Cvičení 1
částečně 2

2014 + 15

1.7 An examination has 15 questions, each with 3 possible answers. Assume that 70% of the students taking the examination are prepared and answer correctly each question with probability 0.8; the remaining 30% answer at random.

a. Characterize the distribution of S, score of a student if one point is attributed to each correct answer.

b. Eight correct answers are necessary to pass the examination. Given that a student has passed the examination, what is the probability that she was prepared?

Example 1.2.2 (Bayes (1764)) A billiard ball W is rolled on a line of length one, with a uniform probability of stopping anywhere. It stops at $\cancel{p}$ u. A second ball O is then rolled n times under the same assumptions and $\cancel{X}$ Y denotes the number of times the ball O stopped on the left of W. *Given* Y $\cancel{X}$, *what inference can we make on* $\cancel{p}$ u?

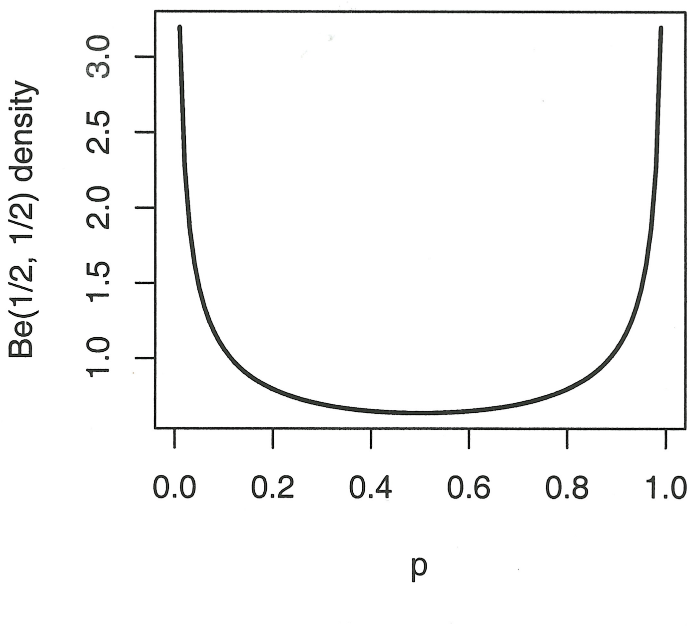
Example 1.2.3 (Laplace (1773)) An urn contains a number n of black and white cards. If the first card drawn out of the urn is white, what is the probability that the proportion $\cancel{p}$ π of white cards is $\cancel{p_0}$ π_0?

Example 1.2.4 (Laplace (1786)) Considering male and female births in Paris, Laplace wants to test whether the probability $\cancel{x}$ π of a male birth is above 1/2. For 251,527 male and 241,945 female births, assuming that x has a uniform prior distribution on [0, 1], Laplace obtains

$$P(\cancel{x}\,\pi \leq 1/2|(251,527; 241,945)) = 1.15 \times 10^{-42}.$$ ← Jak to spočítal?

(see Stigler (1986, p. 134) and Exercise 1.6). He then deduces that this probability x is more than likely to be above 50%. Still assuming a uniform prior distribution on this probability, he also compares the male births in London and Paris and deduces that the probability of a male birth is significantly higher in England. ||

Jeffreys prior pro π



Bayes věta

A, E náhodné jevy takové, že $P(E) \neq 0$

$$\underset{\substack{\uparrow\\ \text{příčina}}}{P(A}\,|\,\underset{\substack{\uparrow\\ \text{následek}}}{E)} = \frac{P(E|A)\,P(A)}{P(E|A)\,P(A) + P(E|A^c)\,P(A^c)} =$$

$$= \frac{P(E|A)\cdot P(A)}{P(E)}$$

$$\underbrace{\frac{P(A|E)}{P(B|E)}}_{\text{odds poté, co pozorujeme data}} = \frac{\dfrac{P(E|A)\,P(A)}{P(E)}}{\dfrac{P(E|B)\,P(B)}{P(E)}} = \frac{P(E|A)}{P(E|B)} \cdot \underbrace{\frac{P(A)}{P(B)}}_{\text{odds před daty}}$$

A, B = příčina (model, parametr)

E = následek (data)

Cvič. 1.7, BCh. str. 33

Test má 15 otázek se 3 možnými odpověďmi (a/b/c).

70% připraveno → prst. správné odpovědi je 0,8

30% odpovídá náhodně

Jaké je rozdělení počtu správných odpovědí, které získá náhodně vybraný student?

S = počet správných odpovědí

a) $P(S=x \mid \text{připraven}) = \binom{15}{x} 0{,}8^x \, 0{,}2^{15-x}$

následek příčina

$\sim Bi(15, 0.8)$

b) $P(S=x \mid \text{nepřipraven}) = \binom{15}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{15-x}$

$\sim Bi(15, 1/3)$

$\Rightarrow P(S=x) = \underbrace{P(S=x \mid \text{připr.}) \underset{0,7}{P(\text{připr.})} + P(S=x \mid \text{nepř.}) \underset{0,3}{P(\text{nepřipr.})}}_{\text{směs dvou binom. rozdělení}}$

jmenovatel v Bayes. větě

Na složení zkoušky je potřeba 8 správných odpovědí. Jaká je pravděpodobnost, že student, který složil zkoušku, je připravený?

