Příklad (házení mincí)

20× hodíme mincí. Padlo nám celkem $n_1 = 5$ líců. Jaká je pravděpodobnost, že nám v daném hodu padne líc?

Klasicky

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

$$X_i \begin{cases} 1 & \dots \text{ v } i\text{-tém hodu padne líc} \\ 0 & \dots \text{ jinak} \end{cases}$$

bodový odhad θ:

$$\hat{\theta} = \overline{X} = \frac{n_1}{n} = \frac{5}{20} = 0{,}25.$$

interval spolehlivosti pro θ je: $\left(\hat{\theta} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}, \hat{\theta} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \right) \overset{\text{pro } \alpha = 0{,}05}{=} (0{,}06; 0{,}44)$

predikce budoucího pozorování X_{n+1}: $X_{n+1} \sim A(\hat{\theta})$: $P(X_{n+1} = 1 | X_1, \dots, X_n) = \hat{\theta} = 0{,}25$,

$P(X_{n+1} = 0 | X_1, \dots, X_n) = 1 - \hat{\theta} = 0{,}75$.

Bayesovsky

– potřebujeme apriorní rozdělení parametru θ (nezávislé na našich datech)

• scénář 1: mince bývají vyvážené, tedy nabývají hodnoty kolem 0,5

$\theta \sim Beta(100, 100)$

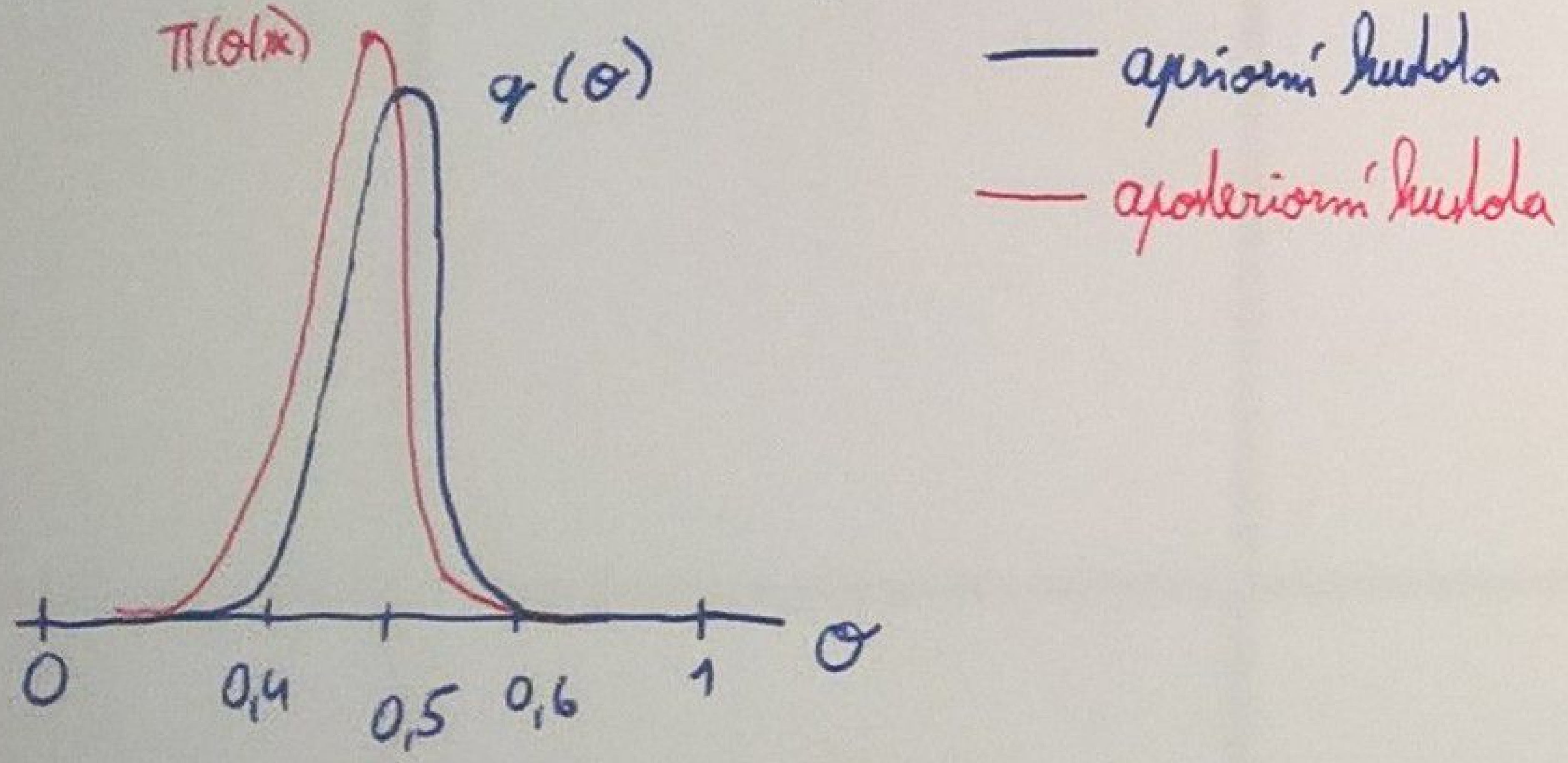
hustota rozdělení $Be(\alpha, \beta)$ je $g(\theta, \alpha, \beta) = \frac{1}{Be(\alpha, \beta)} \cdot \theta^{\alpha - 1} (1-\theta)^{\beta - 1}$, $0 < \theta < 1$, kde $Be(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}$ je beta funkce.

