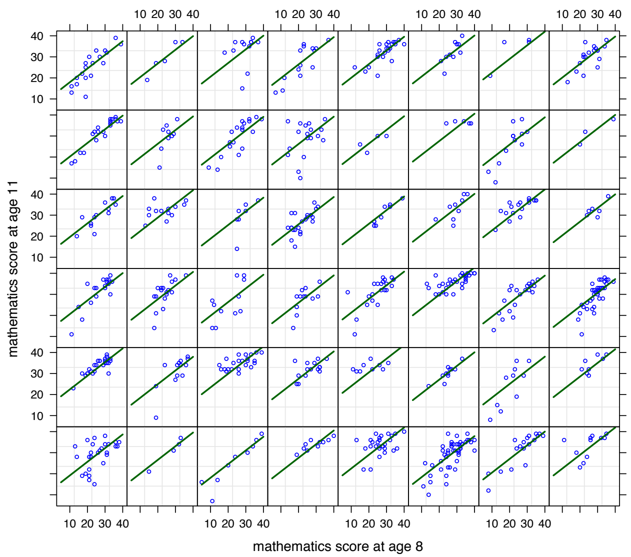
37458

Advanced Bayesian Methods

Bayesian Models for Grouped Data: Additional Aspects



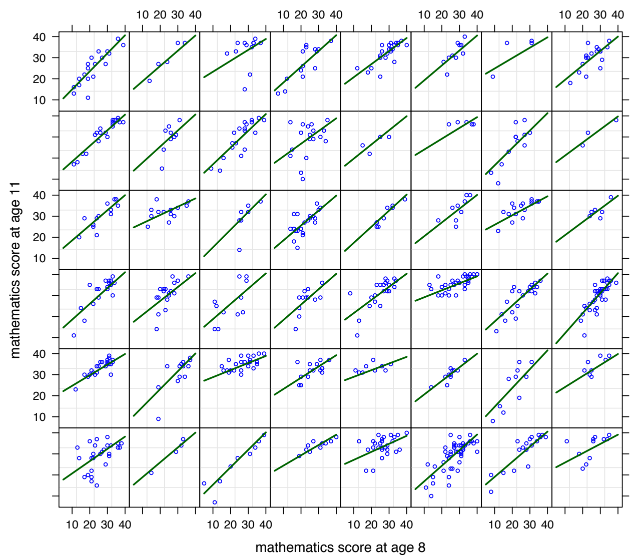
Random Intercepts Model Recap

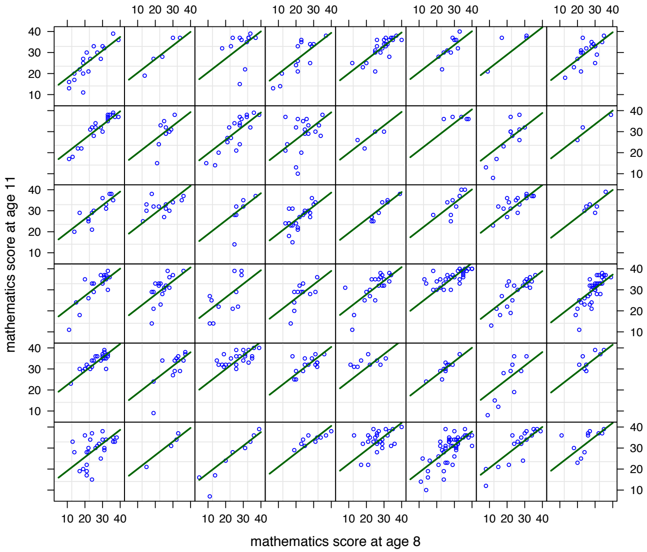
The previous slide fit the Bayesian random intercepts model

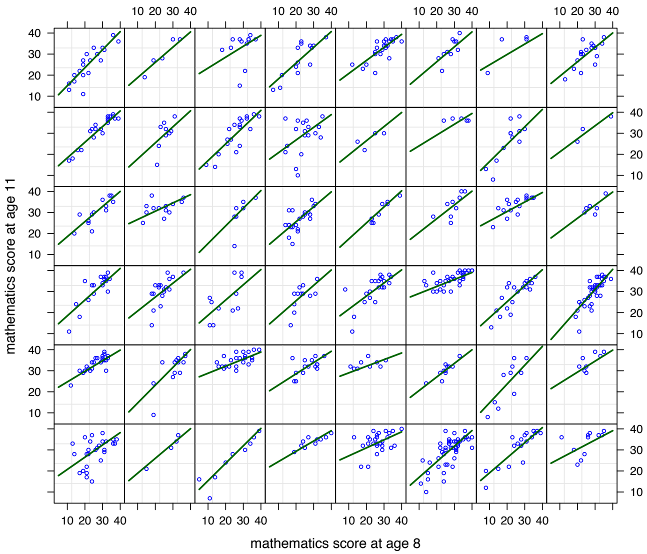
$$y_{ij}|\beta_0,\beta_1,\beta_2,u_i,\sigma_\varepsilon^2 \overset{\text{ind.}}{\sim} N\Big((\beta_0+u_i)+\beta_1\,x_{ij},\sigma_\varepsilon^2\Big)$$

