Příklad:

$Y_1, \dots, Y_n$ nezávislé $\quad Y_i \sim N(x_i^T\beta, \sigma^2)$

$x_1, \dots, x_n \in \mathbb{R}^k$ konstanty

- Pro volbu $x_i \equiv 1$ dostaneme

$Y_1, \dots, Y_n \overset{iid}{\sim} N(\beta, \sigma^2)$

Potřebné veličiny pro ML odhad $\theta = (\beta^T, \sigma^2)^T$.

UŽITEČNÉ OZNAČENÍ: $\quad Y \sim N_n(X\beta, \sigma^2 I) \quad , X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$

PŘEDPOKLAD: $\operatorname{rank}(X) = k$, $n > k$

$$SS(\beta) = \sum_i (Y_i - x_i^T\beta)^2 = (Y - X\beta)^T(Y - X\beta)$$

$$L(\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} SS(\beta)\right)$$

$$l(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2} SS(\beta)$$

DÁLE POZOR $\sigma^2 > 0$ je parametrem, derivujeme podle „σ^2".

$\cancel{U(\theta) = \frac{\partial l}{\partial\beta}}$ $\quad$ PŘEDPOČÍTEJME:

$$\frac{\partial SS}{\partial\beta} = -2X^T(Y - X\beta) = -2(X^TY - X^TX\beta)$$

$$\frac{\partial l}{\partial\beta} = -\frac{1}{2\sigma^2}\frac{\partial SS}{\partial\beta} = -\frac{1}{2\sigma^2}(-2)(X^TY - X^TX\beta) = \frac{1}{\sigma^2}(X^TY - X^TX\beta)$$

$$\frac{\partial l}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} SS(\beta)$$

$$U(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}(X^TY - X^TX\beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} SS(\beta) \end{pmatrix}$$

$$\Rightarrow \quad \hat\beta_{ML} = (X^TX)^{-1}X^TY$$

$$\hat\sigma^2_{ML} = \frac{SS(\hat\beta_{ML})}{n}$$

VÝPOČET $I(\theta)$

$$-\frac{\partial^2 \ell}{\partial\beta\partial\beta^T} = -\left(-\frac{1}{\sigma^2}X^TX\right) = \frac{1}{\sigma^2}X^TX \qquad = I_{\beta,\beta}$$

$$-\frac{\partial^2 \ell}{\partial\sigma^2\partial\sigma^2} = -\left(-\frac{1}{\sigma^6}SS(\beta) + \frac{n}{2\sigma^4}\right) = \frac{1}{\sigma^6}SS(\beta) - \frac{n}{2\sigma^4} = I_{\sigma^2,\sigma^2}$$

$$-\frac{\partial^2 \ell}{\partial\beta\partial\sigma^2} = \cancel{-\left(-\frac{1}{\sigma^4}(Y^TX - \beta\right.} \quad \frac{1}{\sigma^4}(X^TY - X^TX\beta) = I_{\beta,\sigma^2}$$

VÝPOČET $J(\theta)$

$$J_{\beta,\beta} = EI_{\beta,\beta} = E\left(\frac{1}{\sigma^2}X^TX\right) = \frac{1}{\sigma^2}X^TX$$

$$J_{\sigma^2,\sigma^2} = E\left(\frac{1}{\sigma^6}SS(\beta) - \frac{n}{2\sigma^4}\right) = \frac{1}{\sigma^6}\underbrace{ESS(\beta)}_{\substack{E\left(\sum_i (Y_i - x_i^T\beta)^2\right) \\ \underset{\text{nezáv.}}{=} n\sigma^2}} - \frac{n}{2\sigma^4} = \frac{n}{2\sigma^4}$$

$$J_{\beta,\sigma^2} = E\left(\frac{1}{\sigma^4}(X^TY - X^TX\beta)\right) =$$

$$= \frac{1}{\sigma^4}(\underbrace{EX^TY}_{X^TX\beta} - X^TX\beta) = 0$$

$$J(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}X^TX & 0 \\ 0^T & \frac{n}{2\sigma^4} \end{pmatrix}$$

