorname*{argmin}_\delta r(\pi, \delta)$.

$$r(\pi,\delta) = \mathbb{E}^\pi[\mathcal{R}(\theta,\delta)] = \mathbb{E}^\pi\Big[\mathbb{E}^f\big(\theta - \delta(X)\big)^2\Big] = \left|\begin{array}{c}\text{záměna}\\ \mathbb{E}^\pi \text{ a } \mathbb{E}^f\end{array}\right| \overset{\text{viz dříve}}{=\!=\!=} \mathbb{E}^m\Big[\mathbb{E}^\pi\big(\theta - \delta(x)\big)^2|x\Big] =$$

$$= \mathbb{E}^m\Big[\mathbb{E}^\pi(\theta^2|x) - 2\delta(x)\mathbb{E}^\pi(\theta|x) + \delta(x)^2\Big] = \mathbb{E}^m\Big[\underbrace{\mathbb{E}^\pi(\theta^2|x)}_{0<\ldots<+\infty} + \big(\delta(x) - \mathbb{E}^\pi(\theta|x)\big)^2 - \underbrace{\big(\mathbb{E}^\pi(\theta|x)\big)^2}_{0<\ldots<+\infty}\Big]$$

Abychom tento výraz optimalizovali, musíme vzít $\delta^\pi = \delta(x) = \mathbb{E}^\pi(\theta|x)$, abychom výraz minimalizovali.

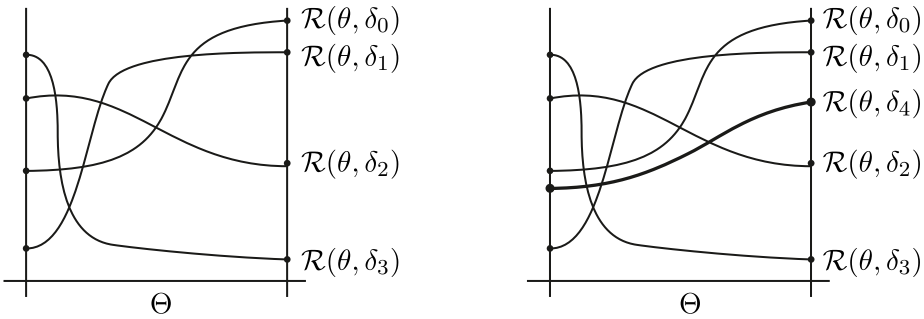
$$r(\pi) = r(\pi, \delta^\pi) = \mathbb{E}^m\Big[\underbrace{\mathbb{E}^\pi\big(\theta - \mathbb{E}^\pi(\theta|x)\big)^2|x}_{\mathrm{D}(\theta|x)}\Big].$$

Záměna je provedena obdobně jako ve větě 5.12. □

10 Přípustnost Bayesovského řešení, Steinův efekt, Bergerův jev

10.1 Přípustnost (Admissibility)

Definice 10.1. Mějme $\mathscr{D}', \Theta, ...$ Rozhodovací funkce δ_0 se nazývá **nepřípustná**, pokud existuje δ_1, která dominuje δ_0, ozn. $\delta_1 \ll \delta_0$, tzn. $\mathcal{R}(\theta, \delta_1) \leqslant \mathcal{R}(\theta, \delta_0)$, $\forall \theta \in \Theta$ (stejnoměrně) a $\exists \theta_0 \in \Theta$ tak, že $\mathcal{R}(\theta_0, \delta_1) < \mathcal{R}(\theta_0, \delta_0)$. V případě, že taková δ_1 neexistuje, nazýváme δ_0 **přípustný**.



Obrázek 10.1: Vlevo jsou přípustné všechny 4 odhady, protože žádný z nich nejde stejnoměrně vylepšit nějakým jiným odhadem. Vpravo už je ale $\mathcal{R}(\theta, \delta_0)$ nepřípustný, protože ho můžeme stejnoměrně vylepšit funkcí $\mathcal{R}(\theta, \delta_4)$.

Věta 10.2.

- $\delta_U = \operatorname{argmin}_\delta \mathcal{R}(\theta, \delta)$, $\forall \theta$, *je vždy přípustný, pokud existuje,*
- $\delta_0 = \operatorname{argmin}_\delta \sup_\theta \mathcal{R}(\theta, \delta)$ *je přípustný, pokud existuje a je jednoznačný,*
- $\delta^\pi = \operatorname{argmin}_\delta r(\pi, \delta)$ *je přípustný, pokud existuje a je jednoznačný.*

Věta 10.3. *Nechť $L(\theta, d)$ je ryze konvexní v d pro $\forall \theta$. Mějme δ_0 jako přípustnou rozhodovací funkci pro $\tau(\theta)$, případně přímo θ. Mějme δ' takovou, že $\mathcal{R}(\theta, \delta') = \mathcal{R}(\theta, \delta_0)$, $\forall \theta$. Pak $\delta' = \delta_0$ s pravděpodobností 1 (s.j.), tedy*

$$\mathbb{E}L\big(\theta, \delta'(X)\big) = \mathbb{E}L\big(\theta, \delta_0(X)\big).$$

Důkaz. Sporem: $\delta_0 \neq \delta'$ s nenulovou pravděpodobností. Sestrojíme $\delta^* = \frac{\delta_0 + \delta'}{2}$ (je to konvexní kombinace). Pak

$$\mathcal{R}(\theta, \delta^*) = \mathbb{E}L\big(\theta, \delta^*(X)\big) = \mathbb{E}\Big[L\Big(\theta, \frac{\delta_0 + \delta'}{2}\Big)\Big] \overset{\forall \theta}{<} \mathbb{E}\Big[\frac{1}{2}L(\theta, \delta_0) + \frac{1}{2}L(\theta, \delta')\Big] =$$
$$= \frac{1}{2}\underbrace{\mathbb{E}L(\theta, \delta_0)}_{\mathcal{R}(\theta, \delta_0)} + \frac{1}{2}\underbrace{\mathbb{E}L(\theta, \delta')}_{\mathcal{R}(\theta, \delta') = \mathcal{R}(\theta, \delta_0)\ \forall \theta} = \mathcal{R}(\theta, \delta_0),\ \forall \theta.$$

To znamená, že $\delta^* \ll \delta_0$, což je spor s přípustností δ_0. □

Věta 10.4. *Jednoznačná δ^π je vždy přípustná rozhodovací funkce, pokud existuje. (Část věty 10.2)*

Důkaz. δ^π jednoznačný Bayes, pak

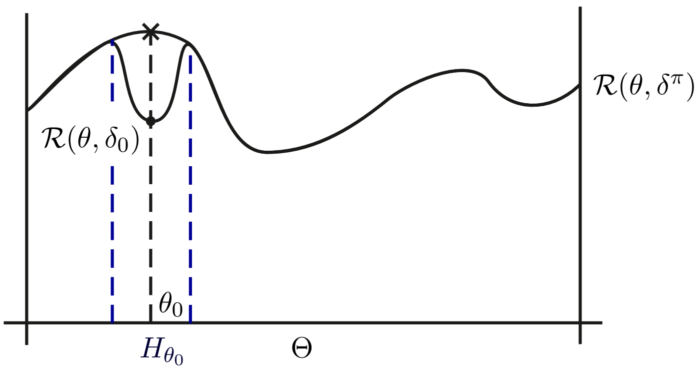
$$
\begin{aligned}
r(\pi, \delta^\pi) &< r(\pi, \delta), \quad \forall \delta,\\
\mathbb{E}\mathcal{R}(\theta, \delta^\pi) &< \mathbb{E}\mathcal{R}(\theta, \delta), \quad \forall \delta,\\
\int \mathcal{R}(\theta, \delta^\pi)\pi(\theta)\mathrm{d}\theta &< \int \mathcal{R}(\theta, \delta)\pi(\theta)\mathrm{d}\theta, \quad \forall \delta.
\end{aligned}
$$

Z toho vyplývá, že existuje bod θ_0 takový, že $\mathcal{R}(\theta_0, \delta^\pi) < \mathcal{R}(\theta_0, \delta)$, $\forall \delta$. Kvůli tomu δ nedominuje δ^π, a proto je δ^π přípustná. □

Poznámka 10.5. Pokud $L(\theta, d)$ je ryze konvexní v d pro $\forall \theta$, pak δ^π je vždy jednoznačná.

