Hierarchické modely 0

1. Hierarchické apriorní rozdělení 1

2

MODEL PRO DATA = věrohodnost 3

$L(\psi) = p(y|\psi)$

APRIORNÍ ROZDĚLENÍ: $p(\psi)$

– typicky voleno z nějaké třídy, např. konjugovaných aprior. rozdělení, tj:

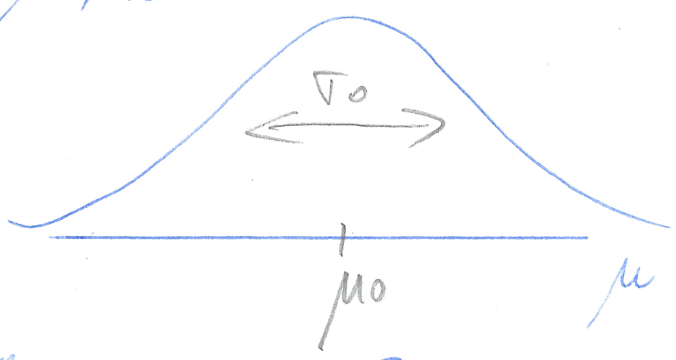
$p(\psi) = p(\psi; \xi)$

ξ hyperparametr, potřeba zvolit

Příklad: MODEL & DATA: $Y_1, \dots, Y_n$ iid $N(\mu, \underline{\sigma^2})$ σ^2 znám

$\psi = \mu$

APRIORNÍ: $\mu \sim N(\mu_0, \sigma_0^2)$



μ_0 má být $0, 1{,}8, \pi, \dots$?

σ_0 má být $\underbrace{0{,}1,\ 1{,}3,\ 10,\ 10^8}$?

další rozdělení na tyto volby

→ hierarchicky specifikované apriorní

Def 3.1 Bayesovský model s hierarchicky specifikovaným apriorním rozdělením

- jednodušeji

MODEL: $L(\psi) = p(y|\psi)$

APRIORNÍ PRO ψ: $p(\psi;\xi)$

$\rightarrow p(\psi|\xi)$ & $p(\xi)$

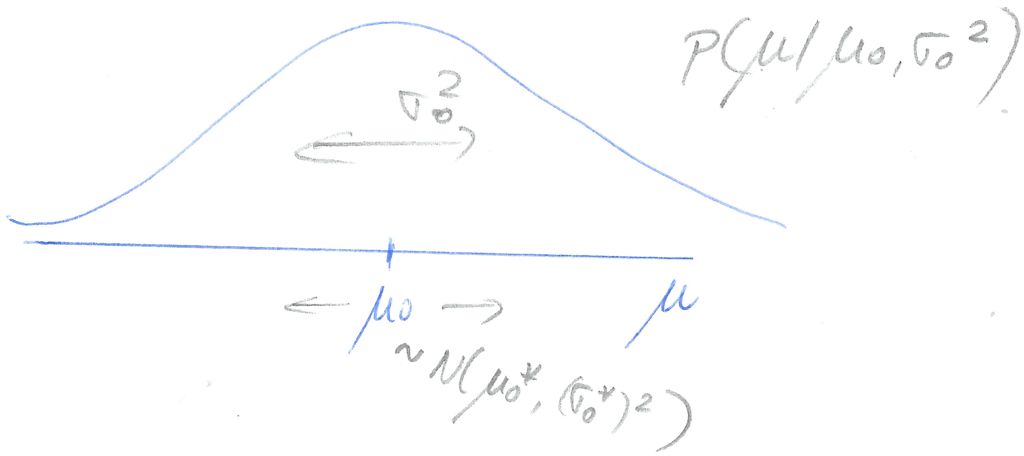
Potom
$$p(\psi) = \int p(\psi,\xi)\,d\xi = \int p(\psi|\xi)\,p(\xi)\,d\xi$$

