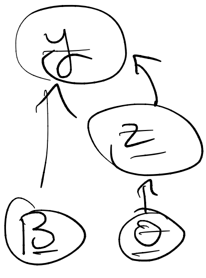
Hierarchical models:

Berliner (1996):



Data: $[\underline{y} \mid \underline{z}, \underline{\beta}]$

process: $[\underline{z} \mid \underline{\theta}]$

parameters: $[\underline{\beta}, \underline{\theta}]$

Account for data-level variation while modeling latent process.

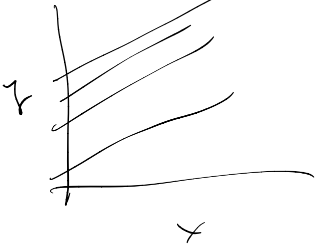
Also called: 1.) multi-level models

2.) mixed effects models
↳ fixed and random effects

3.) state-space models, especially when temporal

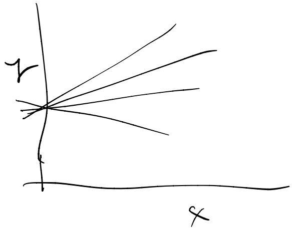
Regression scenarios:

a



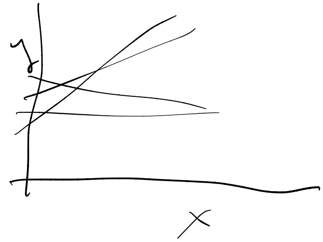
$\beta_{0i} + \beta_1 X_{ij}$

b



$\beta_0 + \beta_{1i} X_{ij}$

c



$\beta_{0i} + \beta_{1i} X_{ij}$

It is common to consider varying coefficients as "random" effects:

For example

a.) $i = 1, \ldots, n$, $j = 1, \ldots, J$

$$y_{ij} \sim N(\beta_{0i} + \beta_1 x_{ij}, \sigma^2) \quad \Big] \text{ data}$$

$$\beta_{0i} \sim N(\mu_0, \sigma_0^2) \quad \Big] \text{ process}$$

$$\left.\begin{aligned} \mu_0 &\sim N(\mu_{00}, \sigma_{00}^2) \\ \sigma_0^2 &\sim IG(q_0, r_0) \\ \beta_1 &\sim N(\mu_1, \sigma_1^2) \\ \sigma^2 &\sim IG(q, r) \end{aligned}\right] \text{ parameters}$$

This is called the "random intercepts" model)

b.)

$$y_{ij} \sim N(\beta_0 + \beta_{1i} x_{ij}, \sigma^2)$$

$$\beta_1 \sim N(\mu_1, \sigma_1^2)$$

$$\mu_1 \sim N(\mu_{10}, \sigma_{10}^2)$$
$$\sigma_1^2 \sim IG(q_1, r_1)$$

$$\beta_0 \sim N(\mu_0, \sigma_0^2)$$
$$\sigma^2 \sim IG(q, r)$$

This called the "random slopes" model

c.)

$$y_{ij} \sim N(\beta_{0i} + \beta_i x_{ij}, \sigma^2) \quad \Big] \text{ data}$$

$$\underline{\beta}_i \sim N(\underline{\mu}, \Sigma), \quad \Sigma = \sigma_\beta^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Big] \text{ process}$$

$$\left.\begin{aligned} \underline{\mu} &\sim N(\underline{\mu}_0, \Sigma_0) \\ \sigma_\beta^2 &\sim IG(q_\beta, r_\beta) \\ \rho &\sim \text{Unif}(-1, 1) \\ \sigma^2 &\sim IG(q, r) \end{aligned}\right] \text{ parameters}$$

All coefficients random!
(no "fixed" effects)

Note: data model can be written as:

$$\underline{y}_i \sim N(X_i \underline{\beta}_i, \sigma^2 I)$$

Posterior:

$$[\underline{\beta}_1, \ldots, \underline{\beta}_n, \underline{\mu}, \sigma_\beta^2, \rho, \sigma^2 \mid \underline{y}_1, \ldots, \underline{y}_n] \propto$$
$$\left(\prod_{i=1}^{n} [\underline{y}_i \mid \underline{\beta}_i, \sigma^2][\underline{\beta}_i \mid \underline{\mu}, \sigma_\beta^2, \rho]\right) [\underline{\mu}][\sigma_\beta^2][\rho][\sigma^2]$$

Full-conditional Distributions:

$$[\underline{\beta}_i \mid \cdot] \propto [\underline{y}_i \mid \underline{\beta}_i, \sigma^2][\underline{\beta}_i \mid \underline{\mu}, \sigma_\beta^2, \rho] \quad \left(\text{from before}\right)$$
$$= N(A^{-1}\underline{b}, A^{-1})$$

$$A = X_i'(\sigma^2 I)^{-1} X_i + \Sigma^{-1} \qquad \text{for } i = 1, \ldots, n$$
$$\underline{b} = X_i'(\sigma^2 I)^{-1} \underline{y}_i + \Sigma^{-1} \underline{\mu}$$

$$[\underline{\mu} \mid \cdot] \propto \left(\prod_{i=1}^{n} [\underline{\beta}_i \mid \underline{\mu}, \sigma_\beta^2, \rho]\right)[\underline{\mu}]$$

