BAYESOVSKÉ ODHADY
V NĚKTERÝCH MODELECH

Michal Friesl

Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

•Slunce •Řidiči •IQ •Regrese •Přežití

Obvyklý model

Pozorování $X = (X_1, \ldots, X_n)$ s hustotou $f(x) = f(x; \theta)$

- známe tvar f až na několik parametrů — θ,
- např. výběr z $\mathrm{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$ neznáme

Chceme učinit závěry týkající se θ

- bodový odhad,
- intervalový odhad,
- testy hypotéz, ...

Klasicky

- θ neznámá, ale pevná konstanta
- k závěrům použijeme tvar hustoty $f(x; \theta)$ a pozorování X
  - » odhad metodou maximální věrohodnosti, momentový, ...

Bayesovský přístup

Bayesovsky

- θ je náhodná veličina se známým, apriorním rozdělením
- $f(x;\theta)$ chápeme jako podmíněnou hustotu $f(x \mid \theta)$
- k závěrům použijeme tvar hustoty $f(x;\theta)$, pozorování X, a navíc $\pi(\theta)$, hustotu parametru θ

Smysl $\pi(\theta)$

- vyjadřuje apriorní informaci o možných hodnotách θ
  » získanou před pokusem, tedy nezávisle na pozorováních X
- mnohdy zvolena objektivně, z minulé zkušenosti
- možno též subjektivně, nebo aby „to šlo spočítat“

Bayesova věta

- pro dvojici veličin (X, Y) je $f(y \mid x) = \dfrac{f(x \mid y) f(y)}{f(x)}$
  - » kde $f(x)$ a $f(y)$ jsou marginální hustoty
- v našem označení $\boxed{\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)}$

Aposteriorní rozdělení $\pi(\theta \mid x)$

- aktualizovaná představa o hodnotách θ
- kombinuje $\pi(\theta)$ a věrohodnostní funkci $L(\theta) = f(x; \theta) = f(x|\theta)$
- používá se i nevlastní apriorní hustota, $\int \pi(\theta) = \infty$

Závěry z aposteriorního rozdělení

- odhad max. věrohodného typu: $\pi(\theta \mid x) = \max_{\theta \in \mathbf{R}}$
- odhad střední hodnotou $\widehat{\theta} = \mathrm{E}(\theta \mid X = x)$, mediánem
  - » podle ztrátové funkce, zde kvadratická, resp. absolutní hodnota
- bayesovská konfidenční oblast, pravděpodobnosti hypotéz

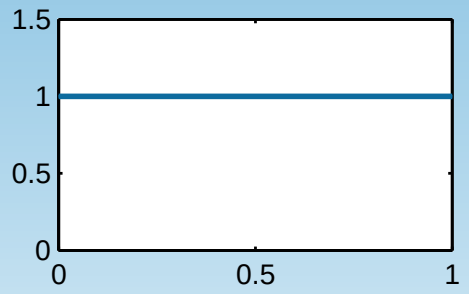
Pravděpodobnosti úspěchů

Určit pravděpodobnost p výskytu jevu v pokusu

- provedeme n pozorování
- pozorujeme k výskytů
  - » problém východu slunce (Bayes/Laplace, 1763/1774)

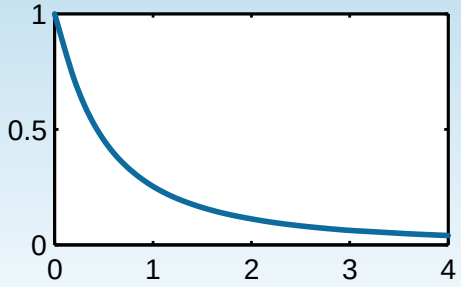
Věrohodnostní funkce

- počet výskytů $S \sim \mathrm{Bi}(n, p)$
- $L(p) = \mathrm{P}(S = k \mid p) = \binom{n}{k} p^k (1-p)^{n-k}$