≡ rozklad variability $p(\psi)$ na

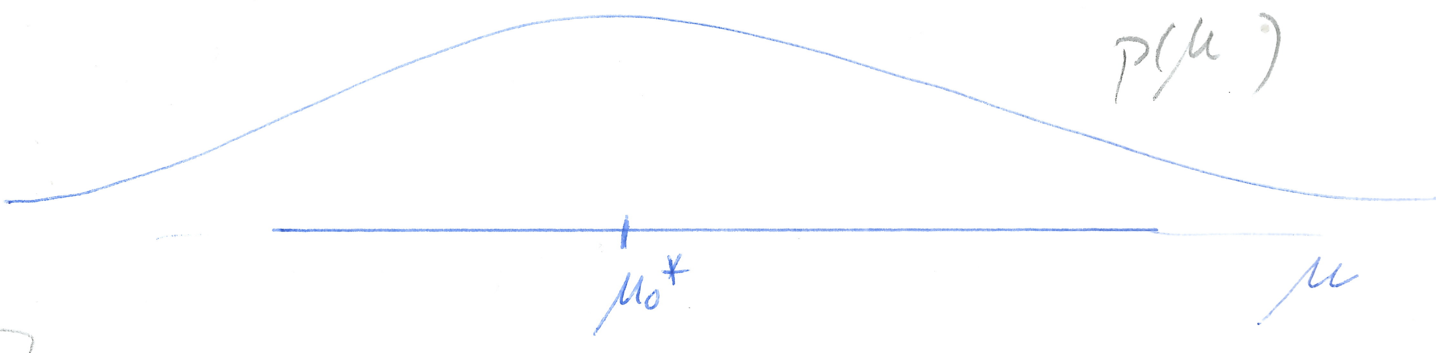
(i) variabilitu $p(\psi|\xi)$

(ii) variabilitu $p(\xi)$

Příklad



→



Def 3.1

apriorní rozdělení pro ψ:

$$\psi \to \xi_1 \to \xi_2 \to \cdots \to \xi_m$$

sdružené:

$$p(\psi, \xi_1, \ldots, \xi_m) = p$$

$$= p_0(\psi / \xi_1, \cancel{\xi_2, \ldots, \xi_m}) \cdot p_1(\xi_1 / \xi_2, \cancel{\xi_3, \ldots, \xi_m})$$

$$\ldots p_{m-1}(\xi_{m-1} / \xi_m) \, p_m(\xi_m)$$

× pořaduje definice

příslušné marginální je

$$p(\psi) = \int_{Z_1 \times \cdots \times Z_m} p_0(\psi / \xi_1) \cdot p_1(\xi_1 / \xi_2) \cdots p_{m-1}(\xi_{m-1} / \xi_m) \cdot p_m(\xi_m) \, d\xi_m \ldots d\xi_1$$

ξ_i = hyperparametr i-té úrovně

Příslušné aposteriorní rozdělení:

$$p(\psi / y) \propto p(y / \psi) \cdot p(\psi) =$$

$$= p(y / \psi) \int p_0(\psi / \xi_1) \cdot p_1(\xi_1 / \xi_2) \ldots p_m(\xi_m) \, d\xi_m \ldots d\xi_1$$

$\underline{\text{Data}} \equiv$ nejvyšší (nejnižší) hierarchická úroveň

$$y \underset{\text{věrohodnost}}{\rightarrow} \psi \rightarrow \xi_1 \rightarrow \xi_2 \rightarrow \cdots \rightarrow \xi_m$$

$$p(y|\psi, \xi_1, \dots \xi_m) = p(y|\psi)$$

předpoklad (hierarchický model)

$$p(\psi, \xi_1, \dots, \xi_m | y) = ?$$

$$\propto p(y|\psi, \cancel{\xi_1, \dots, \xi_m}) \cdot \underbrace{p(\psi, \xi_1, \dots, \xi_m)}_{\text{hierar. prior}} =$$

$$= p(y|\psi) \cdot p_0(\psi|\xi_1) \cdot p_1(\xi_1|\xi_2) \dots p_m(\xi_m)$$

Potom $p(\psi|y) = \int p(\psi, \xi_1, \dots, \xi_m | y) \, d\xi_m \dots d\xi_1$

$$= \int p(y|\psi) \, p_0(\psi|\xi_1) \, p_1(\xi_1|\xi_2) \dots$$

$$\dots p_m(\xi_m) \, d\xi_m \dots d\xi_1$$

= minulá strana

$\underline{\text{Zapamatuj}}$: V hierarchickým apriorním rozdělením je $p(\psi|y)$ $\underline{\text{marginálním}}$ rozdělením z

$$\underbrace{p(\psi, \xi_1, \dots, \xi_m | y)}_{\text{aposteriorní}}$$

aposteriorní, kde $\Theta = (\psi, \xi_1, \dots, \xi_m)$

jsou parametry modelu

Zejména pro Monte Carlo (Markov chain) může být výhodné pracovat primárně s ...