$$u_i|\sigma_u^2 \overset{\text{ind.}}{\sim} N\big(0,\sigma_u^2\big)$$

LIMITATION: The lines for each school are parallel (i.e. same slope).







Random Intercepts and Slopes Models

$$y_{ij}|\beta_0, \beta_1, \beta_2, u_{0i}, u_{1i}, \sigma^2_\varepsilon \overset{\text{ind.}}{\sim} N\Big((\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\, x_{ij}, \sigma^2_\varepsilon\Big),$$

$$\left[\begin{array}{c} u_{0i} \\ u_{1i} \end{array}\right] \Bigg| \sigma^2_{u0}, \sigma^2_{u1}, \rho_u \overset{\text{ind.}}{\sim} N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma^2_{u0} & \rho_u\sigma_{u0}\sigma_{u1} \\ \rho_u\sigma_{u0}\sigma_{u1} & \sigma^2_{u1} \end{array}\right]\right).$$

Note that

$$\left[\begin{array}{cc} \sigma^2_{u0} & \rho_u\sigma_{u0}\sigma_{u1} \\ \rho_u\sigma_{u0}\sigma_{u1} & \sigma^2_{u1} \end{array}\right]$$

is a general 2×2 covariance matrix where

$-1 < \rho_u < 1$ is the correlation between the u_{0i} and u_{1i}.

Look at Stan code

HALF-CAUCHY PRIORS

FOR STANDARD DEVIATION

PARAMETERS

Variance Component Prior Debate (early 2000s)

Until about 2000 most Bayesian analysts used

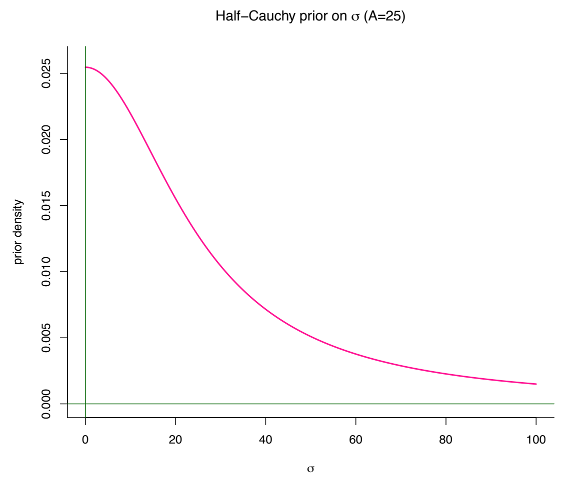
$$p(\sigma^2) \sim \text{Inverse-Gamma}(\varepsilon, \varepsilon) \quad (\varepsilon \text{ 'small'})$$

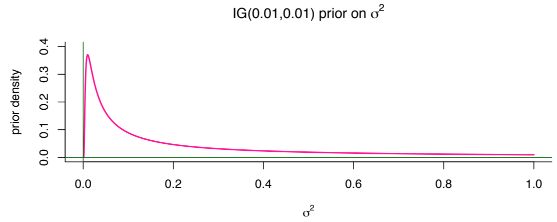
as a 'vague' prior for variance components – because of its conjugate status.

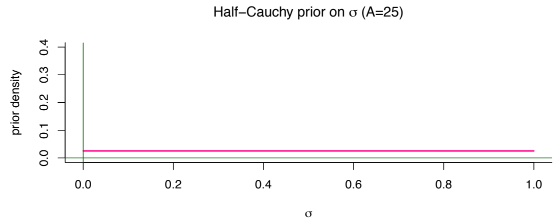
Recent literature (e.g. Gelman, 2006, *Bayesian Analysis*) has criticised this choice since the Inverse-Gamma$(\varepsilon, \varepsilon)$ distribution is not very vague.

Gelman's default recommendation is half-Cauchy with high scale parameter.

$$p(\sigma) = \frac{2A}{\pi(\sigma^2 + A^2)}, \quad \sigma^2 > 0, \quad A \text{ 'big'}$$







Look at Stan code

Covariance Matrix Priors

In 2012 (here in Ultimo, N.S.W.) Alan Huang and Matt Wand extended the Gelman Half-Cauchy prior idea to covariance matrix priors and published:

Huang, A. and Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, **8**, 439–452.