Věta 10.6. *Mějme ÚBM, $X \sim f(x|\theta)$, $\pi(\theta)$. Nechť $\pi(\theta) > 0$ pro $\forall \theta$, nechť dále δ^π existuje a $r(\pi) = r(\pi, \delta^\pi) < +\infty$. Nechť $\mathcal{R}(\theta, \delta)$ je spojitá v θ pro $\forall \delta$. Pak δ^π je přípustná. (Větu lze ukázat i pro $r(\pi, \delta^\pi) = +\infty$)*

Důkaz. Sporem. Nechť δ^π je nepřípustná. Pak by ale existovalo $\delta_0 \ll \delta^\pi$, tzn. $\mathcal{R}(\theta, \delta_0)$ stejnoměrně vylepšuje $\mathcal{R}(\theta, \delta^\pi)$ a $\exists \theta_0$ tak, že $\mathcal{R}(\theta_0, \delta_0) < \mathcal{R}(\theta_0, \delta^\pi)$, viz obr. 10.2. Potom by ale



Obrázek 10.2: Schéma důkazu.

existovalo okolí H_{θ_0} tak, že $\mathcal{R}(\theta, \delta_0) < \mathcal{R}(\theta, \delta^\pi)$ na H_{θ_0}. Potom by platilo, že $\mathbb{E}^\pi\mathcal{R}(\theta, \delta_0) < \mathbb{E}^\pi\mathcal{R}(\theta, \delta^\pi)$. To platí, protože $\pi(\theta) > 0$, $\forall \theta$. Potom tedy

$$
r(\pi, \delta_0) < \underbrace{r(\pi, \delta^\pi)}_{r(\pi) < +\infty},
$$

což je spor s tím, že δ^π je Bayesovská strategie. V δ^π totiž dosahuje minima a my jsme našli menší. □

Poznámka 10.7 (Komentář k předpokladu z předchozí věty).

$$
\mathcal{R}(\theta, \delta) = \mathbb{E}L\big(\theta, \delta(X)\big) = \int_{\mathcal{X}(\mathbb{R}^n)} L\big(\theta, \delta(x)\big) f(x|\theta)\mathrm{d}x.
$$

Zde stačí zajistit, aby měl integrand Lebegovskou majorantu, protože chceme L spojitou v θ.

Poznámka 10.8. Doteď jsme hledali podmínky, za kterých je odhad δ^π přípustný. Nyní budeme hledat možnosti, jak tuto implikaci obrátit.

10.2 Historie hledání odhadu pro přípustné funkce

Stein (1955) Našel nutnou a postačující podmínku (nebudeme ukazovat jakou), za které jsou všechny přípustné rozhodovací funkce limitami posloupnosti δ^{π_n} Bayesovských strategií.

Farrell (1968) (pro zajímavost) Předpokládáme, že

1) $f(x|\theta)$ spojitá v θ, $f > 0$ na Θ,
2) $L(\theta, d)$ je ryze konvexní, spojitá,
3) $\forall E \subset \Theta$ kompaktní platí, že

$$\lim_{\|\delta\|\to+\infty} \inf_{\theta\in E} L(\theta,\delta) = +\infty. \qquad \text{(L je nutně neomezená funkce)}$$

Pak libovolná δ_0 přípustná je limitou δ^{π_n} Bayesovských strategií.

Brown (1986) Předpokládejme, že

1) L ryze konvexní,
2) $f(x|\theta) > 0$, $\mathscr{D}'$ uzavřená a konvexní,
3) L je zdola polospojitá (lower semicontinuous), tzn. f je v x_0 zdola polospojitá, pokud $(\forall\varepsilon > 0)(\exists H_{x_0})\big(f(x) \geqslant f(x_0) - \varepsilon\big)$ na H_{x_0}.
4) $\lim\limits_{\|\delta\|\to+\infty} L(\theta,\delta) = +\infty$ (L je nutně neomezená funkce).

Pak libovolná přípustná strategie je bodovou limitou δ^{π_n} Bayesovské posloupnosti, kde $\pi_n(\theta)$ mají omezený support.

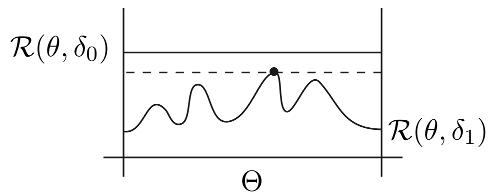
Závěr Libovolná přípustná strategie δ je dosažitelná buď přímo Bayesovským odhadem δ^π nebo zobecněným Bayesovským odhadem δ^π, případně posloupností δ^{π_n}.

Jak funguje přípustnost u klasických statistických strategií?

Věta 10.9. *Pokud existuje jednoznačně minimaxní strategie δ_0, pak je přípustná.*

Důkaz. Ponecháno čtenáři. (Obdobně jako ve větě 10.4) □

Věta 10.10 ("Obrátka"). *Nechť $L(\theta, d)$ je ryze konvexní v d pro $\forall\theta$, δ_0 je přípustná rozhodovací funkce, která má konstantní riziko $\mathcal{R}$. Pak δ_0 je jednoznačná minimaxní strategie.*

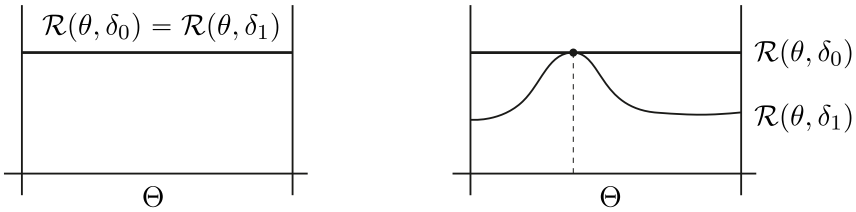


Obrázek 10.3: Schéma důkazu sporem. $\mathcal{R}(\theta,\delta_0)$ je konstantní a není minimaxní.

Důkaz. 1) Ukážeme, že δ_0 je minimaxní. Sporem. Nechť δ_0 není minimax, viz obrázek 10.3. Existuje tedy funkce $\mathcal{R}(\theta,\delta_1)$, která vylepšuje odhad $\mathcal{R}(\theta,\delta_0)$, protože má supremum menší, než hodnota funkce $\mathcal{R}(\theta,\delta_0)$. To je spor s přípustností δ_0.

2) Jednoznačnost sporem. Nechť tedy existuje $\delta_1 \neq \delta_0$ a přitom δ_1 je minimaxní. Možnosti:
   a) $\mathcal{R}(\theta,\delta_0) = \mathcal{R}(\theta,\delta_1), \forall\theta$. Ihned získáváme spor díky větě 10.3 ($\delta_0 = \delta_1$ s.j.), viz obr. 10.4.
   b) $\mathcal{R}(\theta,\delta_0)$ konstantní z předpokladu a $\mathcal{R}(\theta,\delta_1)$ je pod ní, ale mají stejné supremum, což je spor s přípustností δ_0, protože jsme našli δ_1 takovou, že $\delta_1 \ll \delta_0$.

□



Obrázek 10.4: Schéma důkazu sporem. Vlevo je znázorněn případ (a), kde $\mathcal{R}(\theta, \delta_0)$ a $\mathcal{R}(\theta, \delta_1)$ splývají. Vlevo (b) jsou potom odlišné, ale mají stejné supremum.

10.3 Steinův efekt (1955)

Dříve se označoval jako Steinův paradox, ale už se ho podařilo vysvětlit. Mějme tedy běžné statistické $\mathbf{X} \sim f(\mathbf{x}, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$, f nechť je sféricky symetrické rozdělení okolo $\boldsymbol{\theta}$, tzn. $f(\|\mathbf{x}' - \boldsymbol{\theta}\| + \|\mathbf{x}''\|)$, kde $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$, $\mathbf{x}' \in \mathbb{R}^k$, $\mathbf{x}'' \in \mathbb{R}^{n-k}$. To splňuje například $\mathcal{N}_k(\boldsymbol{\theta}, \mathbf{I}_{k\times k})$. Volme pro $\boldsymbol{\theta}$ váženou kvadratickou ztrátovou funkci $L_2^w(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{j=1}^k w_j(\theta_j - \delta_j)^2$.

Označme $\boldsymbol{\delta}^*(\mathbf{x}) = \left(\delta_1^*(\mathbf{x}), ..., \delta_k^*(\mathbf{x})\right)$ odhad (rozhodovací funkci) pro $\boldsymbol{\theta}$ získaný klasickou statistickou procedurou.