Zejména pro Monte Carlo (Markov chain) může být (většinou je) výhodné zpracovat primárně s

$$p(\psi, \xi_1, \dots, \xi_m | y) \propto$$

$$\propto p(y|\psi)\, p_0(\psi|\xi_1)\, p_1(\xi_1|\xi_2) \dots p_m(\xi_m)$$

a integrál $p(\psi|y) = \int \boxed{}\, d\xi_m \dots d\xi_1$

počítat pouze když to je skutečně potřeba (při MC málokdy ...)

→ potřebuji (náhodný) výběr z $p(\psi|y)$

≡ příslušné složky (náhodného) výběru z $p(\psi, \xi_1, \dots, \xi_m | y)$

Příklad: Lineární model

$Y \mid \psi \sim N(X\beta, \sigma^2 I_n)$, $X_{n\times k}$: matice konstant

$\psi = (\beta, \tau)$, $\tau = \sigma^{-2}$

Konjugované apriorní rozdělení

$$p(\beta,\tau \mid \xi) = p(\beta \mid \tau, \xi)\, p(\tau \mid \xi)$$

$$\beta \mid \tau, \xi \sim N(\beta_0, \tau^{-1}\Sigma_0)$$
$$\tau \mid \xi \sim Ga(c_0, d_0)$$

$\xi = \beta_0, \Sigma_0, c_0, d_0$

→ ne vše musíme hned znáhodňovat

- např. pouze d_0 náhodný hyperparametr

  β_0, Σ_0, c_0: pevné hyperparametry

6

hierarchický model:

$$y \to (\beta, \tau) \to d_0$$

$p(y \mid \beta, \tau)$ = věrohodnost $N(X\beta, \tau^{-1} I_n)$

← konjugi systém

$p(\beta, \tau \mid d_0) = p(\beta \mid \tau)\, p(\tau \mid d_0)$

$p(\beta \mid \tau) \sim N(\beta_0, \tau^{-1}\Sigma_0)$, $p(\tau \mid d_0) \sim Ga(c_0, d_0)$

$p(d_0) = ?$

např. $p(d_0) \sim Ga(g_0, h_0)$

pevné / náhodné?

2. Hierarchicky specifikovaná věrohodnost

7

8

NTP TER84073 pokus na myších 9

Y_{ij} = hmotnost zárodku j u myši i

$\mu = \mathbb{E}Y_{ij}$ = primární parametr

obr. 10–11

MODEL 1: $Y_{ij} \overset{iid}{\sim} N(\underbrace{\mu, \sigma^2}_{\psi})$ 12

→ vhodné?

MODEL 2 (realističtější)? 13

$$Y_{ij} | b_i \underset{nezáv.}{\sim} N(b_i, \sigma^2)$$

$$\boxed{\psi \equiv \mu, d^2, \sigma^2}$$

$$b_i \underset{iid}{\sim} N(\mu, d^2)$$

za podmínky ψ:

$$\mathbb{E}Y_{ij} = \mathbb{E}(\mathbb{E}(Y_{ij}|b_i)) = \mathbb{E}(b_i) = \mu$$

Model implikuje (kromě jiného) 14

$$\mathrm{cov}(Y_{ij}, Y_{ik}) = \dots = d^2$$

(SDC)

$$\mathrm{var}\, Y_{ij} = \sigma^2 + d^2$$

$$Y_i = (Y_{i1}, \dots, Y_{in_i})^T \sim N\left(\mu \cdot 1, \begin{pmatrix} \sigma^2+d^2 & & d^2 \\ & \ddots & \\ d^2 & & \sigma^2+d^2 \end{pmatrix}\right)$$

compound symmetry

$\Rightarrow$: $p(y_i/\psi) \equiv \mathcal{N}\left(\mu \cdot \mathbb{1}, \begin{pmatrix} \sigma^2+d^2 & \cdots & d^2 \\ d^2 & \cdots & \sigma^2+d^2 \end{pmatrix}\right)$

tohle jsme ale nenapsali přímo

Model byl specifikován takto:

$$p(y_i/\psi) = \int p(y_i, b_i/\psi)\, db_i =$$

$$= \int \underbrace{p(y_i/b_i;\psi)}_{\sim \prod_{j=1}^{n_i} \underbrace{p(y_{ij}/b_i;\psi)}_{\sim \mathcal{N}(b_i;\psi)}} \underbrace{p(b_i/\psi)}_{\sim \mathcal{N}(\mu, d^2)}\, db_i$$