Přednáška konec cvičení č. 2

SITUACE: $Y_1, \dots, Y_n$ nezávislé, $Y_i \sim \mathcal{N}(x_i^T\beta, \sigma^2)$,

$x_1, \dots, x_n \in \mathbb{R}^k$ konstanty

$\theta = (\beta^T; \sigma^2)^T$

OZNAČ: $\mathbb{X} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$ PŘEDPOKLAD: $\operatorname{rank}(\mathbb{X}) = k < n$

$$SS(\beta) := \sum_{i=1}^{n} (Y_i - x_i^T\beta)^2 = (Y - \mathbb{X}\beta)^T (Y - \mathbb{X}\beta)$$

$$L(\theta) = p(y/\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} SS(\beta)\right)$$

$$l(\theta) = \log L(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2} SS(\beta)$$

$$U(\theta) = \frac{\partial l}{\partial \theta}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}(X^TY - X^TX\beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} SS(\beta) \end{pmatrix}$$

$$I(\theta) = -\frac{\partial^2 l}{\partial\theta\,\partial\theta^T}(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}X^TX & \frac{1}{\sigma^4}(X^TY - X^TX\beta) \\ * & \frac{1}{\sigma^6}SS(\beta) - \frac{n}{2\sigma^4} \end{pmatrix}$$

$$J(\theta) = \mathbb{E}_\theta I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}X^TX & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

SPECIÁLNÍ PŘÍPAD: $x_i = 1 \ \forall i$, tj. $Y_1, \dots, Y_n \overset{iid}{\sim} \mathcal{N}(\beta, \sigma^2)$

$$X^TX = n, \quad J(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Jeffreyova apriorní hustota

(a) σ^2 známé

$$J(\theta) = J(\beta) = \frac{1}{\sigma^2} X^T X$$

$$p(\beta) \propto \sqrt{\det J(\theta)} = \sqrt{\det\left(\frac{1}{\sigma^2} X^T X\right)} \propto 1, \quad \beta \in \mathbb{R}^k$$

nevlastní

(b) β známé

$$J(\theta) = J(\sigma^2) = \frac{n}{2\sigma^4}$$

$$|J(\sigma^2)|^{1/2} = \sqrt{\frac{n}{2}} \frac{1}{\sigma^2} \propto \frac{1}{\sigma^2}$$ parametr $\theta = \sigma^2$

tj: $p(\theta) \propto \frac{1}{\theta}, \quad \theta > 0$

(opět nevlastní)

nechť $\tau := \frac{1}{\theta} = \frac{1}{\sigma^2}$ (inverzní rozptyl)

$\theta = \frac{1}{\tau}$

$\theta' = -\frac{1}{\tau^2}$ $\quad p(\tau) \propto \tau \cdot \frac{1}{\tau^2} = \frac{1}{\tau}$

$= Ga(0,0)$

$\eta := \log \sigma$ (log(směrodatná odchylka))

$e^{2\eta} = \sigma^2 \; (= \theta)$

$\theta' = 2 \cdot e^{2\eta}$ $\quad p(\eta) \propto \frac{1}{e^{2\eta}} \cdot 2 \cdot e^{2\eta} = 2 \propto 1$

tj. Jeffreys pro σ^2 odpovídá $\propto 1$ pro $\log(\sigma)$

Pozn. Pokud specifikujeme sdružené apriorní

$$p(\beta,\sigma^2) \equiv p(\beta,\tau) \propto \frac{1}{\tau} \qquad , \beta \in \mathbb{R}^k, \tau > 0$$

$$\equiv p(\beta,\eta) \propto 1 \qquad \beta \in \mathbb{R}^k, \eta \in \mathbb{R}$$