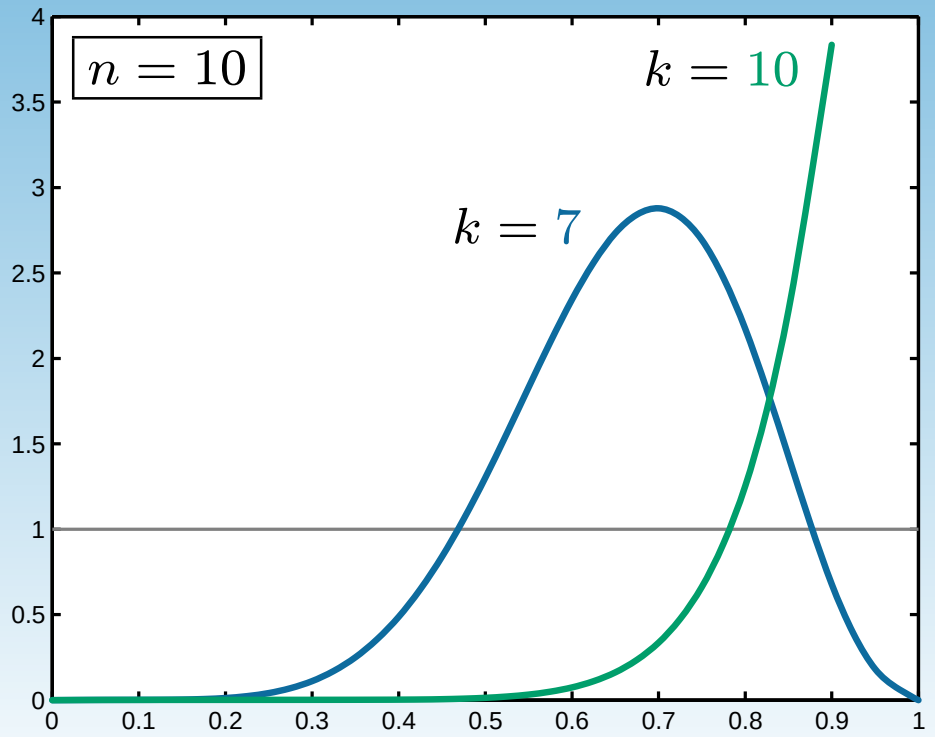
Žádná apriorní informace o p

- neurčitá hustota $\pi(p) = 1$, $p \in (0, 1)$
- ne pro poměr šancí $\gamma = p/(1-p)$,
  kde pak $\pi(\gamma) = (1+\gamma)^{-2}$, $\gamma > 0$



Aposteriorně

- $\pi(p \mid S = k) \propto p^k (1-p)^{n-k} \cdot 1$
- tj. beta rozdělení $\mathrm{B}(k+1, n-k+1)$
- odhad střední hodnotou $\widehat{p} = \mathrm{E}(p \mid S = k) = \dfrac{k+1}{n+2}$

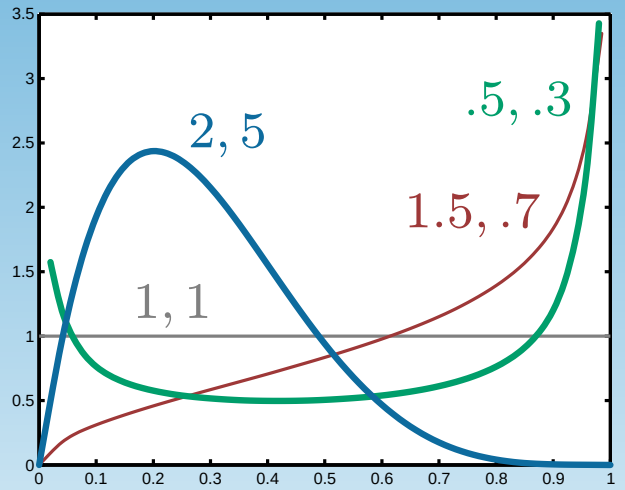


Východ slunce

- v n pozorováních n-krát vyšlo
- pravděpodobnost, že i příště $\widehat{p} = \dfrac{n+1}{n+2}$
- klasicky ML $\widehat{p} = n/n = 1$

Obecněji apriorně $\pi(p) \propto p^{a-1}(1-p)^{b-1}$

- $p \sim \mathrm{B}(a,b)$
- $(p \mid S = k) \sim \mathrm{B}(a+k, b+n-k)$
- odhad $\widehat{p} = \dfrac{a+k}{a+b+n}$



Co když a, b neznáme

- hyperparametry — další apriorní rozdělení pro a, b
- odhadnout — empirické bayesovské metody

Řidiči

Počty ‚nehodových' roků za $n = 10$ let u $N = 20$ řidičů.