$f(x_i,\theta)=\theta^{x_i}\cdot(1-\theta)^{1-x_i}$, $x_i=0,1$

věrohodnost $\pi(\mathbb{x}|\theta)=\pi(x_1,\dots,x_n|\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\theta^{\sum x_i}\cdot(1-\theta)^{n-\sum x_i}=\theta^{n_1}\cdot(1-\theta)^{n-n_1}$

Bayesova věta: $\pi(\theta|\mathbb{x})\propto \underbrace{\frac{1}{Be(100,100)}\cdot\theta^{99}(1-\theta)^{99}}_{q(\theta)}\cdot\underbrace{\theta^{n_1}\cdot(1-\theta)^{n-n_1}}_{\pi(\mathbb{x}|\theta)}\propto\theta^{99+n_1}\cdot(1-\theta)^{99+n-n_1}$

Tedy $\theta|\mathbb{x}\sim Beta(100+n_1,100+n-n_1)=Beta(105,115)$ aposteriorní rozdělení je také beta rozdělením (je to konjugované rozdělení).



My chceme ale bodový odhad parametru θ.

Bayesovská rozhodovací teorie - teorie ztrátových funkcí

$L:\mathbb{R}\times\mathbb{R}\mapsto\mathbb{R}$ je ztrátová funkce, jestliže $L(x,x)=0$, $L(x,y)\geq 0$, $\forall x,y\in\mathbb{R}$

$L(\theta,\hat{\theta})$ ... ztráta, kterou utrpíme, je-li skutečná hodnota parametru rovna θ a my ho odhadujeme pomocí $\hat{\theta}$.

bayesovské riziko: $r(\hat{\theta})=\int_{\Theta} L(\theta,\hat{\theta})\,\pi(\theta|x)\,d\theta$ ... průměrná aposteriorní ztráta ⇒ hledáme odhad, který pro danou ztrátovou funkci L minimalizuje bayesovské riziko

Příklady:

- $L(\theta,\hat{\theta}) = (\theta-\hat{\theta})^2$ ... kvadratická ztrátová funkce
  $\hat{\theta}_1 = \underset{\hat{\theta}}{\arg\min} \int_{\Theta} (\theta-\hat{\theta})^2 \pi(\theta|x)\,d\theta = E(\theta|X_1,\dots,X_n)$ aposteriorní střední hodnota