odpovídá to Jeffreysovi pro β a $\sigma^2/\tau/\eta$ a jejich apriorní nezávislosti, tj:

$$p(\beta,\eta) = p(\beta)\cdot p(\eta)$$

(c) β, σ^2 neznámé

- pouze speciální případ $x_i = 1 \ \forall i$

$\beta = \beta \in \mathbb{R}$

$$J(\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \qquad |J(\theta)| = \frac{n^2}{2\sigma^6}$$

$$p(\beta,\sigma^2) \propto \frac{1}{\sigma^3}$$

$$= p(\beta)\cdot p(\sigma^2) \text{ , kde } p(\beta) \propto 1$$

p

$$p(\sigma^2) \propto \frac{1}{\sigma^3} \left(= \frac{1}{\theta_2^{3/2}}\right)$$

tj. apriorní nezávislost

pro inverzní rozptyl:

$$\tau = \frac{1}{\sigma^2}: \quad p(\tau) \propto \tau^{3/2}\cdot\frac{1}{\tau^2} = \frac{1}{\tau^{1/2}} = Ga\left(\frac{1}{2}, 0\right)$$

(nevlastní)

pro log(směr. odchylky):

$$\eta = \log(\sigma): \quad p(\eta) \propto \frac{1}{e^{3\eta}}\cdot 2\cdot e^{2\eta} \propto \frac{1}{e^{\eta}}$$

Konjugovaný systém apriorních rozdělení (pro lineární model)

$\equiv$ systém apriorních rozdělení
$\{p(\theta;\xi) : \xi \in \boxed{-}\}$ příslušný „modelu"
$p(y/\theta)$ tak, že

$p(\theta/y) \in$ systém

hyperparametr – jeho konkrétní volba zvolí jedno konkrétní apriorní rozdělení

Tvorba konjugovaného systému:

postačující statistika $t(Y)$ (k „modelu")
a faktorizační kritérium:

$t(Y)$ je postačující statistika (tj. $Y/t(Y)$ nezávisí na θ)

$$\Longleftrightarrow \quad p(y/\theta) = g_1(y)\, g_2(t(y);\theta) \quad (\text{pro s.v. } y)$$

hustota z konjugovaného systému:

$$p(\theta;\xi) \propto g_2(\xi;\theta)$$

postač. statistika

$\rightarrow$ hyperparametr

(viz dále)

~~Konj~~ Postačující statistika pro lineární model (normální)

$$p(y|\theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} SS(\beta)\right)$$

$$SS(\beta) = (y - X\beta)^T (y - X\beta)$$

RADA: Označ $b := (X^T X)^{-1} (X^T Y)$ (LSE param. β)

$$SS(\beta) = (Y - X\beta)^T (Y - X\beta) =$$

$$= (\underbrace{Y - Xb} + \underbrace{Xb - X\beta})^T (\underbrace{Y - Xb} + \underbrace{Xb - X\beta})$$

$$= (Y - Xb)^T (Y - Xb) + (Xb - X\beta)^T (Xb - X\beta)$$

$$+ 2 \underbrace{(Y - Xb)^T (Xb - X\beta)}_{= 0 \text{ (použitím } X^T X b = X^T Y)}$$

$$= \underbrace{(Y - Xb)^T (Y - Xb)}_{=: SSe} + (\beta - b)^T X^T X (\beta - b)$$

tedy $p(y|\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} SSe - \frac{1}{2\sigma^2} (\beta - b)^T X^T X (\beta - b)\right)$

$\rightarrow$ postačující statistiky: SSe (> 0)
b $(\in \mathbb{R}^k)$

Konjugovaná apriorní hustota

postač. statistiky → hyperparametry

$$p(\theta;\xi) \propto (2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\cdot S\right)$$

$$\cdot \exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_0)^T X^T X(\beta-\beta_0)\right)$$

$$\xi \equiv (\beta_0, S) \qquad \beta_0 \in \mathbb{R}^k, \quad S > 0$$

---

normující konstanta

$$\int_{\mathbb{R}^k \times (0,\infty)} (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{S}{2\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_0)^T X^T X(\beta-\beta_0)\right) d(\beta,\sigma^2)$$

→ budeme počítat? ---

→ nejde poznat, co je to za rozdělení?