$b_i \equiv$ „skrytá" data

$\equiv$ data vyšší/nižší hierarchické úrovně

– v dalším značena pomocí ξ

8/

OBECNĚ, hierarchicky specifikovaná věrohodnost

$p(y|\psi)$ vznikne

z $\quad p(y, \xi_1, \dots, \xi_m|\psi) \overset{\text{předp.}~\cancel{\text{hierar}}}{=}$

$$= p(y|\xi_1, \dots, \xi_m, \psi) \cdot p(\xi_1|\xi_2, \dots, \xi_m, \psi) \cdots$$

$$\cdots p(\xi_m|\psi)$$

$\overset{\text{předp. hierarchie}}{=}$

$$p(y|\xi_1, \psi)\, p(\xi_1|\xi_2, \psi) \cdots$$

$$\cdots p(\xi_{m-1}|\xi_m, \psi)\, p(\xi_m|\psi)$$

$$p(y|\psi) = \int p(y|\xi_1, \psi)\, p(\xi_1|\xi_2, \psi) \cdots$$

$$\cdots p(\xi_{m-1}|\xi_m, \psi)\, p(\xi_m|\psi)$$

$$d\xi_m\, d\xi_{m-1} \dots d\xi_2\, d\xi_1$$

DATA GENERUJÍCÍ MECHANIZMUS

$$y \leftarrow \xi_1 \leftarrow \xi_2 \leftarrow \dots \leftarrow \xi_m$$

skrytá data 1. úrovně — skrytá data 2. úrovně — skrytá data m-té úrovně

vše závisí na

neznámém parametru ψ

$p(y|\psi) =: L_F(\psi)$ = frekventistická věrohodnost

(se kterou je nutno pracovat při ML atd.)

Pro Bayesovskou inferenci o ψ:

$p(\psi)$ = apriorní rozdělení (také hierarchicky specifikované :-) ZATÍM NE)

(*)
$$p(\psi|y) \propto p(y|\psi)\, p(\psi)$$
$$= \left\{ \int p(y|\xi_1,\psi)\, p(\xi_1|\xi_2,\psi) \ldots p(\xi_m|\psi)\, d\xi_m \ldots d\xi_1 \right\} p(\psi)$$

Co takhle považovat „skrytá" data $\xi_1, \ldots, \xi_m$ za další parametry modelu s apriorním rozdělením (spolu s ψ):

$$p(\xi_1, \ldots, \xi_m, \psi) \overset{\text{VŽDY}}{=} p(\xi_1|\xi_2, \ldots, \xi_m, \psi) \cdot p(\xi_2|\xi_3, \ldots, \xi_m, \psi) \ldots p(\xi_m|\psi)\, p(\psi)$$

POUŽIJEME TOTÉŽ CO V HIERAR. SPECIFIKACI MODELU PRO VĚROHOD.

$$p(\xi_1|\xi_2,\psi) \cdot p(\xi_2|\xi_3,\psi) \ldots p(\xi_m|\psi)\, p(\psi)$$

Potom
$$p(\xi_1, \ldots, \xi_m, \psi|y) \propto p(y|\xi_1, \ldots, \xi_m, \psi) \cdot p(\xi_1, \ldots, \xi_m, \psi) =$$

$$\overset{\text{hierar. model}}{=} p(y|\xi_1,\psi)\, p(\xi_1|\xi_2,\psi) \ldots p(\xi_m|\psi)\, p(\psi)$$

JAK POTOM VYPADÁ (marginální)

$$p(\psi|y) = \int p(\xi_1, \ldots, \xi_m, \psi|y)\, d\xi_m \ldots d\xi_1$$
$$= (*)$$

HLE!