Pak $\exists k_0$ tak, že $\forall k \geqslant k_0$ je $\boldsymbol{\delta}^*$ nepřípustným odhadem (rozhodovací funkce) pro $\boldsymbol{\theta}$, přestože δ_j^* jsou pro odhad θ_j přípustné pro každé $j \in \widehat{k}$.

Důsledek 10.11. *Například klasická rozhodovací funkce metodou jednoznačného minimaxu je od jistého k_0 beznadějná, protože víme (z věty 10.9), že pokud existuje jednoznačný minimax, pak je přípustný. Stein ale říká, že od jistého k_0 se klasickou statistickou procedurou získat přípustné řešení nedá, tzn. od jistého k_0 nemůže existovat jednoznačný minimax. A také stejnoměrně nejlepší strategie δ_U od jisté dimenze k_0 neexistuje.*

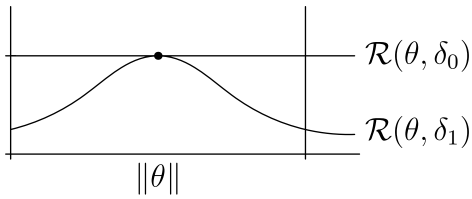
Co s tím? Místo hledání $\delta_U = \underset{\delta \in \mathscr{D}}{\operatorname{argmin}}\, \mathcal{R}(\theta, \delta)$, $\forall \theta$ hledá klasická statistika $\delta_{pu} = \underset{\delta \in \mathscr{D}_0 \subset \mathscr{D}}{\operatorname{argmin}}\, \mathcal{R}(\theta, \delta)\ \forall \theta$, např. $\mathscr{D}_0$ nestranné odhady (UMVUE). Tím jsme zajistili existenci, ale pro vysokou dimenzi opět dostaneme nepřípustné řešení. Další možností je použít $\mathscr{D}_0'$ jako **ekvivariantní** odhady. Ty pak vedou na UMEqE (*Uniformly minimum equivariant estimator*).

Řešení se pak nabízí v podobě Bayese (začlenění apriorní informace π).

10.4 Srovnání přístupů

Tato sekce je brána jako taková poznámka/opakování, které shrnuje některé přístupy z historie.

1955 Steinův efekt: Mějme $X \sim \mathcal{N}_k(\boldsymbol{\theta}, \mathbb{I}_{k\times k})$ a nechť $\delta_0(\mathbf{x}) = \mathbf{x}$ je ML, minimaxní a víme, že $\mathcal{R}(\theta, \delta_0) = konst.$ Do roku 1955 se myslelo, že δ_0 je jednoznačný minimax $\forall k$. Dle věty o jednoznačnosti minimaxu by bylo δ_0 přípustné $\forall k$. Stein však roku 1955 ukázal, že δ_0 je jednoznačný minimax a přípustný pro $k = 1, 2$, a že pro $(\forall k \geqslant 3)(\exists \delta')(\delta' \ll \delta_0)$.



Obrázek 10.5: Kdyby byl $\mathcal{R}(\theta, \delta_0)$ nejednoznačný minimax, mohl by existovat odhad $\mathcal{R}(\theta, \delta_1)$, který ho stejnoměrně vylepšuje, například by byl stejný jen v jednom bodě.

1961 James-Steinův (shrinkage) odhad parametru polohy:

$$\delta_{\mathrm{JS}}(\mathbf{x}) = \underbrace{\left(1 - \frac{k-2}{\|\mathbf{x}\|^2}\right)}_{\to -\infty \text{ při } \|\mathbf{x}\| \to 0^+} \mathbf{x}$$

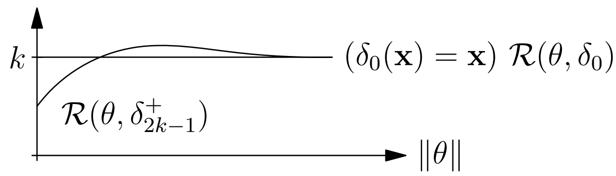
1970 Baranchick:

$$\delta_c^+(\mathbf{x}) = \left(1 - \frac{c}{\|\mathbf{x}\|^2}\right)^+ \mathbf{x} = \begin{cases} \left(1 - \frac{c}{\|\mathbf{x}\|^2}\right)\mathbf{x}, & \text{pokud } c < \|\mathbf{x}\|^2, \\ 0, & \text{jinak,} \end{cases}$$

kde $c \in [k-2, 2(k-2)]$. Pak $\delta_c^+ \ll \delta_{\mathrm{JS}} \ll \delta_0,\ \forall k \geqslant 3$.

1994 Shao et al. $\exists \delta_S \ll \delta_c^+$, ale vylepšení už nestojí za tu složitost (tedy Baranchick je kompromis v rámci použitelnosti).

Poznámka 10.12. Pro $c = 2k - 1 \notin [k-2, 2(k-2)]$ je Baranchickova, ale nedominuje δ_{JS}, $\mathcal{R}(\theta, \delta^+_{2k-1})$.



Obrázek 10.6: Sice má odhad $\mathcal{R}(\theta, \delta^+_{2k-1})$ maximum vyšší, než $\mathcal{R}(\theta, \delta_0)$, ale někdy se vyplatí ho použít, například proto, že chceme předpovídat hodnoty kolem 0 (např. u měření chyb).

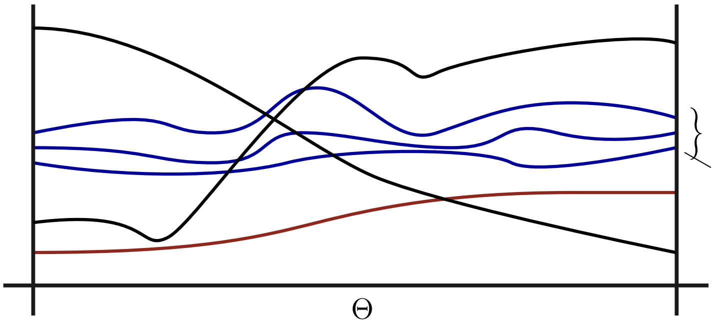
Věta 10.13 (Bergerův jev, 1980)**.** *$\forall k_0 \in \mathbb{N}$ (dimenzi $\boldsymbol{\theta}$), existuje ztrátová funkce $L(\boldsymbol{\theta}, \boldsymbol{d})$ taková, že libovolný odhad $\boldsymbol{\delta}^*(\boldsymbol{x})$ klasickou statistickou procedurou je $\forall k \geqslant k_0$ nepřípustný.*

Poznámka 10.14. Připomínka Rao-Blackwellovy věty: Mějme $L(\theta, d)$ konvexní v d pro $\forall \theta$, $\mathbf{T}(\mathbf{X})$ je postačující statistika, δ_0 je libovolná rozhodovací funkce. Pak

$$\widehat{\delta}(\mathbf{t}) = \mathbb{E}\left[\delta_0(\mathbf{X}) | \mathbf{T}(\mathbf{X}) = \mathbf{t}\right] \ll \delta_0 \quad \text{a} \quad \mathbb{E}\widehat{\delta} = \mathbb{E}\delta_0.$$

Definice 10.15. Označme ztrátovou funkci $L\big(\theta, \delta | \tau(\theta)\big)$. Pak rozhodovací funkci $\delta(\mathbf{x})$ nazýváme **nestrannou** vzhledem k L, pokud

$$\mathbb{E}L\big(\theta, \delta | \tau(\theta)\big) \leqslant \mathbb{E}L\big(\theta, \delta | \varphi(\theta)\big)\big),\ \forall \varphi, \forall \theta.$$



Obrázek 10.7: Modře jsou vyznačeny nestranné odhady, existují také odhady vyznačeny černě, které nejsou nestranné a mezi nimi i červeně vyznačený odhad, který sice není nestranný, ale stejnoměrně vylepšuje všechny nestranné odhady.

