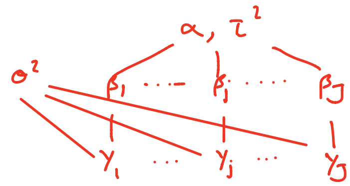
Hierarchical Linear Models

$$\underbrace{Y|\beta,\sigma^2}\sim N(X\beta,\sigma^2 I_n)$$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_J \end{bmatrix} \qquad Y_j|\beta_j,\Sigma_j \sim N(X_j\beta_j,\Sigma_j)$$



$$\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_J \end{pmatrix} \Big| \, \alpha,\tau^2 \sim N(X_\beta\alpha, \tau^2 I_{pJ})$$

$$P(\alpha)\propto 1, \; \sigma^2 \sim Inv\text{-}\chi^2(\nu_0, s_0), \; \tau^2 \sim Inv\text{-}\chi^2(\nu_\tau, s_\tau)$$

Gibbs Sampler

$$\sigma^2|\alpha,\beta,\tau^2,Y \sim Inv\text{-}\chi^2\left(n+\nu_0, \frac{RSS(\beta)+\nu_0 s_0}{n+\nu_0}\right)$$

$$\tau^2|\alpha,\beta,\sigma^2,Y \sim Inv\text{-}\chi^2\left(pJ+\nu_\tau, \frac{(\beta-X_\beta\alpha)^T(\beta-X_\beta\alpha)+\nu_\tau s_\tau}{p+\nu_\tau}\right)$$

$$\beta|\alpha,\tau^2,\sigma^2,Y \sim N\left(\hat\beta=\Sigma_\beta\left(\frac{1}{\sigma^2}X^Ty+\frac{1}{\tau^2}X_\beta\right), \; \Sigma_\beta=\left(\frac{1}{\sigma^2}X^TX+\frac{1}{\tau^2}I\right)^{-1}\right)$$

follow linear model sampling

flat prior derivation

$$\alpha|\beta,\tau^2,\sigma^2,Y\sim N\left(\hat\beta=(X_\beta^TX_\beta)^{-1}X_\beta^T\beta, \; \tau^2(X_\beta^TX_\beta)^{-1}\right)$$

Bayesian GLM

Consider the model

$$Y_i|\beta,\sigma^2 \overset{ind}{\sim} N(x_i^T\beta, \sigma^2/w_i)$$, known weights

$$Y|\beta,\sigma^2 \sim N(X\beta,\sigma^2W^{-1}) \qquad W=(diag(w_i))_{i=1}^n$$

$$\Rightarrow W^{1/2}Y|\beta,\sigma^2\sim N(W^{1/2}X\beta,\sigma^2 I)$$

__Idea__: Fit $W^{1/2}Y\sim W^{1/2}X$ and do standard Bayesian GLM.

$$\boxed{\beta \mid \sigma^2, y \sim N\left(\Sigma_\beta\left(\tfrac{1}{\sigma^2}X^T W Y + \Sigma_0^{-1}\beta_0\right), \Sigma_\beta = \left(\tfrac{1}{\sigma^2}X^T W X + \Sigma_0^{-1}\right)^{-1}\right)}$$

For a GLM

$Y_i \mid \beta, \sigma^2 \overset{ind}{\sim} F(x_i^T\beta, \sigma^2)$ s.t. F satisfies

(i) $\mathbb{E}[Y_i \mid \beta, \sigma^2] = \mu_i(\beta) = \underbrace{g^{-1}}_{\text{link function}}(x_i^T\beta)$

(ii) $\operatorname{Var}(Y_i \mid \beta, \sigma^2) = \underbrace{\sigma^2}_{\text{dispersion}} \underbrace{V}_{\text{variance function}}(\mu_i(\beta))$

Fit these models based on a Laplace Approx.

$\pi(\beta) = \log \mathbb{P}(X \mid \beta, \sigma^2) + \log \mathbb{P}(\beta, \sigma^2)$ and Taylor expand around $\beta^{(t)}$

$$\pi(\beta) \simeq \pi(\beta^{(t)}) + (\beta - \beta^{(t)})^T \underbrace{\frac{d\pi}{d\beta}(\beta^{(t)})}_{\substack{U(\beta^{(t)}) \\ = \text{Score}}} - \frac{1}{2}(\beta - \beta^{(t)})^T \underbrace{\left(-\frac{\partial^2 \pi}{\partial\beta\,\partial\beta^T}\right)}_{\substack{\Sigma_\beta(\beta^{(t)})^{-1} \\ = \text{Hessian}}} (\beta - \beta^{(t)})$$

$$\mathbb{P}(\beta \mid \sigma^2, y) \propto \exp\left\{-\tfrac{1}{2}(\beta - \hat\beta^{(t)})^T \Sigma_\beta^{-1}(\beta^{(t)})(\beta - \hat\beta^{(t)})\right\} \,\dot\sim\, N\left(\hat\beta^{(t)}, \Sigma_\beta(\beta^{(t)})\right)$$

– Used as a proposal in MH or in IRWLS updates

With $\Sigma_\beta(\beta^{(t)})^{-1} = \frac{1}{\sigma^2} X^T W(\beta^{(t)}) X + \Sigma_0^{-1}$

$$\hat\beta^{(t)} = \beta^{(t)} + \Sigma_\beta(\beta^{(t)}) U(\beta^{(t)}) = \beta^{(t)} + \Sigma_\beta(\beta^{(t)})\left(\frac{1}{\sigma^2}X^T(Y - \mu(\beta^{(t)})) - \Sigma_0^{-1}(\beta^{(t)} - \beta_0)\right)$$

Rmk: W depends on g, V

In this case, $\Sigma_\beta(\beta^{(t)})^{-1}\hat\beta^{(t)} = \Sigma_\beta(\beta^{(t)})^{-1}\beta^{(t)} + \frac{1}{\sigma^2}X^T(y - \mu(\beta^{(t)})) - \Sigma_0^{-1}\beta^{(t)} + \Sigma_0^{-1}\beta^{(t)}$

$$= \frac{1}{\sigma^2} X^T W(\beta^{(t)}) \underbrace{\left[X\beta^{(t)} + W(\beta^{(t)})^{-1}(Y - \mu(\beta^{(t)}))\right]}_{Z = \text{“Working Residual”}} + \Sigma_0^{-1}\beta_0$$

$$\hat{\beta}^{(t)} = \Sigma_\beta(\beta^{(t)})\left[{}^{1}/_{\sigma^2}\, W(\beta^{(t)})\, z^{(t)} + \Sigma_0^{-1}\beta_0 \right]$$

$\boxed{\text{rstanarm}}$