- pozorování 0 0 2 0 0 2 2 0 6 4 3 1 1 1 0 0 5 1 1 0
- u j-tého řidiče počet $S_j \sim \mathrm{Bi}(n, p_j)$
  » nejsou všechna p_j stejná
- individuální odhady $\widehat{p}_j = S_j/n$ vs. kolektivní $\overline{S}/n$

Zkombinujeme

- rozložení parametru p mezi řidiči $p_j \sim \mathrm{B}(a, b)$
- aposteriorně $(p_j \mid S_j = k) \sim \mathrm{B}(a + k, b + n - k)$, tedy

$$\widehat{p}_j^{bay} = \frac{a+k}{a+b+n} = \frac{a+b}{a+b+n}\underbrace{\frac{a}{a+b}}_{p_0} + \frac{n}{\underbrace{a+b}_{n_0}+n}\frac{k}{n}$$

- a, b odhadneme z dat všech řidičů

Odhad apriorních parametrů

Využijeme

$$\frac{1}{n}\,\mathrm{E}\,S = \frac{1}{n}\,\mathrm{E}\,\mathrm{E}(S\mid p) = \mathrm{E}\,p = \frac{a}{a+b} = p_0$$

$$n\frac{\mathrm{E}\,\mathrm{var}(S\mid p)}{\mathrm{var}\,\mathrm{E}(S\mid p)} = \frac{\mathrm{E}(p(1-p))}{\mathrm{var}(p)} = a+b = n_0$$