$P(\text{připraven} \mid S \geq 8) = \dfrac{P(S \geq 8 \mid \text{připr.}) \cdot P(\text{připr.})}{P(S \geq 8)} =$

Dosaď 0,9634

2

Bayes věta (spojitá verze)

$$g(x|y) = \frac{f(y|x)\, g(x)}{\int \underbrace{f(y|x)\, g(x)}_{h(y,x)}\, d\lambda(x)} = \frac{f(y|x)\, g(x)}{f(y)},$$

pokud $f(y) \neq 0$

$= 0$ jinak

JINÉ SYMBOLY:

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{\int_{\Theta} p(y|\theta) \cdot p(\theta)\, d\lambda(\theta)} = \frac{p(y|\theta)\, p(\theta)}{p(y)},$$

pokud $p(y) \neq 0$

$p(y|\theta) = L(\theta)$ = věrohodnost (likelihood)

$p(\theta)$ = apriorní rozdělení

Θ

$$p(y) = \int_{\Theta} p(y|\theta)\, p(\theta)\, d\lambda(\theta)$$

= marginální / integrovaná věrohodnost,
(marginal) (integrated)

prediktivní hustota

Je v praxi potřeba zabývat se podmínkou $p(y) \neq 0$?

UŽITEČNÉ SI UVĚDOMIT (pro počítání):

$$p(\theta|y) \propto p(y|\theta)\, p(\theta) = L(\theta)\, p(\theta)$$

proporcionální vzhledem k θ

"normující" konstantu $\int L(\theta) p(\theta)\, d\mu(\theta)$ stačí spočítat na konec

---

TYPICKY: $L(\theta) = \prod_{i=1}^{n} L_i(\theta)$

DATA: $Y = (Y_1^T, \dots, Y_n^T)^T$, $L_i(\theta) = p(y_i|\theta)$

$$p(\theta|y) \propto \underbrace{\prod_{i=1}^{n} L_i(\theta)}_{\text{info nových dat}}\; \underbrace{p(\theta)}_{\text{„věrohodnost" starých dat (dříve známá informace o } \theta)}$$

Př. 1 $\quad Y = (Y_1, \dots, Y_n)$, $\quad Y_i \sim N(\underbrace{\mu, \sigma^2}_{\theta})$ (MODEL)

$$L_i(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

Př. 2 $\quad Y = (Y_1, \dots, Y_n)$, $\quad Y_i \sim N(x_i^T \beta, \sigma^2)$, $\quad x_i \in \mathbb{R}^k$ pevné

$$L_i(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i^T\beta)^2}{2\sigma^2}\right) \qquad \theta = (\beta^T, \sigma^2)$$

Př. 3 $\quad Y = (Y_1, \dots, Y_n)$, $\quad Y_i \sim N_{n_i}(X_i\beta, \underbrace{Z_i^T D Z_i + \sigma^2 I_{n_i}}_{\Sigma_i})$

X_i, Z_i: matice konstant $\qquad$ (LMM: lineár. smíš. model)

$\theta \equiv \beta, D, \sigma^2$

$$L_i(\theta) = (2\pi)^{-\frac{n_i}{2}} |\Sigma_i|^{-1/2} \exp\left(-\frac{1}{2}(y_i - X_i\beta)^T \Sigma_i^{-1} (y_i - X_i\beta)\right)$$

Cvič. 1.2.2, BCh str. 10

$U \sim U(0,1)$ (rovnoměrné rozdel.) $U \equiv$ parametr

$\equiv$ apriorní rozdel.

$V_i =$ místo, kde m zastaví při i-tém pokusu

$$Y = \sum_{i=1}^{n} \mathbb{I}(V_i < U) \equiv \text{data}$$

$\mathbb{I}(V_i < U) \mid U \sim Alt(U)$ (alternativní)

$P(V_i < U \mid U) = U$

$Y = \sum_{i=1}^{n} \mathbb{I}(V_i < U)$, $Y \mid U \sim Bi(n, U)$ (binomické)

$P(Y = y \mid U) = \binom{n}{y} U^y (1-U)^{n-y}$ $\equiv$ věrohodnost

na okraj:

$$P(Y=y) = \int_0^1 P(Y=y \mid U=u)\, p(u)\, du =$$

$$= \int_0^1 \binom{n}{y} u^y (1-u)^{n-y} \cdot 1 \, du = \int_0^1 \binom{n}{y} u^y (1-u)^{n-y} \, du$$

$$= \binom{n}{y} \int_0^1 u^y (1-u)^{n-y} \, du = \binom{n}{y} B(y+1, n-y+1)$$

$$p(u \mid y) \propto P(Y=y \mid U=u)\, p(u) = \binom{n}{y} u^y (1-u)^{n-y} \cdot 1$$

$$\propto \underbrace{u^y (1-u)^{n-y}}$$

poznávám jádro Beta hustoty

→ netřeba se explicitně trápit s normující konstantou tj: $U | Y \sim Be(Y+1, n-Y+1)$

$\Rightarrow P(a < U < b \mid Y=y) = F_y(b) - F_y(a)$,

kde F_y = distr. fce $Be(y+1, n-y+1)$

$$E(U \mid Y=y) = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2} =: \hat{U}$$

BUDE: věrohodnostní (credible) interval:

$1-\alpha = P(U \in I_\alpha \mid Y=y)$

jak např. zde?