Omezili jsme se na prostor nestranných rizikových funkcí $\mathscr{D}_0 \subset \mathscr{D}$ vzhledem k L. Dále $\delta_U^0 = \operatorname*{argmin}_{\delta \in \mathscr{D}_0} \mathcal{R}(\theta, \delta), \ \forall \theta$ jako UBU$_L$E (*Uniformly best unbiased estimators*), pokud existuje. Pak

$$L = L_2(\theta, \boldsymbol{\delta}) = (\theta - \delta)^2 \overset{\text{případně}}{=\!=\!=} \big(\tau(\theta) - \delta\big)^2 \quad \Rightarrow \quad \mathbb{E}L_2(\theta, \delta) = \mathbb{E}(\delta - \mathbb{E}\delta) = \mathrm{D}\delta.$$

Věta 10.16. *Mějme L konvexní v δ, $T(\boldsymbol{X})$ je úplná postačující statistika, δ_0 nestranná rozhodovací funkce vzhledem k L_2. Pak $\widehat{\delta}_{\mathrm{R-B}}(\boldsymbol{t})$ je UMVUE (potenciálně nepřípustný kvůli Steinovi).*

11 Grupy transformací, ekvivariantní odhady

11.1 Equivariantní rozhodovací funkce (MRE$_q$E)

Definice 11.1. Mějme grupu $(G, *)$, tj. máme k dispozici inverzní prvek, neutrální prvek, atd. Pro každé $g_1, g_2 \in G$ tedy máme definováno, co to je $g_1 * g_2$. Volme speciálně grupu $(\mathbb{R}, *)$, potom máme definováno $g * x$, $\forall g \in G$, $\forall x \in \mathbb{R}$ a také $g * \mathbf{x}$, $\forall g \in G$, $\forall \mathbf{x} \in \mathbb{R}^n$ (po složkách). Vezmeme pravděpodobnostní model $X_j = \theta * \varepsilon_j$, kde ε_j značí chyby měření s $\varepsilon_j \sim \mathbb{P}_{0j}$, $\forall j \in \hat{n}$. Potom $\mathbf{X} = (X_j)_{j=1}^n = \theta * \boldsymbol{\varepsilon}$, kde

$$\boldsymbol{\varepsilon} \overset{id}{\sim} \prod_{j=1}^n \mathbb{P}_{0j} \overset{iid}{=\!=} (\mathbb{P}_{01})^n.$$

Dále platí, že

$$\mathbf{X} \sim \mathbb{P}_\theta \in \mathcal{P}_* = \{\mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}\},\ g \in G \ \Rightarrow\ g * \mathbf{X} = g * (\theta * \boldsymbol{\varepsilon}) = (g * \theta) * \boldsymbol{\varepsilon} \sim \mathbb{P}_{g*\theta} \in \mathcal{P}_*.$$

Definice 11.2. Rozhodovací funkce δ se nazývá **ekvivariantní** pro θ vzhledem ke grupě $(G, *)$, pokud platí, že

$$\delta(g * \mathbf{x}) = g * \delta(\mathbf{x}), \quad \forall g \in G,\ \forall \mathbf{x} \in \chi(\mathbb{R}^n).$$

Definice 11.3. Ztrátová funkce $L(\theta, d)$ se nazývá **invariantní** vzhledem ke grupě $(G, *)$, pokud

$$L\big(g * \theta, g * \delta(x)\big) = L\big(\theta, \delta(x)\big) \quad \forall\theta, \forall g, \forall\delta.$$

Důsledek 11.4.

$$L(\theta, \delta) \overset{\forall g}{=\!=} L(g * \theta, g * \delta) = \left| \begin{array}{l} \textit{volme } g = \widetilde{\theta} \textit{ inverzní prvek v } (G, *) \\ \textit{tzn. } \widetilde{\theta} * \theta = I/O \textit{ neutrální prvek v } (G, *) \end{array} \right| = L(\widetilde{\theta} * \theta, \widetilde{\theta} * \delta) =$$
$$= L(I/O, \widetilde{\theta} * \delta) =: \varrho(\widetilde{\theta} * \delta), \quad \forall\theta, \forall\delta.$$

Značení I/O se používá proto, že např. pro operaci $+$ je to 0, kdežto pro $\cdot$ je to 1.

Důsledek 11.5.

$$\mathcal{R}(\theta, \delta) = \mathbb{E}_\theta L\big(\theta, \delta(\boldsymbol{X})\big) \overset{inv.\ L}{=\!=\!=} \mathbb{E}_\theta \varrho\big(\widetilde{\theta} * \delta(\boldsymbol{X})\big) = \left| \begin{array}{c} \boldsymbol{X} = \theta * \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \sim \mathbb{P}_0 \end{array} \right| = \mathbb{E}_{\mathbb{P}_0} \varrho\big(\widetilde{\theta} * \delta(\theta * \boldsymbol{\varepsilon})\big) \overset{\delta\ ekvi.}{=\!=\!=}$$
$$= \mathbb{E}_{\mathbb{P}_0} \varrho\big(\widetilde{\theta} * \theta * \delta(\boldsymbol{\varepsilon})\big) = \mathbb{E}_{\mathbb{P}_0} \varrho\big(I/O * \delta(\boldsymbol{\varepsilon})\big) = \textit{konst v } \theta,\ \forall\theta,$$

tedy nezávisí na θ. Hledá se proto jednodušeji odhad s nejnižším rizikem (mezi ekvivariantními). Vezme se takový odhad, který má nejnižší hodnotu rizikové funkce.

Definice 11.6. Definujeme MRE$_q$E *Minimum risk equivariant estimator* jako

$$\delta^* := \operatorname*{argmin}_{\delta \in \mathscr{D}_{Eq}} \mathcal{R}(\theta, \delta),\ \forall\theta.$$

Takový odhad ale nemusí být přípustný a nemusí ani existovat.

POZNÁMKA 11.7. Bayesovské řešení $\widehat{\delta_{\mathrm{B}}} = \operatorname{argmin}_\delta r(\pi, \delta)$ nebo $\widehat{\delta_{\mathrm{B}}}(\mathbf{x}) = \operatorname{argmin}_{\delta(\mathbf{x})} \varrho(\pi, \delta|\mathbf{x})$, kde $\varrho(\pi, \delta|\mathbf{x})$, se nazývá **aposteriorní Bayesovské riziko** a splňuje (toto ϱ není to stejné jako v textu výše)

$$\varrho(\pi, \delta|\mathbf{x}) = \mathbb{E}^\pi[L|\mathbf{x}] = \int L(\theta, \delta)\pi(\theta|\mathbf{x})\mathrm{d}\theta = \int L(\theta, \delta)\frac{f\pi(\theta)}{\int f\pi(\theta)\mathrm{d}\theta}\mathrm{d}\theta.$$

Zde stačí volit $\pi(\theta) = 1$ na Θ (potenciálně nevlastní). Pokud zvolíme konkrétně $L = L_2$, pak $\delta^*(\mathbf{x}) = \frac{\int \theta f(\mathbf{x}-\theta)\mathrm{d}\theta}{\int f(\mathbf{x}-\theta)\mathrm{d}\theta}$, což se nazývá **Pitmanův odhad**.

PŘÍKLAD 11.8. Volme $(G, *) = (\mathbb{R}, +)$, tedy $g*\mathbf{x} = (x_1+g, ..., x_n+g) = g+\mathbf{x}$. Pravděpodobnostní model se nazývá model s měřítkem (*Location family*) (např. $\mathcal{N}(\mu, \sigma^2$ známé)), pokud

$$\begin{aligned} X_j &= \theta + \varepsilon_j \quad \text{(aditivní chyba)}, \\ \mathbf{X} &= \theta + \boldsymbol{\varepsilon} \sim \mathbb{P}_\theta \text{ a platí, že} \\ g * \mathbf{X} &= g + \mathbf{X} \sim \mathbb{P}_{g+\theta}. \end{aligned}$$

δ je ekvivariantní, pokud $\delta(g + \mathbf{x}) = g + \delta(\mathbf{x})$. Například $\delta(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} X_j$ (nebo také $\delta(\mathbf{x}) = x_{\text{medián}}$) je ekvivariantní rozhodovací funkce, protože

$$\delta(g + \mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n}(X_j + g) = \frac{1}{n}\sum_{j=1}^{n} X_j + g = \delta(\mathbf{x}) + g.$$

L je invariantní, pokud

$$L(\theta, \delta) = \varrho(\widetilde{\theta} * \delta) = \varrho(\delta - \theta) \overset{\text{např.}}{=\!=} (\delta - \theta)^2 = L_2.$$

Máme tedy zajištěno, že $\mathcal{R}(\theta, \delta_{Eq}) = konst$ v θ.
Závěr: Existuje procedura pro hledání MRE_qE pro $(G, *) = (\mathbb{R}, +)$ ve tvaru

$$\delta^*(\mathbf{x}) = \operatorname*{argmin}_{\delta \in \mathscr{D}_{Eq}} \frac{\int \varrho(\delta - \theta)f(\mathbf{x} - \theta)\mathrm{d}\theta}{\int f(\mathbf{x} - \theta)\mathrm{d}\theta} = \delta_{\mathrm{B}}^{\pi=1}(\mathbf{x}).$$

PŘÍKLAD 11.9. Mějme $\varepsilon_j \sim iid\,\mathrm{Exp}(1)$, $f(\mathbf{x}-\theta) = \mathrm{e}^{-\sum_{j=1}^n x_j + n\theta}$, pokud $\theta < x_{(1)} = \min\{(x_j)_1^n\}$. Dále pak

$$\delta^*(\mathbf{x}) \overset{L_2}{=\!=} ... = x_{(1)} - \frac{1}{n},$$

což je Pitmanův odhad pro Exp(1), tedy MRE_qE.