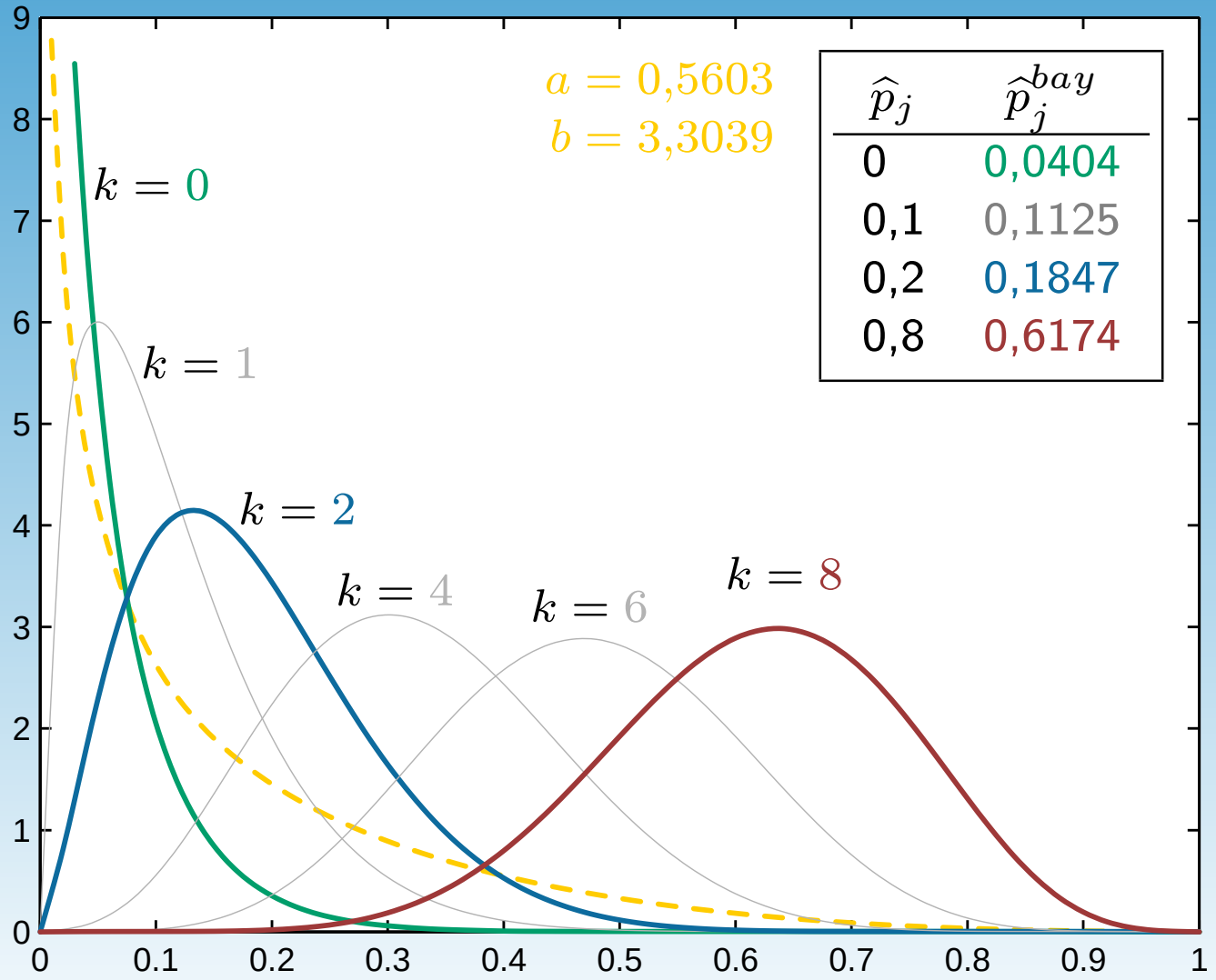
a odhadneme

$$\mathrm{E}\,S \approx \overline{S}, \qquad \mathrm{E}\,\mathrm{var}(S\mid p) \approx N^{-1}\Big(\frac{n}{n-1}\sum \widehat{p}_j(1-\widehat{p}_j)\Big)$$

$$\mathrm{var}\,\mathrm{E}(S\mid p) \approx n^2(s^2_{\widehat{p}_j} - N^{-1}\dots)$$

Pro naše data $n_0 \approx 3{,}8643$, $p_0 \approx 0{,}1450$

$$\widehat{p}_j^{bay} = 0{,}154\cdot 0{,}145 + 0{,}846\cdot \widehat{p}_j$$



$\widehat{p}_j$	$\widehat{p}_j^{bay}$
0	0,0404
0,1	0,1125
0,2	0,1847
0,8	0,6174

Normální IQ

Hledáme IQ u dítěte

- skutečnou hodnotu θ neznáme
- známe X výsledek testu, $X \sim \mathrm{N}(\theta, 10^2)$ při daném θ

Apriorní informace

- z dlouhodobých výzkumů
- rozložení IQ u dětí $\theta \sim \mathrm{N}(100, 225)$