Zapamatuj 2: Y hierarchicky specifikovanou věrohodností

$$p(y|\psi) = \int p(y|\xi_1,\psi)\, p(\xi_1|\xi_2,\psi) \ldots p(\xi_{m-1}|\xi_m,\psi) \cdot p(\xi_m|\psi)\, d\xi_m \ldots d\xi_1$$

a sdruženým apriorním rozdělením

$$p(\xi_1, \ldots, \xi_m, \psi) = p(\xi_1|\xi_2,\psi) \ldots p(\xi_m|\psi)\, p(\psi)$$

(jehož ξ část odpovídá modelu)

je ~~$p(\psi|y)$ max~~

$p(\psi|y)$ marginálním rozdělením

z $p(\xi_1, \ldots, \xi_m, \psi|y)$

→ Zejména pro Monte Carlo (Markov chain) může být (většinou je) výhodné pracovat primárně s

$$p(\xi_1, \ldots, \xi_m, \psi|y) \propto p(y|\xi_1,\psi)\, p(\xi_1|\xi_2,\psi) \ldots p(\xi_m|\psi)\, p(\psi)$$

a integrál $p(\psi|y) = \int \boxed{\quad} \, d\xi_m \ldots d\xi_1$

počítat pouze když to je skutečně potřeba (při MC málokdy)

→ potřebuji (náhodný) výběr z $p(\psi|y)$

≡ příslušné složky (náhodného) výběru z $p(\xi_1, \ldots, \xi_m, \psi|y)$

Shrnutí (zjednodušené na pouze 1 úroveň skrytých dat)

DATA: y

SKRYTÁ DATA: ξ

PARAMETRY (SKUTEČNÉ): ψ

$$p(y|\psi) = \int p(y|\xi,\psi)\, p(\xi|\psi)\, d\xi$$

$p(\psi)$ = apriorní rozdělení „skutečné"

užitečné uvažovat jako parametry též ξ s apriorním $p(\xi,\psi) = p(\xi|\psi)\,p(\psi)$

strukturální (modelová) část aprior. rozděl.

$\theta := (\xi,\psi)$ ≡ parametry pro Bayes. model

$\Rightarrow$ $p(\psi|y)$ je marginálním z $p(\xi,\psi|y)$

tj. $p(\psi|y) = \int p(\xi,\psi|y)\, d\xi$

Též „skutečné“ apriorní rozdělení může být specifikováno hierarchicky
(dále opět pouze 1 úroveň)

tj: $$p(\psi) = \int \underline{p(\psi|\xi)}\, \underline{p(\xi)}\, d\xi$$

Při počítání je užitečné uvažovat jako parametry též ξ s apriorním

$$p(\underset{\sim}{\xi}, \psi, \xi) = \underline{p(\underset{\sim}{\xi}|\psi)}\, p(\psi|\xi)\, p(\xi)$$

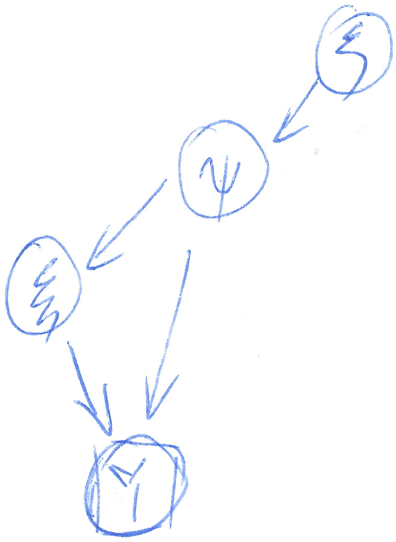
jako na minulé str

$\theta = (\underset{\sim}{\xi}, \psi, \xi)$ = parametry pro Bayes model

$\Rightarrow$ $p(\psi|y)$ je marginálním z $p(\underset{\sim}{\xi}, \psi, \xi | y)$

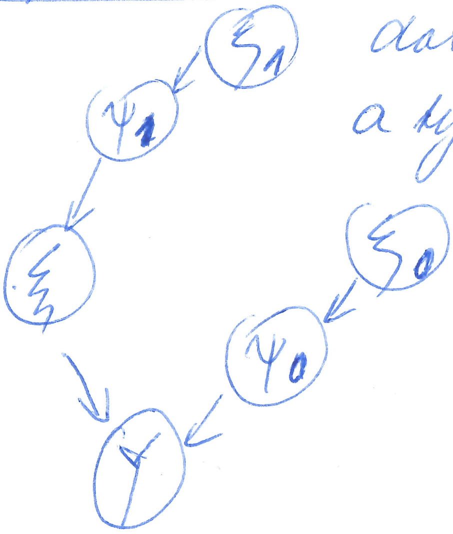
tj: $$p(\psi|y) = \iint p(\underset{\sim}{\xi}, \psi, \xi | y)\, d\xi\, d\underset{\sim}{\xi}$$