12 Bayesovské odhady pro kvadratickou a multilinearni L

Poznámka 12.1 (Opakování vyhazovací věty 9.12). Mějme ÚBM, $X \sim f_X(x|\theta)$, $\pi(\theta) \sim \theta$, $L = L_2$. Pokud $0 < \mathbb{E}(\theta^2|\mathbf{x}) < +\infty$, pak

$$\delta^\pi(\mathbf{x}) = \mathbb{E}^\pi(\theta|\mathbf{x})$$

a Bayesovské riziko je

$$r(\pi) = \mathbb{E}^m[\mathrm{D}^\pi(\theta|\mathbf{x})].$$

Věta 12.2. *Mějme f_X, $\pi(\theta)$, $L_2^w(\theta,\delta) := w(\theta)(\theta-\delta)^2$. Pokud $0 < \mathbb{E}\left[\theta^2 w(\theta)|\boldsymbol{x}\right] < +\infty$, pak*

$$\delta^\pi(\boldsymbol{x}) = \frac{\mathbb{E}^\pi[\theta w(\theta)|\boldsymbol{x}]}{\mathbb{E}^\pi[w(\theta)|\boldsymbol{x}]}$$

a $r(\pi) = c \cdot \mathbb{E}^{m_w}[\mathrm{D}^{\pi_w}(\theta|\boldsymbol{x})]$, kde $m_w^{(*)}$ a $\pi_w(\theta|\boldsymbol{x})$ jsou marginální rozdělení $\boldsymbol{X}$ a $\theta|\boldsymbol{x}$ příslušné $\pi_w(\theta) = \frac{\pi(\theta)w(\theta)}{c}$, $c = \int \pi w \mathrm{d}\theta$.

Důkaz.

$$r_w(\pi,\delta) \overset{L_2^w}{\equiv} \mathbb{E}^\pi\left[\mathbb{E}^f[L_2^w(\theta,\delta(\mathbf{X}))]\right] = \int_\Theta \left(\int_{\mathcal{X}(\mathbb{R}^n)} w(\theta)\big(\theta-\delta(\mathbf{x})\big)^2 f(\mathbf{x}|\theta)\mathrm{d}\mathbf{x}\right)\pi(\theta)\mathrm{d}\theta =$$

$$= \int_\Theta \mathbb{E}^f\left[L_2\big(\theta,\delta(\mathbf{X})\big)\right] \underbrace{\pi(\theta)w(\theta)}_{\propto \pi_w(\theta)\cdot c} \mathrm{d}\theta = c \cdot \mathbb{E}^{\pi_w} \underbrace{\mathbb{E}^f(L_2)}_{\mathcal{R}(\theta,\delta)} = c \cdot r_2(\pi_w,\delta).$$

Ukázali jsme tedy, že

$$\delta^\pi = \operatorname*{argmin}_\delta r_w(\pi,\delta) = \operatorname*{argmin}_\delta r_2(\pi_w,\delta) \overset{12.1}{=} \mathbb{E}^{\pi_w}(\theta|\mathbf{x}) = \cdots = \frac{\mathbb{E}^\pi[\theta w(\theta)|\mathbf{x}]}{\mathbb{E}^\pi[w(\theta)|\mathbf{x}]}$$

kde aposteriorní rozdělení $\pi_w(\theta|\mathbf{x})$ příslušné apriorní hustotě $\pi_w(\theta)$. Dále $r(\pi) \overset{12.1}{=} \mathbb{E}^{m_w}\mathrm{D}^{\pi_w}(\theta|\mathbf{x})$. ◻

Poznámka 12.3.
- $w(\theta)$ se zde přidává, protože například můžeme chtít, aby byl odhad v okolí nuly opravdu dobrý, ale vzdálené body nás nezajímají.
- Pokud je δ^π jednoznačný, potom je přípustný. Pokud $\mathcal{R}(\theta,\delta)$ je spojitá v θ a $\pi > 0$, potom je δ^π přípustná. Přípustná δ^π není ovlivněna $w(\theta)$ v L_2^w (robustnost δ^π z hlediska volby váhy).

Věta 12.4 (Vyhazovací 2)**.** *Mějme $\left\{f(x|\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k\right\}$, $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$, $L = L_2^{\boldsymbol{Q}}(\boldsymbol{\theta},\boldsymbol{\delta}) = (\boldsymbol{\theta}-\boldsymbol{\delta})^T\boldsymbol{Q}(\boldsymbol{\theta}-\boldsymbol{\delta})$, kde $\boldsymbol{Q}$ je PD matice o rozměru $k \times k$ a $\boldsymbol{Q}$ nezávisí na $\boldsymbol{\theta}$. Nechť $\mathbb{C}\mathrm{ov}(\boldsymbol{\theta}|\boldsymbol{x})$ je konečná (ideálně regulární) a nenulová. Pak pro Bayesovské řešení platí, že $\boldsymbol{\delta}^\pi(\boldsymbol{x}) = \mathbb{E}^\pi(\boldsymbol{\theta}|\boldsymbol{x})$.*

Důkaz. Obdodně jako v $\mathbb{R}^1$ (viz 9.12). ◻

Poznámka 12.5. a) δ^π nezávisí na tvaru $\mathbf{Q}$, tzn. volíme $\mathbf{Q} = \mathbf{I}$, tedy $L_2^{\mathbf{Q}} = \sum_{i=1}^k(\theta_i-\delta_i)^2$, a nemusíme proto užívat $L_2^{\mathbf{Q}} = \sum_i\sum_j q_{ij}(\theta_i-\delta_i)(\theta_j-\delta_j)$.

b) $L_2^{\mathbf{Q},\mathbf{w}}(\boldsymbol{\theta},\boldsymbol{\delta}) = \sum_1^k w_i(\boldsymbol{\theta})\cdot(\theta_i-\delta_i)^2$. V tomto případě tedy věta 12.4 neplatí!

12.1 Další volby L

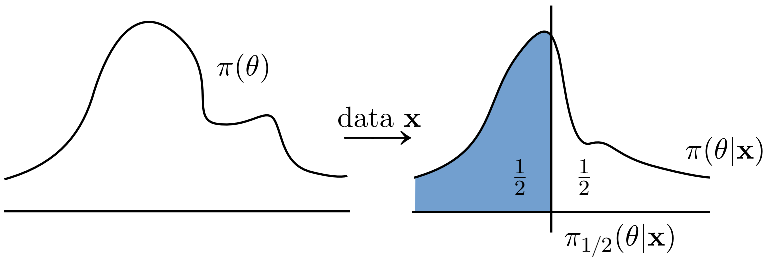
L_2 není robustní na odchylky (1964 Huber, Hampel, Ronchetti), což někdy může být problém. Historicky se proto začaly zkoušet jiné ztrátové funkce, například obecněji $L_p = (\theta - \delta)^p$, $p \in (1, 2)$. Najednou ale nevíme, v jakém tvaru bude δ^π. Víme, že podle definice je

$$\delta^\pi = \operatorname*{argmin}_{\delta} \underbrace{r(\pi, \delta)}_{\mathbb{E}^\pi \mathbb{E}^f L} \overset{\forall \mathbf{x}}{=} \operatorname*{argmin}_{\delta(\mathbf{x})} \underbrace{\int L\big(\theta, \delta(\mathbf{x})\big)\pi(\theta|\mathbf{x})\mathrm{d}\theta}_{\mathbb{E}^\pi(L|\mathbf{x})} = \operatorname*{argmin}_{\delta(\mathbf{x})} \varrho(\pi, \delta|\mathbf{x}),$$

kde ϱ je aposteriorní Bayesovské riziko. Takový způsob ale vede na minimalizaci integrálu, což většinou není možné vyřešit.