Hustoty

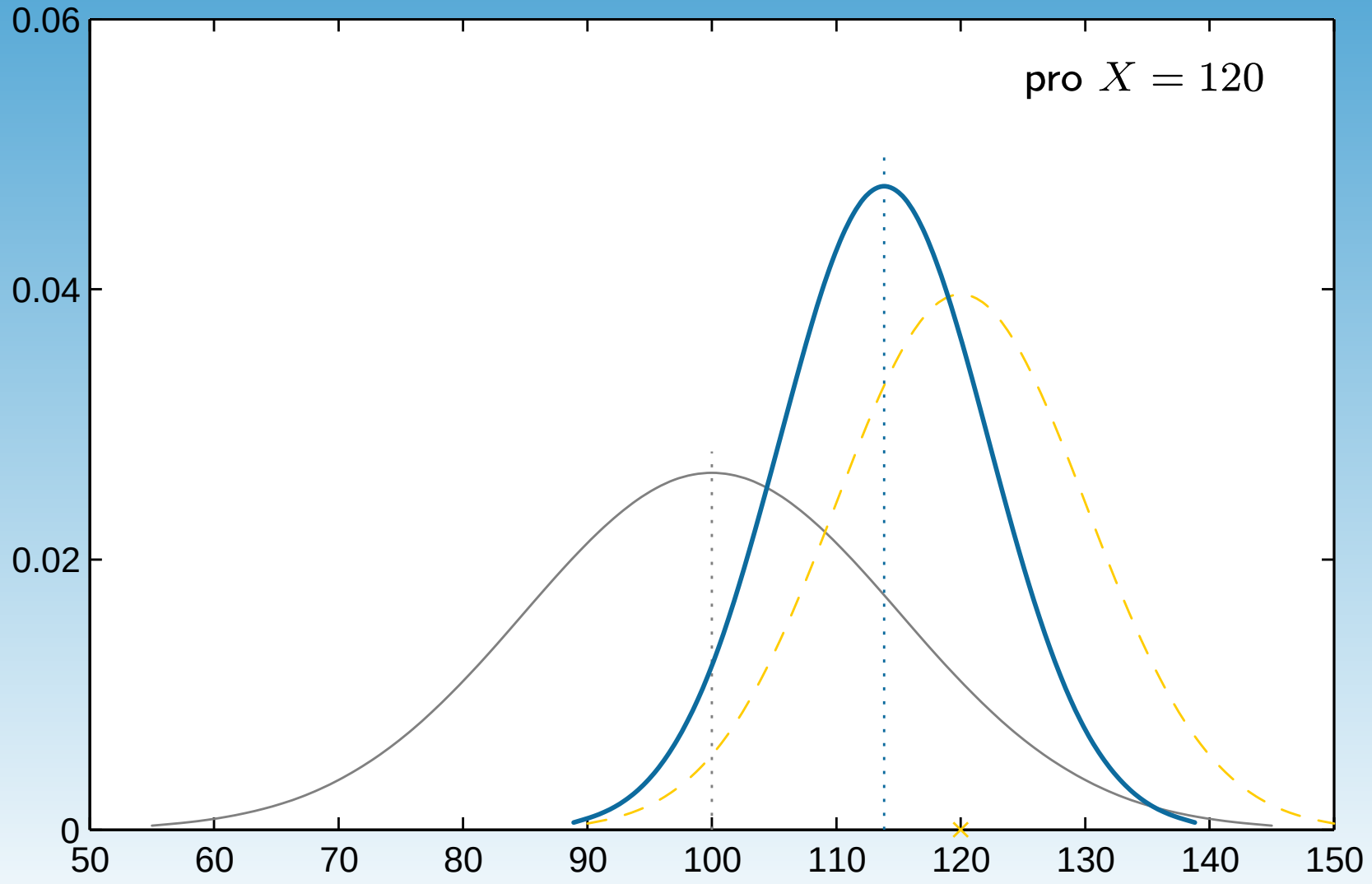
- $f(x \mid \theta) = \dfrac{1}{\sqrt{2\pi \cdot 10^2}} \exp\Big(-\dfrac{(x-\theta)^2}{2 \cdot 10^2}\Big)$
- $\pi(\theta) = \dfrac{1}{\sqrt{2\pi \cdot 225}} \exp\Big(-\dfrac{(\theta-100)^2}{2 \cdot 225}\Big)$

Počítáme

- $$\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) \propto \exp\Big(-\frac{(x-\theta)^2}{2\cdot 10^2} - \frac{(\theta-100)^2}{2\cdot 225}\Big)$$
$$\propto \exp\Big(-\theta^2\big(\frac{1}{10^2}+\frac{1}{225}\big) + 2\theta\big(\frac{x}{10^2}+\frac{100}{225}\big)\Big)$$
$$\sim \mathrm{N}\Big(\frac{x\cdot 225 + 100\cdot 10^2}{225+10^2}, \frac{225\cdot 10^2}{225+10^2}\Big)$$
$$= \mathrm{N}\Big(\frac{9}{13}x + \frac{4}{13}100, \frac{900}{13}\Big)$$

Odhadujeme

- bodově $\widehat{\theta} = \mathrm{E}(\theta \mid X) = \frac{9}{13}X + \frac{4}{13}100 \doteq 0{,}69X + 0{,}31\cdot 100$
- interval $\widehat{\theta} \pm u_{1-\alpha/2}\sqrt{900/13} \doteq \widehat{\theta} \pm 8{,}32\cdot u_{1-\alpha/2}$


pro $X = 120$
0.06
0.04
0.02
0
50
60
70
80
90
100
110
120
130
140
150

Více pozorování

Více pozorování

- $X_1, \ldots, X_n$ náhodný výběr z $\mathrm{N}(\mu, \sigma^2)$
- odhadnout μ, rozptyl σ^2 známe

Oblíbený model

- apriornímu $\mu \sim \mathrm{N}(a, b^2)$
- odpovídá $(\mu \mid X_1, \ldots, X_n) \sim \mathrm{N}(\underbrace{w\overline{X} + (1-w)a}_{\widehat{\mu}}, w\sigma^2/n)$
  kde $w = nb^2/(nb^2 + \sigma^2)$.