- $L(\theta,\hat{\theta}) = |\theta-\hat{\theta}|$ ... absolutní ztrátová funkce
  $\hat{\theta}_2 = \text{medián}\,(\theta|X_1,\dots,X_n)$ aposteriorní medián

- $L(\theta,\hat{\theta}) = \mathbb{1}\{\theta \neq \hat{\theta}\}$ ... zero-one ztrátová funkce
  $\hat{\theta}_3 = \underset{\theta\in\Theta}{\arg\max}\, \pi(\theta|X_1,\dots,X_n)$ aposteriorní modus (metoda maximální aposteriorní věrohodnosti)

Intervalový odhad θ

Klasicky: θ je neznámé: $P(L(X_1,\dots,X_n) \le \theta \le H(X_1,\dots,X_n)) = 1-\alpha, \ \forall\theta$ $[L,H]$ je $(1-\alpha)\cdot 100\%$ interval spolehlivosti pro θ

Bayesovsky: θ je náh. veličina $\Rightarrow$ má smysl počítat pravděpodobnost, že parametr θ náleží danému intervalu (věrohodnostní interval, credibility interval)

$[a,b] = [\overset{a(X)}{\overset{\shortparallel}{a(X_1,\dots,X_n)}}, \overset{b(X)}{\overset{\shortparallel}{b(X_1,\dots,X_n)}}]$ je $(1-\alpha)\cdot 100\%$ věrohodnostní interval pro parametr θ, jestliže $P(a(X)\le\theta\le b(X)\,|\,X) = 1-\alpha$.

- $100(1-\alpha)\%$ symetrický věrohodnostní interval (equal tail) je interval $[a,b]$, kde $a = \frac{\alpha}{2}$ - kvantil aposteriorního rozdělení $\theta|X$
  $b = 1-\frac{\alpha}{2}$ - kvantil −||−

- $100(1-\alpha)\%$ věrohodnostní interval o nejvyšší aposteriorní hustotě (HPD) je interval $[a,b]$ takový, že $\pi(\theta|X) \ge c, \ \forall\theta\in[a,b]$ a c je největší číslo takové, že

$P(\pi(\theta|X)\ge c) = 1-\alpha$.

Tvrzení

Je-li $\pi(\theta|x)$ spojitá a unimodální, pak HPD interval je nejkratší mezi všemi $100\cdot(1-\alpha)\%$ věrohodnostními intervaly.

Predikce budoucího pozorování

chceme odhadnout hodnotu následující náh. veličiny X_{n+1} splňující stejný model jako původní data $X_1,\dots,X_n$.

Tedy X_{n+1} má hustotu $f(x_{n+1},\theta)$

aposteriorní prediktivní hustota $f(x_{n+1}|x)=\int_{\Theta} f(x_{n+1},\theta)\cdot\pi(\theta|x)\,d\theta$.

Příklad - pokračování

$X\sim Beta(\alpha,\beta) \Rightarrow EX=\frac{\alpha}{\alpha+\beta}$, modus $X=\frac{\alpha-1}{\alpha+\beta-2}$, medián $X\doteq\frac{\alpha-\frac{1}{3}}{\alpha+\beta-\frac{2}{3}}$

v našem případě: aposteriorní rozdělení $\theta|x\sim Beta(105,115)$

$\hat{\theta}_1=E(\theta|x)=\frac{105}{220}\doteq 0,477$

$\hat{\theta}_2=\text{medián}\,(\theta|x)\doteq 0,477$

$\hat{\theta}_3=\text{modus}\,(\theta|x)\doteq 0,477$

95% equal tail věr. interval je $[0,412;0,543]$,

95% HPD věr. interval je $[0,419;0,535]$.

predikce budoucího pozorování X_{n+1}: $f(x_{n+1},\theta)=\theta^{x_{n+1}}\cdot(1-\theta)^{1-x_{n+1}}$, pro $x_{n+1}=0,1$.