Věta 12.6. *Mějme $f(x|\theta)$, $\pi(\theta)$, $L_1(\theta, \delta) := |\theta - \delta|$. Nechť $0 < \mathbb{E}\big[|\theta|\big|\boldsymbol{x}\big] < +\infty$, pak $\delta^\pi(\boldsymbol{x}) = \pi_{1/2}(\theta|\boldsymbol{x})$, což je medián aposteriorního rozdělení.*

Důkaz. Viz obecnější důkaz věty 12.7 □



Obrázek 12.1: Vyznačení mediánu aposteriorní hustoty.

Věta 12.7. *Mějme f, π, $L = L_{k_1,k_2}(\theta, d) = \begin{cases} k_1(d - \theta), & d \geqslant \theta, \\ k_2(\theta - d), & d < \theta. \end{cases}$ (multilineární loss). Nechť $0 < \mathbb{E}\big(|\theta|\big|\boldsymbol{x}\big) < +\infty$. Pak $\delta^\pi(\boldsymbol{x}) = \pi_{\frac{k_2}{k_1+k_2}}(\theta|\boldsymbol{x})$, což je $\frac{k_2}{k_1+k_2}$-kvantil aposteriorního rozdělení.*

Důkaz. Spočítáme si, jak vypadá aposteriorní Bayesovské riziko:

$$\begin{aligned}
\varrho(\pi,\delta|\mathbf{x}) &= \mathbb{E}^\pi\big[L_{k_1,k_2}\big(\theta, \delta(\mathbf{x})\big)\big|\mathbf{x}\big] = k_1 \int_{-\infty}^{\delta} (\delta - \theta)\pi(\theta|\mathbf{x})\mathrm{d}\theta + k_2 \int_{\delta}^{+\infty} (\theta - \delta)\pi(\theta|\mathbf{x})\mathrm{d}\theta \overset{P.P.}{=} \\
&= k_1\Big(\big[(\delta - \theta)\mathrm{F}_\pi(\theta|\mathbf{x})\big]_{-\infty}^{\delta} + \int_{-\infty}^{\delta} \mathrm{F}_\pi(\theta|\mathbf{x})\mathrm{d}\theta\Big) + k_2 \cdot (\dots) = \\
&= k_1\Big(0 - \underbrace{\delta \cdot \mathrm{F}_\pi(-\infty|\mathbf{x})}_{0} + \underbrace{\lim_{\theta \to -\infty} \theta\mathrm{F}_\pi(\theta|\mathbf{x})}_{0,\text{ protože } \int_\Theta |\theta|\pi(\theta|\mathbf{x}) < \infty} + \int_{-\infty}^{\delta} \mathrm{F}_\pi(\theta|\mathbf{x})\mathrm{d}\theta\Big) + k_2 \cdot (\dots^{*}) = \\
&= k_1 \int_{-\infty}^{\delta} \mathrm{F}_\pi(\theta|\mathbf{x})\mathrm{d}\theta + k_2 \int_{\delta}^{+\infty} \big(1 - \mathrm{F}_\pi(\theta|\mathbf{x})\big)\mathrm{d}\theta.
\end{aligned}$$

Využíváme vztahu $\lim_{\theta \to +\infty} \theta\mathrm{F}_\pi(\theta|\mathbf{x}) = 0 = \lim_{\theta \to +\infty} \theta\big(1 - \mathrm{F}_\pi(\theta|\mathbf{x})\big)$, který platí díky předpokladu. Proto pak

$$\varrho'_\delta = k_1 \cdot 1 \cdot \mathrm{F}_\pi(\delta|\mathbf{x}) + k_2 \cdot (-1)\big(1 - \mathrm{F}_\pi(\delta|\mathbf{x})\big) = 0,$$

z toho

$$\mathrm{F}_\pi(\delta|\mathbf{x}) = \frac{k_2}{k_1 + k_2} \in (0,1), \quad \text{a proto } \delta(\mathbf{x}) = \pi_{\frac{k_2}{k_1+k_2}}(\theta|\mathbf{x}).$$

□

Věta 12.8. *Mějme* f, π, $L_{\{0,1\}}(\theta,\delta) \overset{ozn}{=\!=} L_{0,1}(\theta,\delta) = \begin{cases} 0, & |\theta-\delta| \leqslant c, \\ 1, & |\theta-\delta| > c. \end{cases}$ *Pak Bayesovské řešení* $\delta^\pi(\boldsymbol{x})$ *je středem intervalu* I_{2c} *délky* $2c$*, který maximalizuje* $\mathbb{P}^\pi(\theta \in I_{2c}|\boldsymbol{x})$.

Důkaz. Přenecháno čtenáři (Zkouškový důkaz).
Nápověda: Vyjít z $\varrho(\pi, \delta(\mathbf{x})|\mathbf{x})$ a vyrobit tam argmin ϱ. □

Definice 12.9. Definujeme **maximum aposteriori odhad rozhodovací funkce** (někdy také Bayesovská maximálně věrohodný odhad) vztahem $\delta_{\mathrm{MAP}}(\mathbf{x}) := \operatorname{argmax}_\theta \pi(\theta|\mathbf{x})$ nebo alternativně, pokud $\int f \cdot \pi \, \mathrm{d}\theta = +\infty$, definujeme vztahem $\delta_{\mathrm{MAP}}(\mathbf{x}) := \operatorname{argmax}_\theta f(\mathbf{x}|\theta)\pi(\theta)$.

13 Výpočetní aspekty Bayesovských metod

Obecně, když hledáme Bayesovské řešení pro naše naměřená data $\mathbf{x}$, provádíme

$$\underset{\delta \in \mathscr{D}}{\operatorname{argmin}}\, \varrho(\pi, \delta|\mathbf{x}) \propto \int_{\Theta} L(\theta, \big(\delta(\mathbf{x})\big)) \pi(\theta) f(\mathbf{x}|\theta) \mathrm{d}\theta,$$

což je možné počítat numerickými metodami. Pro L_2 jsme však ukázali, že umíme δ najít ve tvaru

$$\delta^{\pi}(\mathbf{x}) = \mathbb{E}^{\pi}[\tau(\theta)|\mathbf{x}] = \frac{\int_{\Theta} \tau(\theta)\pi(\theta) f(\mathbf{x}|\theta)\mathrm{d}\theta}{\int_{\Theta} \pi(\theta) f(\mathbf{x}|\theta)\mathrm{d}\theta},$$

na což je opět možné použít numerické metody a získat tedy alespoň aproximaci řešení. Pro L_1 ztrátové funkce pak můžeme získat (opět např. numericky) odhad

$$\delta^{\pi}(\mathbf{x}) = \pi_{1/2}(\theta|\mathbf{x}).$$

Hledáme tedy $\pi_{1/2}$ takové, že

$$\frac{1}{2} = \int_{-\infty}^{\pi_{1/2}} \pi(\theta|\mathbf{x})\mathrm{d}\theta = \int_{-\infty}^{\pi_{1/2}} \pi(\theta) f(\mathbf{x}|\theta)\mathrm{d}\theta.$$

Pro $L_{0,1}$ pak hledáme

$$\delta^{\pi}(\mathbf{x}) = \underset{\theta}{\operatorname{argmax}}\, \pi(\theta) f(\mathbf{x}|\theta),$$

což opět není snadná úloha, i když dnes existují metody, jak to počítat.

13.1 Numerická integrace

Existuje mnoho metod, pomocí kterých se dají již zmíněné integrály vypočítat. Za zmínku stojí např. Simpson, Aproximace na základě vhodných OG bází, Aproximace wavelety, metoda Monte Carlo, Laplaceova asymptotická aproximace apod. Zajímavé jsou tzv. Polynomiální kvadratury (Gauss, Newton-Costes,...), které postupují následovně:

$$\int_{\Theta} \tau(\theta)\pi(\theta) f(x|\theta)\mathrm{d}\theta \sim \int_{-\infty}^{+\infty} \mathrm{e}^{-t^2/2} f(t)\mathrm{d}t \approx \sum_{i=1}^{n} w_i f(t_i),$$

kde

$$w_i = \frac{2^{n-1} n! \sqrt{n}}{n^2 [H_{n-1}(t_i)]^2}.$$

Zde označujeme t_i jako uzly v $[a, b]$ a w_i jako váhy. Potom t_i je i-tý kořen $H_n(t)$ jakožto Hermitovského polynomu.

