IV. Bayesovské odhady (Bayesovská statistika)

parametrický model

$X_1, \dots, X_n$ je náhodný výběr s distr. fcí $F(x, \theta)$ a hustotou $f(x, \theta)$

apriorní rozdělení parametru θ (θ je náhodná veličina)

další informace o parametru θ:
domněnky
zkušenosti
dřívější pozorování

aposteriorní rozdělení parametru θ $\Longrightarrow$ závěry o parametru θ (bodové odhady, intervalové odhady, ...)

Bayesovský přístup využívá kromě původních dat ještě informaci o parametru θ, kterou máme k dispozici nezávisle na našich datech:

- objektivní (informace z podobných úloh, z minulosti)
- subjektivní (názor, zkušenost)

Tato mimodatová informace o hodnotě neznámého parametru θ je vyjádřena pomocí pravd. rozdělení $\Rightarrow$ param. θ považujeme za náhodnou veličinu.

Výhody: využití apriorní informace; lze použít i pro malý počet pozorování

Nevýhody: volba apriorního rozdělení (subjektivita) $\Rightarrow$ rozdílné výsledky, nemožnost ověření správnosti

Příklad (lékařská diagnostika)

Jistou nemocí trpí 3 % populace. Nemoc je diagnostikována na základě testu, který má senzitivitu 0,7 a specificitu 0,98. Určete pravděpodobnost správné diagnózy, pokud byl výsledek testu pozitivní, resp. negativní.

N = testovaná osoba je nemocná $\quad P(N)=0{,}03$ ... apriorní pravděpodobnost (apriorní informace)

$+$ = výsledek testu je pozitivní $\quad P(+|N)=0{,}7$ ... senzitivita testu

$-$ = výsledek testu je negativní $\quad P(-|\bar{N})=0{,}98$ ... specificita testu

$$\left.\begin{aligned} P(N|+) &= \frac{P(+|N)\cdot P(N)}{P(+|N)\cdot P(N)+P(+|\bar{N})\cdot P(\bar{N})} \doteq 0{,}52 \\[2ex] P(\bar{N}|-) &= \frac{P(-|\bar{N})\cdot P(\bar{N})}{P(-)} \doteq 0{,}99 \end{aligned}\right\} \text{aposteriorní pravděpodobnosti}$$

<u>Matematický model</u>

$X_1,\dots,X_n$ je náh. výběr, kde X_i má hustotu $f(x,\theta)$, kde $\theta\in\Theta\subset\mathbb{R}^p$ je neznámý parametr

Θ je nyní p-rozměrný náh. vektor s hustotou $g(\theta)$ ... apriorní hustota

věrohodnost: $\ell(\mathbf{x}|\theta)=\ell(x_1,\dots,x_n|\theta)=\prod_{i=1}^{n} f(x_i,\theta)$ ... podmíněná hustota n. v. $(X_1,\dots,X_n)^T$ při daném Θ.

<u>Věta</u> (Bayesova)

Pro podmíněnou hustotu náh. vektoru Θ při daném $\mathbf{X}=(X_1,\dots,X_n)^T$ platí:

$$\pi(\theta|\mathbf{x})=\frac{\ell(\mathbf{x}|\theta)\cdot g(\theta)}{\int_\Theta \ell(\mathbf{x}|\theta)\cdot g(\theta)\,d\theta} \qquad \text{je-li } \int_\Theta \ell(\mathbf{x}|\theta)\cdot g(\theta)\,d\theta\neq 0$$

... aposteriorní hustota

Poznámka

Je-li Θ diskrétní náh. vektor s prav. funkcí $g(\Theta)$.

$$\pi(\Theta|x) = \frac{\pi(x|\Theta)\cdot g(\Theta)}{\sum_{\Theta\in\Theta} \pi(x|\Theta)\cdot g(\Theta)}$$ je aposteriorní prav. funkce

Příklad (IQ daného člověka)

Chceme určit IQ daného člověka. Necháme jej udělat test, jeho výsledek označíme X. X = výsledek je výsledek testu u daného člověka

$$X \sim N(\underset{\substack{\uparrow \\ \text{IQ daného člověka}}}{\Theta}, \underset{\substack{\uparrow \\ \text{variabilita výsledku}}}{100})$$

Cíl: Odhad parametru Θ.

Klasicky: $\hat{\Theta} = X \sim N(\Theta, 100)$ je nestranný odhad parametru Θ s rozptylem 100.