$$f(x_{n+1}|x)=\int_0^1\theta^{x_{n+1}}\cdot(1-\theta)^{1-x_{n+1}}\cdot\frac{\theta^{104}\cdot(1-\theta)^{114}}{Be(105,115)}\,d\theta=\frac{1}{Be(105,115)}\cdot\int_0^1\theta^{104+x_{n+1}}\cdot(1-\theta)^{115-x_{n+1}}\,d\theta=\frac{Be(105+x_{n+1},116-x_{n+1})}{Be(105,115)},\ x_{n+1}=0,1.$$

$f(0|\mathbf{x}) = P(X_{n+1} = 0|\mathbf{x}) = 0{,}522$

$f(1|\mathbf{x}) = P(X_{n+1} = 1|\mathbf{x}) = 0{,}478$

scénář 2: žádná apriorní informace o parametru θ

$g(\theta) = 1, \quad 0<\theta<1 \ldots \quad \theta \sim Rs(0,1) = Beta(1,1)$

scénář 3: Jeffreysova apriorní hustota

$J(\theta) = \dfrac{1}{\theta(1-\theta)}$ je Fisherova informace příslušná X_i

$g(\theta) \propto \sqrt{|J(\theta)|} = \theta^{-\frac{1}{2}} \cdot (1-\theta)^{-\frac{1}{2}}, \quad 0<\theta<1 \quad \ldots \quad \theta \sim Beta\left(\frac{1}{2}, \frac{1}{2}\right)$

Závěrečné poznámky

Bayesova věta: $\pi(\theta|x) = \dfrac{\pi(x|\theta)\cdot g(\theta)}{\underbrace{\int_{\Theta} \pi(x|\theta)\cdot g(\theta)\, d\theta}_{\text{výpočet tohoto integrálu může být náročný}}}$

- numerická integrace
- simulace pomocí Monte Carlo metody

Příklad

výpočet integrálu $\int_0^1 x^9 \, dx = \frac{1}{10}$

$X_1, \dots, X_m$ je náh. výběr z $Rs(0,1)$, definujeme $I_m = \frac{1}{m} \sum_{i=1}^{m} X_i^9 \xrightarrow[m \to \infty]{s.j.} \int_0^1 x^9 \, dx = \frac{1}{10}$. Proč to funguje?

Silný zákon velkých čísel: $X_1, \dots, X_m$ je náh. výběr, pak $\frac{1}{m} \sum_{i=1}^{m} h(X_i) \xrightarrow[m \to \infty]{s.j.} E\, h(X_1)$.

Použití:

- Nechť $\Theta_1, \dots, \Theta_m$ je náh. výběr z apriorního rozdělení s hustotou $q(\theta)$.
- Integrál $I = \int_{\Theta} \pi(x|\theta) \cdot q(\theta) \, d\theta$ aproximujeme pomocí $I_m = \frac{1}{m} \sum_{i=1}^{m} \pi(x|\Theta_i)$ (SZVČ: $I_m \xrightarrow[m \to \infty]{s.j.} I$).

Výpočet aposteriorní střední hodnoty

$$E(\Theta|x) = \int_{\Theta} \theta \cdot \pi(\theta|x) \, d\theta = \int_{\Theta} \theta \frac{\pi(x|\theta) \cdot q(\theta)}{\int_{\Theta} \pi(x|\theta) \cdot q(\theta) \, d\theta} \, d\theta = \frac{\overbrace{\int_{\Theta} \theta \cdot \pi(x|\theta) \cdot q(\theta) \, d\theta}^{J}}{\underbrace{\int_{\Theta} \pi(x|\theta) \cdot q(\theta) \, d\theta}_{I}}.$$

Integrál J aproximujeme pomocí $J_m = \frac{1}{m} \cdot \sum_{i=1}^{m} \Theta_i \cdot \pi(x|\Theta_i)$.