$$\propto \exp\left\{-\frac{1}{2}\sum_i (\underline{\beta}_i - \underline{\mu})' \Sigma^{-1} (\underline{\beta}_i - \underline{\mu})\right\} \exp\left\{-\frac{1}{2}(\underline{\mu} - \underline{\mu}_0)' \Sigma_0^{-1} (\underline{\mu} - \underline{\mu}_0)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left(-2\underbrace{\left(\sum_{i=1}^{n} \underline{\beta}_i' \Sigma^{-1} + \underline{\mu}_0' \Sigma_0^{-1}\right)}_{\underline{b}'}\underline{\mu} + \underline{\mu}'\underbrace{\left(n\Sigma^{-1} + \Sigma_0^{-1}\right)}_{A}\underline{\mu}\right)\right\}$$

$$= N(A^{-1}\underline{b}, A^{-1})$$

$$[\sigma_\beta^2 \mid \cdot] \propto \left(\prod_{i=1}^{n} [\underline{\beta}_i \mid \underline{\mu}, \sigma_\beta^2, \rho]\right)[\sigma_\beta^2], \quad R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

$$\propto (\sigma_\beta^2)^{-n} \exp\left\{-\frac{1}{\sigma_\beta^2} \frac{\sum_{i=1}^{n} (\underline{\beta}_i - \underline{\mu})' R^{-1} (\underline{\beta}_i - \underline{\mu})}{2}\right\} (\sigma_\beta^2)^{-(q_\beta+1)} \exp\left\{-\frac{1}{\sigma_\beta^2 r_\beta}\right\}$$

$$\propto (\sigma_\beta^2)^{-(\underbrace{n+q_\beta}_{\tilde{q}}+1)} \exp\left\{-\frac{1}{\sigma_\beta^2}\underbrace{\left(\frac{\sum_{i=1}^{n} (\underline{\beta}_i - \underline{\mu})' R^{-1} (\underline{\beta}_i - \underline{\mu})}{2} + \frac{1}{r_\beta}\right)}_{\tilde{r}^{-1}}\right\}$$

$$= IG(\tilde{q}, \tilde{r})$$

$$[\rho \mid \cdot] \propto \left(\prod_{i=1}^{n} [\underline{\beta}_i \mid \underline{\mu}, \sigma_\beta^2, \rho]\right)[\rho] \qquad (\text{use } M\text{-}H)$$

$$[\sigma^2 \mid \cdot] \propto \left(\prod_{i=1}^{n} [\underline{y}_i \mid \underline{\beta}_i, \sigma^2]\right)[\sigma^2]$$

$$\propto (\sigma^2)^{-\frac{nJ}{2}} \exp\left\{-\frac{1}{\sigma^2} \frac{\sum_{i=1}^{n} (\underline{y}_i - X_i\underline{\beta}_i)'(\underline{y}_i - X_i\underline{\beta}_i)}{2}\right\} (\sigma^2)^{-(q+1)} \exp\left\{-\frac{1}{\sigma^2 r}\right\}$$

$$\propto (\sigma^2)^{-(\underbrace{\frac{nJ}{2}+q}_{\tilde{q}}+1)} \exp\left\{-\frac{1}{\sigma^2}\underbrace{\left(\frac{\sum_i (\underline{y}_i - X_i\underline{\beta}_i)'(\underline{y}_i - X_i\underline{\beta}_i)}{2} + \frac{1}{r}\right)}_{\tilde{r}}\right\}$$

$$= IG(\tilde{q}, \tilde{r})$$

Generalizations to hierarchical model:

1.) Let data-level variance vary $\sigma_i^2 \sim IG(q, r)$

2.) Let data-level variance be random:

$$\log \sigma_i \sim N(\mu_\sigma, \sigma_\sigma^2)$$
$$\mu_\sigma \sim N(\mu_{\sigma_0}, \sigma_{\sigma_0}^2)$$
$$\sigma_\sigma^2 \sim IG(q_\sigma, r_\sigma)$$

3.) Let β_{0i}, β_{1i} have diff. variances:

$$\beta_{0i} \sim N(\mu_0, \sigma_0^2)$$
$$\beta_{1i} \sim N(\mu_1, \sigma_1^2) \Rightarrow \rho = 0$$

4.) Let $\underline{\beta}_i$ have fully unknown covariance:

$$\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho \\ \sigma_0\sigma_1\rho & \sigma_1^2 \end{bmatrix}$$

recall correlation:

$$\rho = \frac{\sigma_0\sigma_1\rho}{\sqrt{\sigma_0^2\sigma_1^2}}$$

5.) Scalable prior for Σ: $\Sigma^{-1} \sim \text{Wish}((S\nu)^{-1}, \nu)$
"wishart" w/ mean:

$$E(\Sigma^{-1}) = (S\nu)^{-1}\nu = S^{-1}$$

6.) GLMMs: $y_i \sim \text{Pois}(\lambda_i)$

$$\log(\lambda_i) = X_i\underline{\beta}_i, \quad \underline{\beta}_i \sim N(\underline{\mu}, \Sigma)$$

7.) More covariates: $\underset{p\times 1}{\underline{\beta}_i} \sim N(\underline{\mu}, \underset{p\times p}{\Sigma})$