---

už vynechám

$$p(\beta,\sigma^2;\xi) \propto (\sigma^2)^{-n/2}\exp\left(-\frac{S}{2\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_0)^T X^T X\cdot(\beta-\beta_0)\right)$$

lze faktorizovat

$$p(\beta|\sigma^2)\cdot p(\sigma^2), \text{ kde}$$

$$p(\sigma^2) \propto (\sigma^2)^{-\frac{n-k}{2}}\exp\left(-\frac{S}{2\sigma^2}\right)$$

$$p(\beta|\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}(\beta-\beta_0)^T X^T X(\beta-\beta_0)\right)(\sigma^2)^{-\frac{k}{2}}$$

POZNÁVÁME: $\beta|\sigma^2 \sim N(\beta_0, \sigma^2(X^TX)^{-1})$

$\sigma^2 \sim$ INV-GAMMA

$$\tau := \frac{1}{\sigma^2} \qquad p(\tau) \propto \tau^{\frac{n-k}{2}} \exp\left(-\frac{S}{2}\tau\right)\tau^{-2}$$

$$\sigma^2 = \frac{1}{\tau}$$

$$(\sigma^2)' = -\frac{1}{\tau^2} \qquad = \tau^{\left(\frac{n-k}{2}-1\right)-1} \exp\left(-\frac{S}{2}\tau\right)$$

$$\text{tj. } \tau \sim Ga\left(\frac{n-k}{2}-1, \frac{S}{2}\right)$$

Tedy konjugovaný systém pro normální lin model

$$p(\beta,\tau) = p(\beta|\tau)\,p(\tau) \qquad \tau \sim Ga\left(\frac{n-k}{2}-1, \frac{S}{2}\right)$$

$$\beta|\tau \sim N(\beta_0, \tau^{-1}(X^TX)^{-1})$$

$\beta_0 \in \mathbb{R}^k$, $S>0$: hyperparametry

Vyjde stále konjugovaný:

$\tau \sim Ga(c_0, d_0)$, $c_0, d_0 > 0$

$\beta|\tau \sim N(\beta_0, \tau^{-1}\Sigma_0)$, $\beta_0 \in \mathbb{R}^k$, $\Sigma_0 > 0$ } hyperparam.

SPECIÁLNĚ: normální iid případ

$X^TX = n$, β = střední hodnota výběru

původní konjug. systém: $\tau \sim Ga\left(\frac{n-1}{2}-1, \frac{S}{2}\right)$

$$\beta|\tau \sim N\Big(\beta_0, \underbrace{\tau^{-1}\frac{1}{n}}_{\frac{\sigma^2}{n}}\Big)$$

→ Normální-gama rozdělení

$N\text{-}Ga(\beta_0, \Sigma_0, c_0, d_0)$

V praxi se spíše používá tzv. <u>semikonjugovaný</u> systém

$p(\beta, \tau) = p(\beta) \cdot p(\tau)$ (apriorní nezávislost β a τ)

$\tau \sim Ga(c_0, d_0)$, $\beta \sim N(\beta_0, \Sigma_0)$

$\rightarrow$ vyjde (spočítáme)

$p(\beta, \tau | y) \equiv N\text{-}Ga(\cdot, \cdot, \cdot, \cdot)$

Princip neurčitosti (motivovaný Jeffreysem)

$p(\tau) \propto \frac{1}{\tau} \equiv Ga(0, 0)$

$p(\beta) \propto 1 \equiv N(0, 0^{-1})$