14 Bayesovská strategie testování hypotéz

Testujeme hypotézy $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, kde $\Theta = \Theta_0 \uplus \Theta_1$. Znáhodníme $\theta \sim \pi(\theta)$, kde $\pi(\theta)$ jako vždy představuje apriorní informaci (vlastní/nevlastní/konstantu). Oproti klasické statistice je testování u Bayesovské statistiky o poznání jednodušší. Vždy jsme totiž museli najít kritický obor W a zkoumali jsme, zda-li se v ní nachází naše naměřená data. Nezkoumali jsme tedy vlastně pravděpodobnost, že zkoumaná hypotéza platí, ale spíše pravděpodobnost, že data padly do kritického oboru.
V Bayesovské statistice se však rovnou ptáme na pravděpodobnost, že daná hypotéza platí. Máme tedy

Apriorní test H_0**:** H_0 přijmeme, pokud $\mathbb{P}^\pi(H_0) = \mathbb{P}^\pi(\theta \in \Theta_0) \overset{?}{>} \frac{1}{2}$ (můžeme nastavit i jiný threshold, než 1/2).

Aposteriorní test H_0**:** Máme $\pi(\theta)$, naměříme data $\mathbf{X} \sim f_\mathbf{X}$ a z toho získáme $\pi(\theta|\mathbf{x})$. Poté můžeme spočítat pravděpodobnost, že platí H_0 jako

$$\mathbb{P}^\pi(H_0|\mathbf{x}) = \mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}) \overset{?}{>} \frac{1}{2}, \text{ což je vlastně}$$
$$\int_{\Theta_0} \pi(\theta|\mathbf{x}) \mathrm{d}\theta \lessgtr \frac{1}{2},$$

kde $\pi(\theta|\mathbf{x}) = \frac{f\pi(\theta)}{\int_\Theta f\pi(\theta)\mathrm{d}\theta}$.

Vezmeme tedy rozhodovací prostor $\mathscr{D} = \{0,\ 1\} = \{\text{no } H_0,\ \text{yes } H_0\}$ a nastavujeme parametrickou funkci

$$\tau(\theta) = I_{\Theta_0} = \begin{cases} 1, & \theta \in \Theta_0 \ (H_0 \text{ platí}), \\ 0, & \theta \in \Theta_1 \ (H_1 \text{ platí}). \end{cases}$$

S ní přišel Neyman s Pearsonem (a Fisherem) už kolem roku 1950. Nyní ale ukážeme, že tato na první pohled slibná metoda je vlastně špatná a vede na tzv. Jeffreys-Lindleyův paradox.

Věta 14.1. *Volme ztrátovou funkci* $(a_0, a_1 > 0)$

$$L_{a_0,a_1}(\theta,\delta) = \begin{cases} 0, & \delta = I_{\Theta_0}(\theta) = \tau(\theta) \ (\textit{trefa}), \\ a_0, & \theta \in \Theta_0(H_0) \wedge \delta = 0 \ (\textit{chyba I. druhu}), \\ a_1, & \theta \in \Theta_1(H_1) \wedge \delta = 1 \ (\textit{chyba II. druhu}). \end{cases}$$

Potom

$$\delta^\pi = \begin{cases} 1, & \textit{pokud } \mathbb{P}^\pi(\theta \in \Theta_0|\boldsymbol{x}) > \frac{a_1}{a_0+a_1}, \\ 0, & \textit{pokud } \mathbb{P}^\pi(\theta \in \Theta_0|\boldsymbol{x}) \leqslant \frac{a_1}{a_0+a_1}, \end{cases}$$

kde $\frac{a_1}{a_0+a_1} = \frac{1}{1+\frac{a_0}{a_1}}$ *je závislá na poměru* $\frac{a_0}{a_1}$. *Pro* $a_0 = a_1 = 1$ *je pak hranice* $\frac{1}{2}$.

Důkaz. Vezmeme aposteriorní Bayesovské riziko (za L bereme $L_{a_0,a_1}(\theta,\delta)$)

$$\varrho(\pi,\delta|\mathbf{x}) = \mathbb{E}^\pi\big[L\big(\theta,\delta(\mathbf{x})\big)\big|\mathbf{x}\big] = 0\cdot\mathbb{P}^\pi(...) + a_0\cdot\mathbb{P}^\pi(\theta\in\Theta_0|\mathbf{x})\cdot I_{\{0\}}(\delta) + a_1\cdot\mathbb{P}^\pi(\theta\in\Theta_1|\mathbf{x})\cdot I_{\{1\}}(\delta) =$$
$$= \begin{cases} a_0\cdot\mathbb{P}^\pi(\theta\in\Theta_0|\mathbf{x}), & \text{pokud } \delta(\mathbf{x}) = 0, \\ a_1\cdot\mathbb{P}^\pi(\theta\in\Theta_1|\mathbf{x}), & \text{pokud } \delta(\mathbf{x}) = 1. \end{cases}$$

Z toho vyplývá, že

$$\delta^\pi(\mathbf{x}) = \underset{\delta}{\operatorname{argmin}}\, \varrho = \begin{cases} 1 & \text{pokud } a_1\mathbb{P}^\pi(\theta \in \Theta_1|\mathbf{x}) < a_0\mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}), \\ 0 & \text{pokud } a_1\mathbb{P}^\pi(\theta \in \Theta_1|\mathbf{x}) \geqslant a_0\mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}), \end{cases}$$

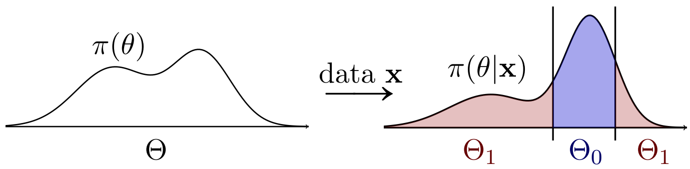
kde horní řádek přepíšeme na

$$a_1\mathbb{P}^\pi(\theta \in \Theta_1|\mathbf{x}) = a_1\big(1 - \mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x})\big) < a_0\mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}),$$

z čehož spočítáme

$$\mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}) > \frac{a_1}{a_0 + a_1}.$$

□



Obrázek 14.1: Význam Bayesovského testování H_0 oproti H_1. Vypočítáme tedy pravděpodobnost, že se parametr θ nachází v intervalu Θ_0 a zjistíme, jestli je tato hodnota větší, než $\frac{a_1}{a_0+a_1}$. (Spočítáme plochu aposteriorní hustoty nad intervalem Θ_0.)

Poznámka 14.2. Bayes umožňuje symetrizovat rozhodovací úlohu ($a_0 = a_1$) nebo penalizovat v závislosti na úloze, experimentu nebo "datech".
Bayesovské testování hypotéz se vyhýbá nutnosti nastavení $\alpha \in (0,1)$, kde α ovlivňuje β sílu testu.

Definice 14.3. Definujeme **Bayesovský faktor** jako

$$B^\pi(\mathbf{x}) := \frac{\frac{\mathbb{P}^\pi(\theta\in\Theta_0|\mathbf{x})}{\mathbb{P}^\pi(\theta\in\Theta_1|\mathbf{x})}}{\frac{\mathbb{P}^\pi(\theta\in\Theta_0)}{\mathbb{P}^\pi(\theta\in\Theta_1)}}.$$

Poznámka 14.4. Říká nám, jaký je vliv dat $\mathbf{x}$ na testování hypotéz oproti tomu, když $\mathbf{x}$ vůbec nepoužiji. Pokud $B^\pi(\mathbf{x}) > 1$, tak se data oproti apriornímu testování spíše přiklánějí k H_0. Pokud by ovšem vyšlo například velmi vysoké číslo, neznamená to nutně, že přijímáme hypotézu H_0, protože tato hodnota je jen relativní příklon k hypotéze H_0 oproti apriornímu testování.