Bayesovský intervalový odhad

- $\widehat{\mu} \pm u_{1-\alpha/2}\sqrt{w}\frac{\sigma}{\sqrt{n}}$
- oproti klasickému kratší $(w < 1)$

Porovnání MSE

Střední čtvercová chyba při daném μ

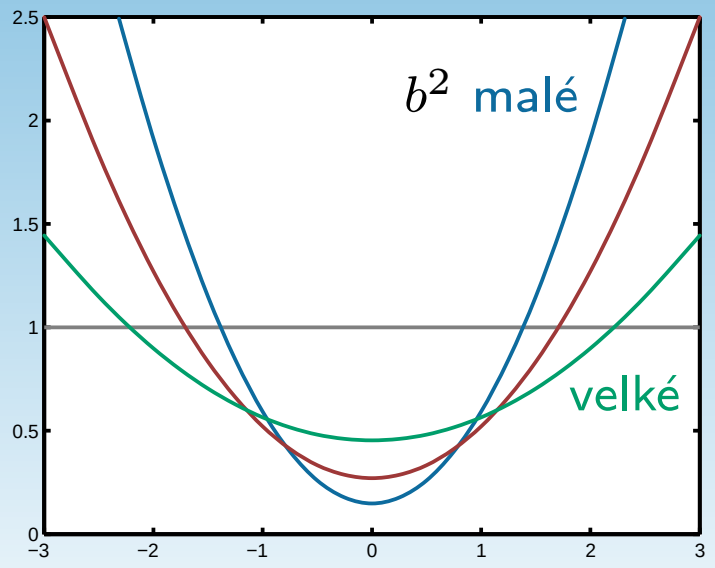
- $\mathrm{MSE}\,\overline{X} = \mathrm{E}(\overline{X} - \mu)^2 = \mathrm{var}\,\overline{X} = \sigma^2/n$
- $\mathrm{MSE}\,\widehat{\mu} = \mathrm{E}(\widehat{\mu} - \mu)^2 = \mathrm{var}\,\widehat{\mu} + (\mathrm{E}\,\widehat{\mu} - \mu)^2$
  $= w^2\,\mathrm{var}\,\overline{X} + (w\,\mathrm{E}\,\overline{X} + (1-w)a - \mu)^2$

Kdy je $\widehat{\mu}$ lepší než $\overline{X}$

- $w^2\sigma^2/n + (1-w)^2(a-\mu)^2 < \sigma^2/n$
- tj. když $|a - \mu| < \sqrt{2b^2 + \sigma^2/n}$

Prakticky

- hodnoty μ omezené
- zvolíme b^2 velké
- ... a $\widehat{\mu}$ je lepší než $\overline{X}$



Regrese

Standardní model n pozorování, k parametrů

- po složkách $y_i = \sum_j x_{ij}\beta_j + e_i$, $i = 1, \ldots, n$
- maticově $y = X\beta + e$

Klasický odhad metodou nejmenších čtverců

- $b = (X'X)^{-1}X'y$

Normální chyby

- $y \sim \mathrm{N}(X\beta, \sigma^2 I)$
- $f(y; \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2}\,\mathrm{e}^{-(y-X\beta)'(y-X\beta)/2\sigma^2}$

Neurčitostní apriorní hustota (nevlastní)

- parametry β a σ^{-2}
- $\pi(\beta, \sigma^{-2}) \propto 1 \cdot (\sigma^{-2})^{-1} = (\sigma^{-2})^{-1}, \quad \beta \in \mathbf{R},\ \sigma^{-2} > 0$

Aposteriorní

- $\pi(\beta, \sigma^{-2} \mid y) = f(y \mid \beta, \sigma^{-2}) \cdot \pi(\beta, \sigma^{-2})$

$$\propto (\sigma^{-2})^{n/2} \exp\big[(y - X\beta)'(y - X\beta)\sigma^{-2}/2\big] \cdot (\sigma^{-2})^{-1}$$

$$= (\sigma^{-2})^{k/2} \exp\big[(\beta - b)' X' X (\beta - b)\sigma^{-2}/2\big]$$

$$\times (\sigma^{-2})^{(n-k)/2-1} \exp\big[\underbrace{(y - Xb)'(y - Xb)}_{S_b}\sigma^{-2}/2\big]$$

. . . normální-gama rozdělení

Při daném y

- $(\beta \mid \sigma^{-2}, y) \sim \mathrm{N}(b, \sigma^2 (X'X)^{-1})$

$$(\sigma^{-2} \mid y) \sim \mathrm{G}(S_b/2, (n-k)/2)$$

- a odhad $\widehat{\beta} = \mathrm{E}(\beta \mid y) = b = (X'X)^{-1} X'y$ (= MNČ)

Marginálně aposteriorně

- $\pi(\beta \mid y) \propto ((\beta - b)' X' X (\beta - b) + S_b)^{-n/2}$
- až na lineární transformaci k-rozměrné t_{n-k}
- složky po znormování t_{n-k} (1-rozm.)