Bayesovsky: Hodnota IQ v populaci má normální rozdělení $N(100, 225)$, tedy $\Theta \sim N(100, 225)$ apriorní rozdělení parametru Θ

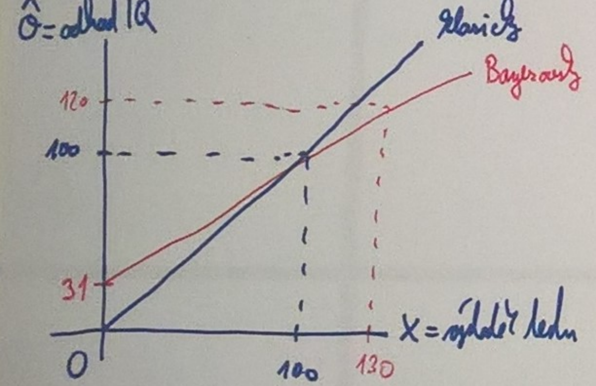
$$g(\Theta) = \frac{1}{\sqrt{2\pi\cdot 225}}\cdot \exp\left\{-\frac{(\Theta-100)^2}{2\cdot 225}\right\}, \Theta\in\mathbb{R}$$ ... apriorní hustota

$$\pi(x|\Theta) = f(x,\Theta) = \frac{1}{\sqrt{2\pi\cdot 100}}\cdot \exp\left\{-\frac{(x-\Theta)^2}{2\cdot 100}\right\}, x\in\mathbb{R}$$

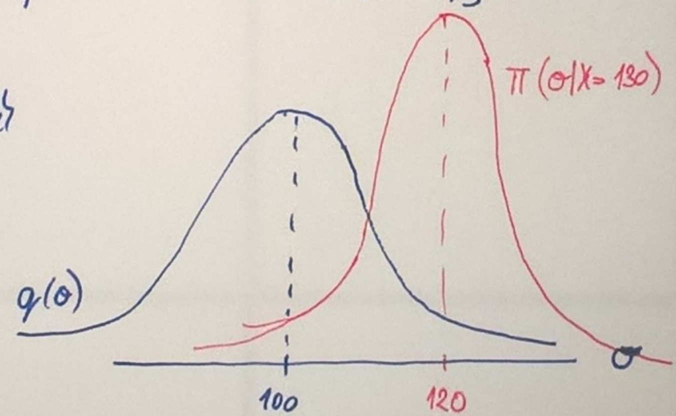
$$\pi(\theta|x) = \frac{\pi(x|\theta)\cdot q(\theta)}{\int_{\Theta} \pi(x|\theta)\cdot q(\theta)\,d\theta} = \frac{\frac{1}{\sqrt{2\pi\cdot 100}}\cdot \exp\left\{-\frac{(x-\theta)^2}{2\cdot 100}\right\}\cdot \frac{1}{\sqrt{2\pi\cdot 225}}\cdot \exp\left\{-\frac{(\theta-100)^2}{2\cdot 225}\right\}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\cdot 100}}\cdot \exp\left\{-\frac{(x-\theta)^2}{2\cdot 100}\right\}\cdot \frac{1}{\sqrt{2\pi\cdot 225}}\cdot \exp\left\{-\frac{(\theta-100)^2}{2\cdot 225}\right\} d\theta} = \dots =$$

$$= \frac{1}{\sqrt{2\pi\cdot\frac{900}{13}}}\cdot \exp\left\{-\frac{\left(\theta-\frac{400+9x}{13}\right)^2}{2\cdot\frac{900}{13}}\right\}$$ aposteriorní hustota, hustota $N\left(\frac{400+9x}{13}, \frac{900}{13}\right)$.

Bayesovsky bychom tedy IQ daného člověka odhadli pomocí střední hodnoty aposteriorního rozdělení, tj. $\frac{400+9x}{13}$, kde x je výsledek testu (realizace n. v. X)



Nechť daná osoba má výsledek testu $x = 130$.



Variabilita původního (klasického) odhadu byla 100, v bayesovském případě klesla na $\frac{900}{13} \doteq 69$.