Poznámka 14.5. Mějme například $\Theta_0 = \{\theta_0\}$ a $\Theta_1 = \{\theta_1\}$ jednoprvkové. Potom

$$\mathbb{P}^\pi(\theta \in \Theta_0) = \mathbb{P}(\theta = \theta_0) =: \pi_0, \quad \mathbb{P}^\pi(\theta \in \Theta_1) = \mathbb{P}(\theta = \theta_1) =: \pi_1$$

a

$$\mathbb{P}^\pi(\theta = \theta_0|\mathbf{x}) = \pi(\theta_0|\mathbf{x}) = \frac{f_0\pi_0}{f_0\pi_0 + f_1\pi_1}, \quad \mathbb{P}^\pi(\theta = \theta_1|\mathbf{x}) = \pi(\theta_1|\mathbf{x}) = \frac{f_1\pi_1}{f_0\pi_0 + f_1\pi_1}.$$

Z toho pak $B^\pi(\mathbf{x}) = \frac{\frac{f_0\pi_0}{f_1\pi_1}}{\frac{\pi_0}{\pi_1}} = \frac{f_0}{f_1} = \frac{L_0}{L_1}$, což je věrohodnostní poměr - *Likelihood ratio.*

Pro obecnou úlohu označme $\varrho_0 := \mathbb{P}^\pi(\theta \in \Theta_0)$ a $\varrho_1 := \mathbb{P}^\pi(\theta \in \Theta_1) = 1 - \varrho_0$. Přijímáme H_0, tedy $\delta(\mathbf{x}) = 1$, pokud

$$a_1 \cdot \mathbb{P}^\pi(\theta \in \Theta_1|\mathbf{x}) < a_0 \cdot \mathbb{P}^\pi(\theta \in \Theta_0|\mathbf{x}).$$

Potom po úpravě a vydělení $\frac{\varrho_0}{\varrho_1}$ dostáváme

$$\frac{\frac{a_1}{a_0}}{\frac{\varrho_0}{\varrho_1}} < \frac{\frac{\mathbb{P}^\pi(\theta\in\Theta_0|\mathbf{x})}{\mathbb{P}^\pi(\theta\in\Theta_1|\mathbf{x})}}{\frac{\varrho_0}{\varrho_1}} = B^\pi(\mathbf{x}).$$

Pro rozhodování o H_0 tedy můžeme použít Bayesovský faktor, ale v tomto formátu.

Poznámka 14.6 (Sekvenční Bayesovské testování hypotéz). Máme apriorní pravděpodobnosti ϱ_0, ϱ_1, naměříme data $\mathbf{x}$. Z toho pak vyrobíme aposteriorní hustotu $\pi(\theta|\mathbf{x})$, a pak se rozhodujeme mezi H_0 a H_1. Poté můžeme sekvenčně pokračovat dál tak, že vezmeme $\widetilde{\pi}(\theta) = \pi(\theta|\mathbf{x})$ a naměříme další data $\widetilde{\mathbf{x}} \sim \widetilde{f}$. Potom získáme aposteriorní hustotu $\widetilde{\pi}(\theta|\widetilde{\mathbf{x}}, \mathbf{x})$ a rozhodneme se mezi H_0 a H_1. Znovu můžeme sekvenčně pokračovat tak, že $\widetilde{\widetilde{\pi}}(\theta) = \widetilde{\pi}(\theta|\widetilde{\mathbf{x}}, \mathbf{x})$ a pokračujeme dále analogicky.

14.1 Aspekt Bayesovského testování hypotéz

Testujeme $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Nastavíme $\theta \sim \pi(\theta)$ a předpokládáme, že je tato hustota spojitá. Ve spojitém modelu je pravděpodobnost, že náhodná veličina θ nabude jedné izolované hodnoty, nulová, tedy

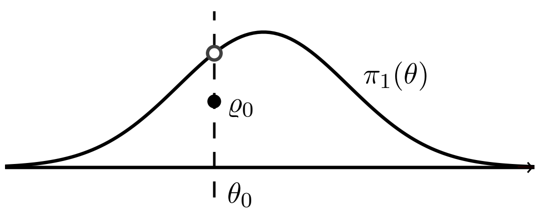
$$\varrho_0 = \mathbb{P}^\pi(\theta = \theta_0) = 0 \quad \Rightarrow \quad \mathbb{P}^\pi(\theta = \theta_0|\mathbf{x}) = \frac{f_0\pi_0}{c} = 0.$$

Totéž se dá analogicky ukázat pro 1. Z toho zároveň vyplývá, že $\varrho_0 = \begin{cases} 0 \\ 1 \end{cases}$ je absorbující stav, protože se z něj nedostaneme, ať už naměříme libovolná nová data. Dále víme, že $\pi(\theta)$ je konstantní, tudíž nepreferujeme žádnou konkrétní hodnotu $\theta \in \Theta$.

Příklad 14.7. Mějme H_0 hypotézu, že zítra bude pršet s pravděpodobností rovnu 0.7 a H_1 jako $\neq 0.7$. Toto testování nemá smysl, protože např. $0.700001 \neq 0.7$ a není možné při kvalitních datech testovat jednu hodnotu z celého kontinua.
Jak na to? Buďto můžeme testovat interval ve tvaru

$$H_0 : \theta \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon) \text{ vs. } H_1 : \theta \notin (\theta_0 - \varepsilon, \theta_0 + \varepsilon),$$

nebo můžeme volit $\pi(\theta) = \varrho_0 \cdot \delta_{\theta_0} + (1 - \varrho_0) \cdot \pi_1(\theta)$, kde ϱ_0 volíme větší než 0 (to si nastavujeme sami), případně menší než 1 (pak $\varrho_1 = 1 - \varrho_0$), a δ značí Diracovo δ, viz obr. 14.2.



Obrázek 14.2: Hodnota ϱ_0 vyjadřuje pravděpodobnost, že věříme právě této hodnotě θ_0, a na zbytku definičního oboru máme k dispozici apriorní hustotu $\pi_1(\theta)$. Tímto způsobem tedy vkládáme do BTH dodatečnou apriorní informaci.

Při takto upravené apriorní hustotě můžeme spočítat aposteriorní hustotu jako

$$\pi(\theta_0|\mathbf{x}) = \mathbb{P}^\pi(\theta = \theta_0|\mathbf{x}) = \frac{f_0 \cdot \pi(\theta_0)}{\int_\Theta f_X \cdot \pi \,\mathrm{d}\theta} = \frac{f_0 \cdot \varrho_0}{f_0 \cdot \varrho_0 + (1-\varrho_0)\underbrace{\int_{\Theta_1} f \cdot \pi(\theta)\,\mathrm{d}\theta}_{m_1(\mathbf{x})}} =$$

$$= \left[1 + \frac{1-\varrho_0}{\varrho_0}\frac{m_1(\mathbf{x})}{f_0(\mathbf{x})}\right]^{-1} < \frac{a_1}{a_0 + a_1}.$$

Z toho se dá potom vypočítat Bayesovský faktor jako $B^\pi(\mathbf{x}) = \frac{f_0(\mathbf{x})}{m_1(\mathbf{x})}$.

Příklad 14.8. Mějme $X \sim \mathcal{N}(\mu, 1)$ a chceme testovat $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$, kde $\mu \in \mathbb{R}$. Víme, že $\varrho_0 = \frac{1}{2}$, a tedy $\varrho_1 = \frac{1}{2}$ (tzn. H_0 věříme na 50%). Potom volíme

a) nejprve

$$\pi(\mu) = \frac{1}{2}\delta_0 + \frac{1}{2}\pi_1(\mu)\ , \text{ kde } \pi_1(\mu) = 1 \text{ na } \mu \neq 0 \quad \text{(nevlastní apriorní hustota).}$$

Potom po dosazení dostaneme

$$\pi(\mu = 0|\underbrace{\mathbf{x}}_{=x}) = \frac{1}{1 + \sqrt{2\pi}\mathrm{e}^{x^2/2}} \overset{\forall x \in \mathbb{R}}{\leqslant} \frac{1}{1+\sqrt{2\pi}} = 0.285, \quad \text{tzn. vždy zamítáme } H_0!$$

b) Nyní budeme předpokládat, že $\pi_1(\mu) = \mathcal{N}(m = 0, \tau^2)$ (vlastní), $\varrho = \frac{1}{2}$. Oproti minulému příkladu, kdy jsme o rozdělení μ vůbec nic nevěděli najednou máme k dispozici alespoň normální rozdělení. Z toho už se něco dá vyvozovat. Platí tedy, že

$$\pi(\mu = 0|\mathbf{x}) = [...]^{-1} \overset{\tau \to +\infty}{\longrightarrow} 1.$$