Bayesovské konfidenční oblasti

- pro $\beta_i \ldots$ $b_i \pm t_{1-\alpha/2}(n-k)\sqrt{(X'X)^{-1}_{ii} S_b/(n-k)}$
- pro $\beta \ldots$ $\left\{\beta; \frac{(\beta-b)' X' X (\beta-b)/k}{S_b/(n-k)} \leqq F_{1-\alpha}(k, n-k)\right\}$

(jako klasické intervaly spolehlivosti)

Obecněji

Apriorní normální-gama

- $(\beta \mid \sigma^{-2}) \sim \mathrm{N}(a_0, \sigma^2 M_0^{-1})$
- $\sigma^{-2} \sim \mathrm{G}(S_0, n_0)$

Aposteriorní normální-gama s parametry

$$a_1 = M_1^{-1}(M_0 a_0 + X'Xb) = \widehat{\beta},$$

$$M_1 = M_0 + X'X$$

$$n_1 = n_0 + n,$$

$$S_1 = S_0 + S_b + (b - a_0)'(M_0^{-1} + (X'X)^{-1})^{-1}(b - a_0)$$

Předchozí bylo formálně s

- $M_0 = 0$ (resp. $M_0^{-1} = \infty$), $S_0 = 0$, $n_0 = -k$

Jamesův-Steinův odhad

Regresní model, $k > 2$

- pro jednoduchost σ^2 známé, odhadujeme β
- apriorně $\beta \sim \mathrm{N}(0, \tau^2 (X'X)^{-1})$
- bayesovský odhad $\widehat{\beta} = \mathrm{E}(\beta \mid y) = \Big(1 - \dfrac{\sigma^2}{\sigma^2 + \tau^2}\Big) b$

Využijeme

- marginálně $z = (X'X)^{-1/2} X'y \sim \mathrm{N}(0, (\sigma^2 + \tau^2) I_k)$
- tedy $\mathrm{E}(\sigma^2 + \tau^2)/(z'z) = \mathrm{E}(1/\chi^2_k) = 1/(k-2)$
- odhadneme $\frac{1}{\sigma^2+\tau^2} \approx \frac{k-2}{z'z}$

Dosadíme ...

- odhad $\widehat{\beta}_{JS} = \Big(1 - \dfrac{(k-2)\sigma^2}{b'X'Xb}\Big) b$
- pro každé β je $\mathrm{MSE}\, \widehat{\beta}_{JS} < \mathrm{MSE}\, b$, kde MSE s vahou $X'X$

Neparametrické bayesovské odhady

Pozorování $X_1, \ldots, X_n$

- parametr celá distribuční funkce $F(x)$, $x \in \mathbf{R}$
- neparametrický odhad $F_e(x) = \#\{i; X_i \leqq x\}/n$

Bayesovsky

- hodnoty F jsou náhodné
- apriorní rozdělení pro nekonečněrozměrný parametr F

Dirichletův proces $F \sim \mathcal{D}(n_0 F_0)$

- pro každé $-\infty = t_0 < \cdots < t_k = \infty$

$$(U_1, \ldots, U_k) \sim \mathrm{D}(a_1, \ldots, a_k) \propto \prod u_i^{a_i - 1}, \quad \sum u_i = 1,$$

  kde $U_i = F(t_i) - F(t_{i-1})$, $a_i = n_0 F(t_i) - n_0 F(t_{i-1})$
- složky $U_i \sim \mathrm{B}(a_i, n_0 - a_i)$, $\mathrm{E}\, F(t) = F_0(t)$

Aposteriorně

- $(F \mid X_1, \ldots, X_n) \sim \mathcal{D}(n_0 F_0 + n F_e)$
- bayesovský odhad $\widehat{F}(t) = \dfrac{n_0 F_0(t) + n F_e(t)}{n_0 + n}$