Poznámka

$$\underbrace{\pi(\theta|x)}_{\text{funkce } \theta} = \frac{\pi(x|\theta)\cdot q(\theta)}{\underbrace{\int_{\Theta} \pi(x|\theta)\cdot q(\theta)\,d\theta}_{\text{nezávisí na } \theta \text{ (normující konstanta, aby } \pi(\theta|x) \text{ byla hustota)}}}$$

$\Rightarrow$ proto se Bayesova věta píše ve tvaru $\pi(\theta|x) \propto \pi(x|\theta)\cdot q(\theta)$

je přímo úměrná (je násobkem)

Ke mám umožněno je použití tzv. nevlastních hustot:

$q(\theta) \geq 0$, ale $\int_{\Theta} q(\theta) \neq 1$ (je typicky $+\infty$)

např. $q(\theta) = 1,\ \theta \in \mathbb{R}$, nebo $q(\theta) = 1$ pro $\theta > 0$.

Postup při bayesovském modelování:

1) $X_1, \dots, X_n$ je náh. výběr z rozdělení s hustotou $f(x, \theta)$, kde θ je neznámý parametr. Nechť $\mathbf{x} = (x_1, \dots, x_n)^T$ je realizace tohoto náh. výběru (naše data).
Definujeme sdruženou hustotu $(X_1, \dots, X_n)^T$ jako $\pi(\mathbf{x} | \theta) = \prod_{i=1}^{n} f(x_i, \theta)$.

2) Zvolíme apriorní hustotu $q(\theta)$.

3) Pomocí Bayesovy věty získáme aposteriorní rozdělení $\pi(\theta | \mathbf{x})$.

4) Pomocí $\pi(\theta | \mathbf{x})$ provádíme statistickou inferenci (odhady, testy, ...)

5) Posoudíme vhodnost našeho modelu.

Volba apriorního rozdělení

apriorní hustota $q(\theta)$ by měla odrážet naši apriorní informaci o parametru θ ← buď $q(\theta)$ vyplývá přímo z apriorní informace
- $q(\theta)$ odhadneme pomocí histogramu, nebo jádrového odhadu
- $q(\theta)$ splňuje nějaký parametrický model; my odhadneme jen jeho parametry

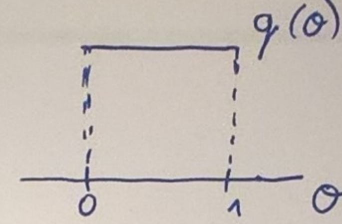
1) Rovnoměrné apriorní rozdělení – neinformativní

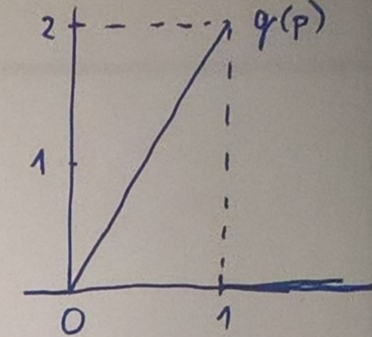
– nemáme žádnou informaci o parametru θ (všechny hodnoty jsou stejně „pravděpodobné")

$q(\theta) = 1$, $\theta \in \Theta$ může být i nevlastní hustota

Příklad

$X_1, \dots, X_n$ je náh. výběr z $A(\theta)$, $\theta \in (0,1)$





neinformativní apriorní hustota: $q(\theta) = 1$, $\theta \in (0,1)$.

změníme parametrizaci $p := \sqrt{\theta}$, $p \in (0,1)$ ... věta o transformaci $q(p) = 2p$, $p \in (0,1)$.

prostá transformace

příslušné apriorní hustoty se liší a aposteriorní taky => rozdílné výsledky.

2) Jeffreysova apriorní hustota - neinformativní

$q(\theta) \propto \sqrt{\det J(\theta)}$, kde $J(\theta)$ je Fisherova informační matice o parametru θ příslušná X_i.

Poznámka

Pro tuto volbu $q(\theta)$ je aposteriorní rozdělení $\pi(\theta|x)$ stejné pro všechny reparametrizace $\pi(\theta)$.

3) Konjugované apriorní rozdělení

Nechť apriorní hustota $q(\theta)$ patří do nějaké třídy hustot $\mathcal{F}$. Řekneme, že $q(\theta)$ je konjugovaná vzhledem k věrohodnosti $\pi(x|\theta)$, pokud i aposteriorní hustota $\pi(\theta|x)$ patří do $\mathcal{F}$.

Poznámka

$\mathcal{F}$ se volí tak, aby obsahovala co nejméně hustot.

Hledání se provádí např. za pomoci postačujících statistik.