Pozn: struktura hierarchického modelu
$\Leftrightarrow$ Directed Acyclic Graph (DAG)



17

Obvykle (ne nutně): Každá hierarchická úroveň dat má své „skutečné" parametry a ty mají své hyperparametry



$$\Theta = (\underbrace{\xi}_{\text{skrytá data}}, \overbrace{\psi_0, \psi_1}^{\text{„skutečné" parametry}}, \underbrace{\zeta_0, \zeta_1}_{\text{hyperparametry}})$$

→ MODEL (věrohodnost pro Bayes model)

$$p(y|\Theta) = p(y|\xi,\psi_0)$$

APRIORNÍ pro Θ

$$p(\Theta) = p(\xi,\psi_0,\psi_1,\zeta_0,\zeta_1) =$$

$$= \underbrace{p(\xi|\psi_1)}_{\substack{\text{strukturální} \\ \text{část apriorního}}} \cdot p(\psi_1|\zeta_1)\,p(\zeta_1) \cdot p(\psi_0|\zeta_0)\,p(\zeta_0)$$

(= též „model" část apriorního rozdělení)

Frekventistická věrohodnost $(\psi = (\psi_0,\psi_1))$

$$L_F(\psi) = p(y|\psi) = \int p(y,\xi|\psi)\,d\xi =$$

$$= \int p(y|\xi,\psi)\,p(\xi|\psi)\,d\xi =$$

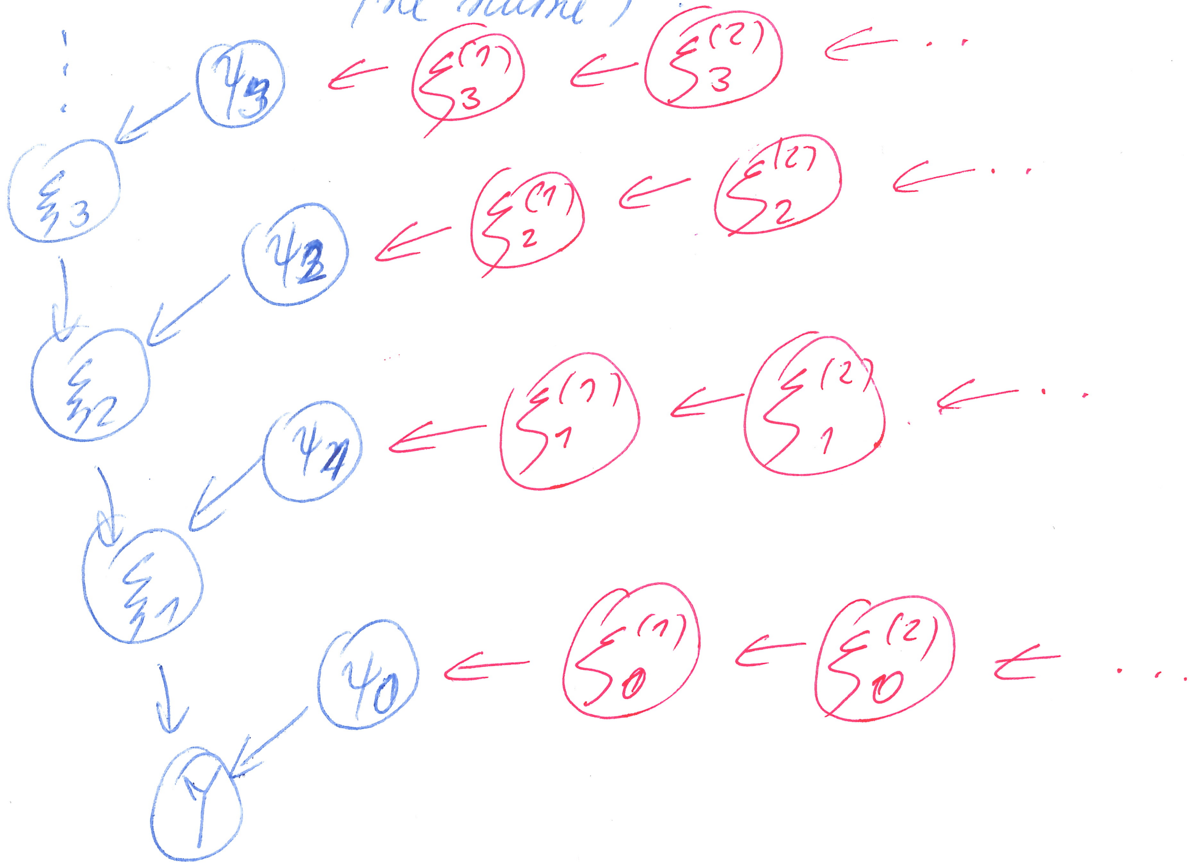
$$\overset{\text{hierarchie}}{\underset{\text{v DAG}}{=}} \int p(y|\xi,\psi_0)\,p(\xi|\psi_1)\,d\xi$$