Analýza dat o přežití

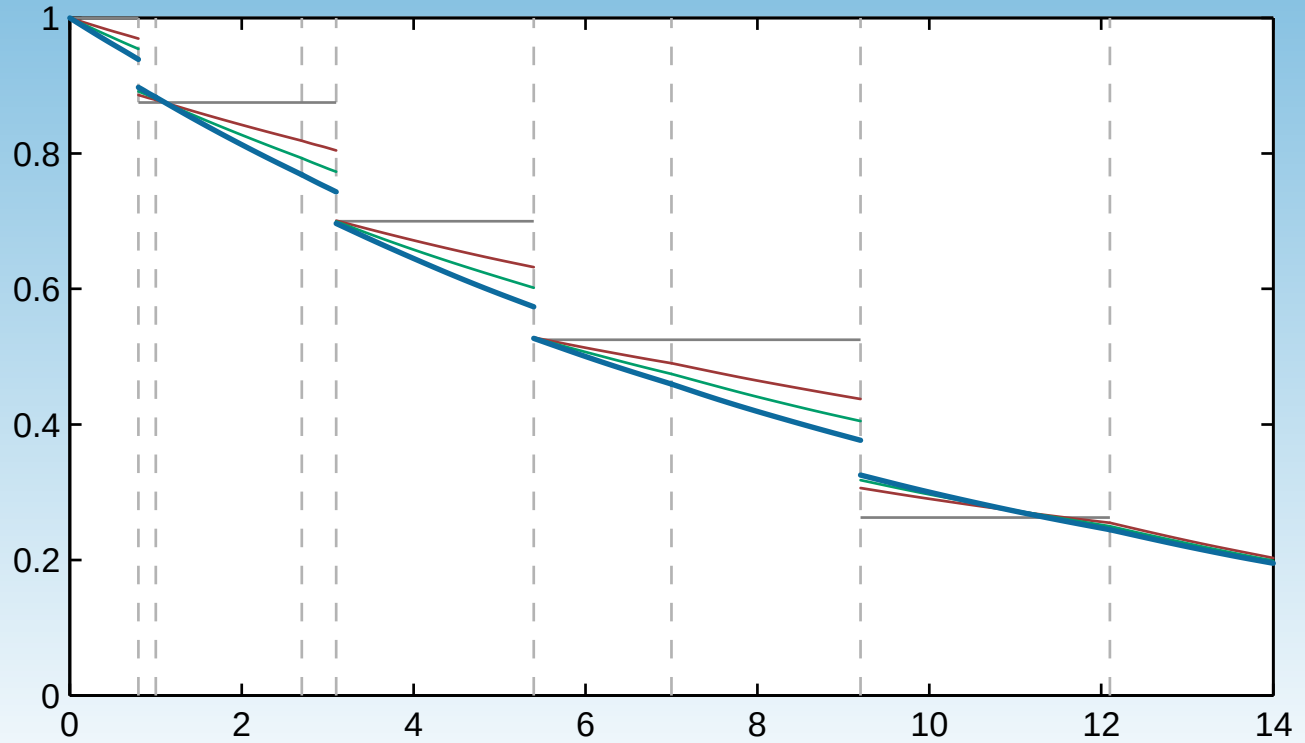
- některá X_i cenzorována
- pozorujeme $Z_i = \min(X_i, Y_i)$ a $\delta_i = I_{[X_i \leqq Y_i]}$
- odhad

$$1 - \widehat{F}(t) = \frac{n_0 S_0(t) + N_t}{n_0 + n} \prod_s \frac{n_0 S_0(s-) + N_{s-} - u(s)}{n_0 S_0(s-) + N_s},$$

kde součin je přes $\{s, \exists_i\, Z_i \leqq t, \delta_i = 0\}$,
$N_s = \#\{i; Z_i \leqq s\}$ je počet pozorování nepřekračujících s,
$u(s) =$ počet necenzorovaných pozorování v okamžiku s

Příklad

- Miniaturní data (Kaplan a Meier): necenzorovaná 0,8, 3,1, 5,4, 9,2, cenzorovaná 1,0, 2,7, 7,0 12,1.
- apriorní představa $1 - F_0(x) = \mathrm{e}^{-0{,}12x}$, $n_0 = 4, 8, 16$



BAYESOVSKÉ ODHADY
V NĚKTERÝCH MODELECH

Michal Friesl

Katedra matematiky
Fakulta aplikovaných věd
Západočeská univerzita v Plzni

•http://home.zcu.cz/~friesl/Archiv/Bohd04.html