$I_\alpha = \left(Q_y\left(\frac{\alpha}{2}\right), Q_y\left(1-\frac{\alpha}{2}\right)\right)$, $Q_y(p)$ = p-kvantil $Be(\dots)$ rozdělení

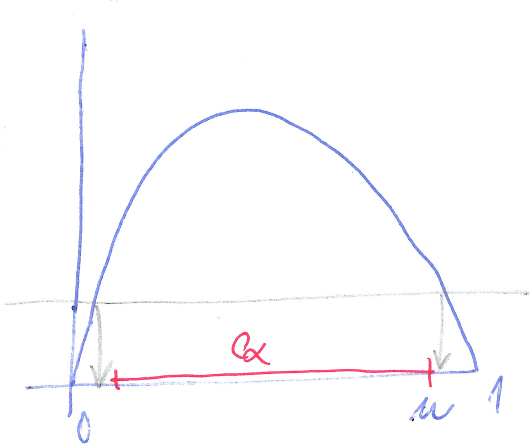
= equal-tail (ET) interval

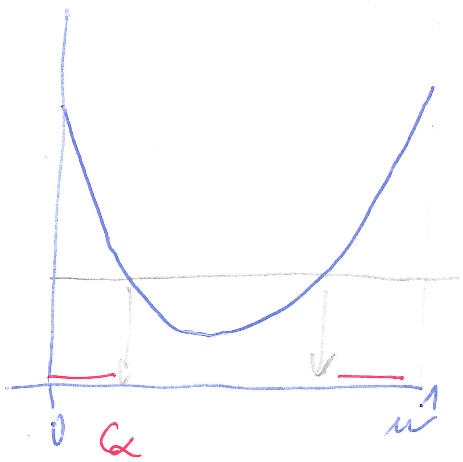
highest posterior density (HPD) věrohodn. množina:

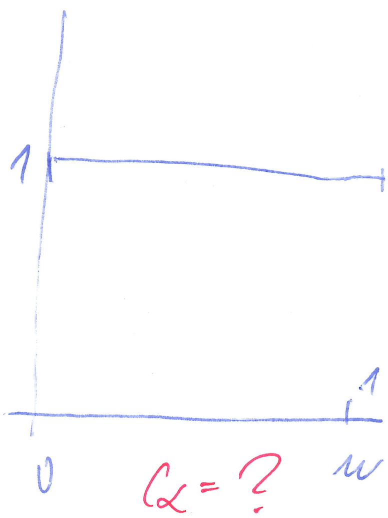
C_α taková, že $P(U \in C_\alpha \mid Y=y)$ &

pro skoro všechna $u_1 \in C_\alpha$, $u_2 \notin C_\alpha$

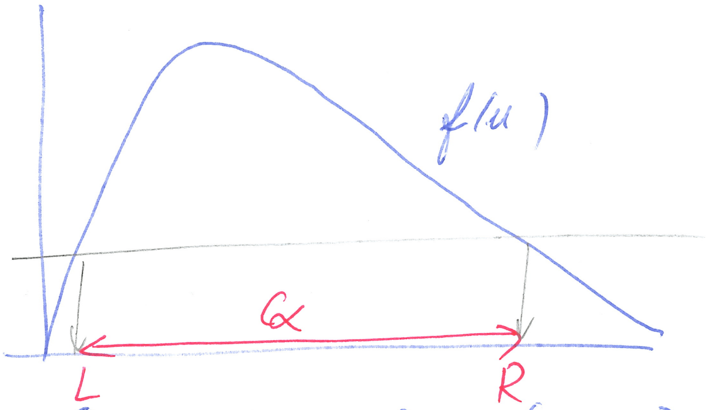
$p(u_1 \mid y) \geq p(u_2 \mid y)$







Jak spočítat HPD věrohodn. množinu, pokud vím, že se jedná o interval?



$$F(u) := \int_0^u f(s)\,ds$$

$0 \le u \le 1$

VÍM: $C_\alpha = (L, R)$

1) $f(L) = f(R)$

2) $F(R) - F(L) = 1 - \alpha$ } 2 rovnice o 2 neznámých

---

drobná komplikace?

‖

1) $f(L) - f(R) = 0$

2) $\int_L^R f(s)\,ds = 1 - \alpha$

~~další drobná komplikace~~?

NUMERICKY:

hledej $l < r$, které $\min\ |f(l) - f(r)|^2 + |\int_l^r f(s)\,ds - 1 + \alpha|^2$

atp.