APOSTERIORNÍ PRO Θ:

$p(\Theta|y) \propto$ SOUČIN ŠIPEK Z DAG $= p(y|\xi,\psi_0)\,p(\xi|\psi_1)\,p(\psi_1|\zeta_1) \cdot p(\zeta_1) \cdot p(\psi_0|\zeta_0)\,p(\zeta_0)$

TOP: Jak data generující mechanismus, tak apriorní rozdělení mohou být specifikovány hierarchicky s více než jednou úrovní.

Typicky: Každá úroveň má své parametry (ne nutně)



$$\theta = (\xi_1, \xi_2, \xi_3, \dots \psi_0, \psi_1, \psi_2, \psi_3, \dots, \xi_0^{(1)}, \dots)$$

= "parametry" pro Bayes. model

aposteriorní pro θ:

$p(\theta | y) \propto$ SOUČIN ŠIPEK Z DAG

$$= \underbrace{p(y | \xi_1, \psi_0)\, p(\xi_1 | \xi_2, \psi_1)\, p(\xi_2 | \xi_3, \psi_2)\, p(\xi_3 | \psi_3) \dots}_{\text{MODEL (MECHANISMUS PRO DATA)}}$$

APRIORNÍ (HIERARCH.)

$$p(\psi_0 | \xi_0^{(1)})\, p(\xi_0^{(1)} | \xi_0^{(2)}) \dots$$

$$p(\psi_3 | \xi_3^{(1)})\, p(\xi_3^{(1)} | \xi_3^{(2)}) \dots$$

"ANOVA" hierarchický model

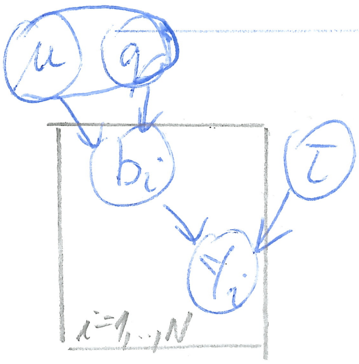
$Y_{ij} | b_i \sim N(b_i, \sigma^2)$

$b_i \sim N(\mu, d^2)$

$Y = (Y_1, \dots, Y_N)$

$Y_i = (Y_{i1}, \dots, Y_{in_i})$

$\xi = (b_1, \dots, b_N)$, $\psi = (\tau, \mu, q)$ $\quad \tau = \sigma^{-2}$, $q = d^{-2}$



$$\Theta = (\underbrace{b_1, \dots, b_N}_{\xi}, \underbrace{\tau, \mu, q}_{\psi})$$

šipky

$$p(\Theta | y) \propto \underbrace{\prod_{i=1}^{N} \{ p(y_i | b_i, \tau) \, p(b_i | \mu, q) \}}_{\text{dáno modelem } N(\dots)} \underbrace{p(\mu, \cancel{\xi}, q) \, p(\tau)}_{= p(\mu, q, \tau)}$$

$$p(\mu, q, \tau) = p(\mu, q) \, p(\tau)$$

↑ předpokládáme

např. konjugované rozdělení

$$p(\mu, q) = \underbrace{p(\mu | q)}_{\sim N(\mu_0, k_0^{-1} q^{-1})} \, p(q) \sim Ga(a_0, b_0)$$

$$p(\tau) \sim Ga(c_0, d_0)$$

hyperparametry: ~~k_0~~ $\mu_0, k_0, a_0, \underline{b_0}, c_0, \underline{d_0}$

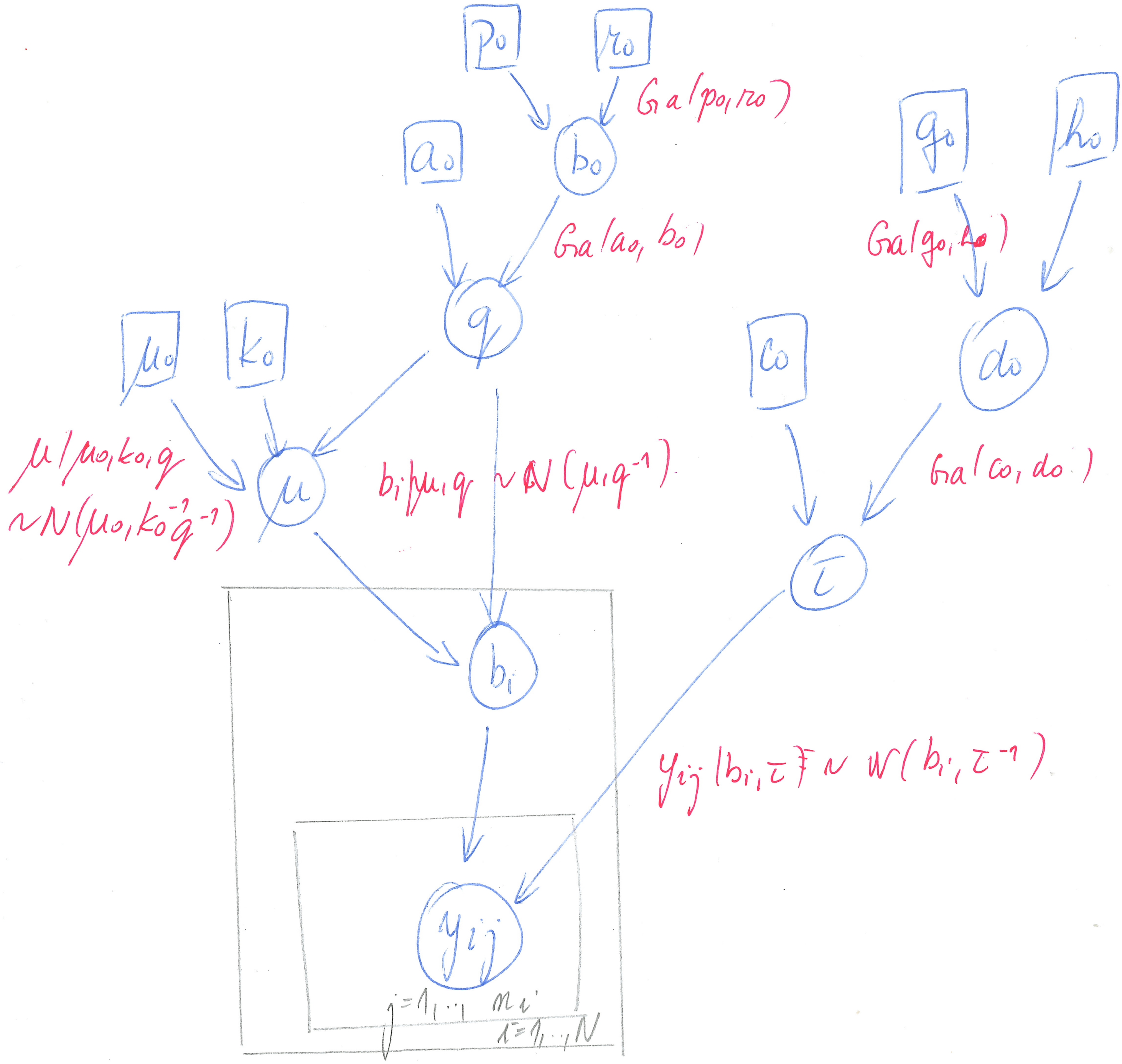
nechť náhodné

$\xi = (b_0, d_0)$

$p(\xi) = p(b_0) \cdot p(d_0)$

↑ předp. nezáv.

DAG včetně pevných hyperparametrů



navíc: $p(b_0) \sim Ga(p_0, r_0)$

$p(d_0) \sim Ga(g_0, h_0)$

$$\Theta = (\underbrace{b_1, \dots, b_N}_{\xi}; \underbrace{\tau, \mu, q}_{\psi}, \underbrace{b_0, d_0}_{\xi})$$

$p(\Theta | y) \propto$ SOUČIN ČIPEK Z DAG . . . . .

Jak volit pevné hyperparametry?

→ potřeba chápat význam jednotlivých parametrů

20

21–22