(Naturally) the class's Stan scripts use this prior – but technical details left aside (but Huang & Wand paper on lecturer's web-pages).

SCALING IN

BAYESIAN ANALYSIS

A New Species of Snake!

A zoologist discovers a new species of snake in a remote part of the Amazon. The snakes are very long and the zoologist measures 100 of them.

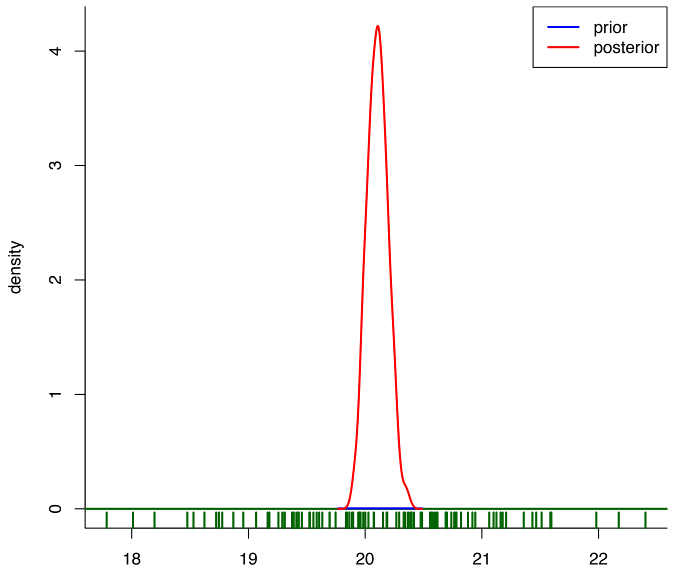
The data are sent back to her lab and a statistician fits the following Bayesian model in Stan:

$$x_i|\mu, \sigma^2 \overset{\text{ind.}}{\sim} N(\mu, \sigma^2)$$

$$\mu \sim N(0, 100^2), \quad \sigma \sim \text{Half-Cauchy}(0, 100^2)$$

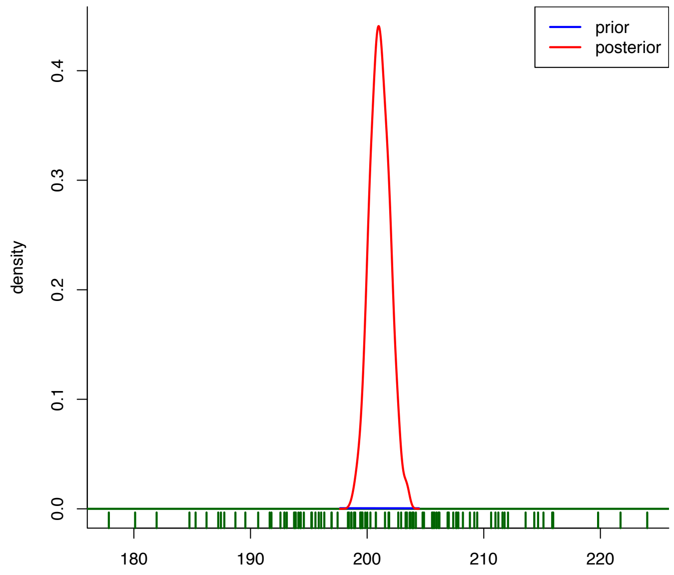
where $x_1, \ldots, x_{100}$ are the snake lengths in metres. The statistician wants to make Bayesian inference about:

$$\mu = \text{mean length of snake species.}$$



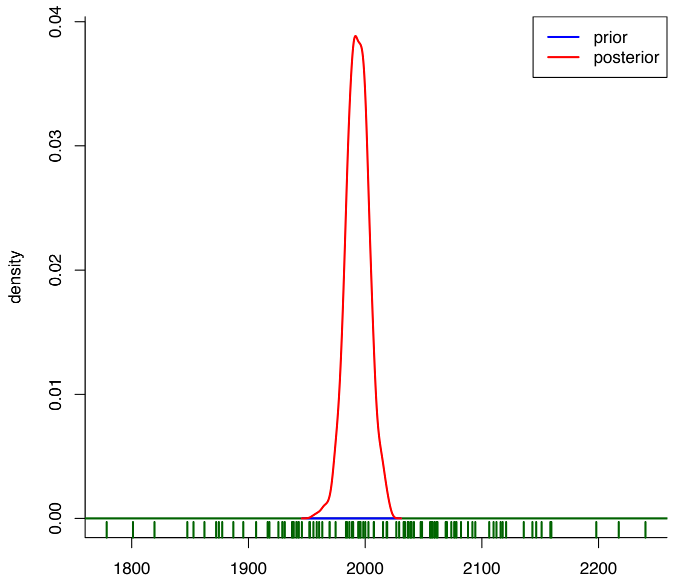
INSTEAD, MEASURE LENGTH

IN DECIMETRES



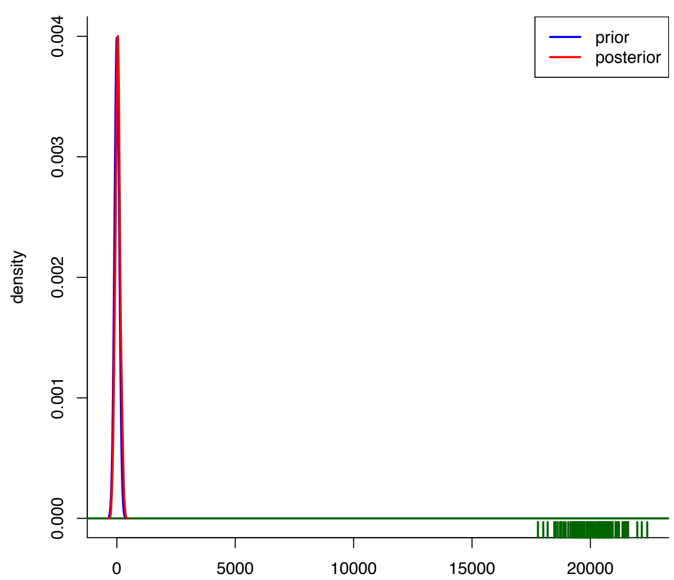
INSTEAD, MEASURE LENGTH

IN CENTIMETRES



INSTEAD, MEASURE LENGTH

IN MILLIMETRES



Summary of Bayesian Analyses With Different Length Units

length units	classical sample mean	Bayes estimate of μ
metres	20.11	20.11
decimetres	201.1	201.1
centimetres	2011	1993
millimetres	20110	44.7

CONCLUSION:

Bayesian inference with a fixed prior specification (such as $\mu \sim N(0, 100^2)$) depends on the units of measurement – sometimes crucially.

For big spreadsheets of data with dozens of columns this is a concern, since there can be data of many variable types on different scales and with different units.

Remedy

If

$$x_1^{\text{orig}}, \ldots, x_{100}^{\text{orig}}$$

are the original snake length data then first standardise to

$$x_i \equiv \frac{x_i^{\text{orig}} - \text{sample mean}}{\text{sample standard deviation}}.$$

Then alright to use:

$$x_i|\mu \overset{\text{ind.}}{\sim} N(\mu, \sigma^2)$$

$$\mu \sim N(0, 100), \quad \sigma \sim \text{Half-Cauchy}(0, 100)$$

Conversion Back to Original Units

If we use a Bayesian inference engine to do regression analysis for a model such as:

$$\text{price}_i|\beta_0, \beta_1, \sigma^2 \overset{\text{ind.}}{\sim} N\left(\beta_0 + \beta_1 \, \text{age}_i, \sigma^2\right)$$

for data on used cars then

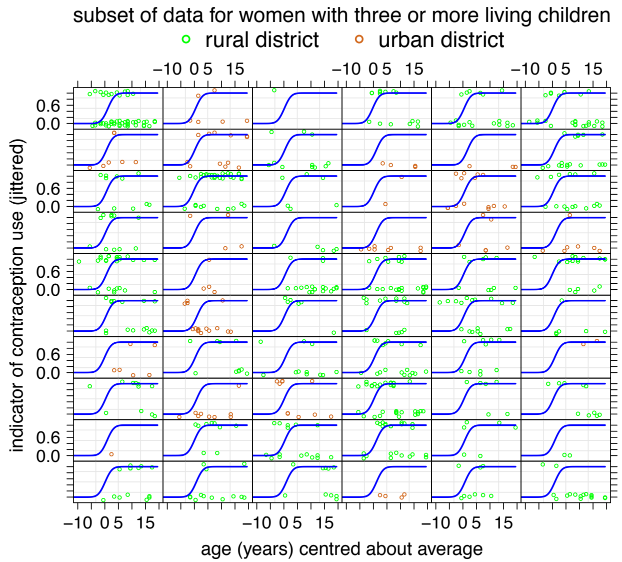
$$\beta_1 = \text{depreciation rate.}$$

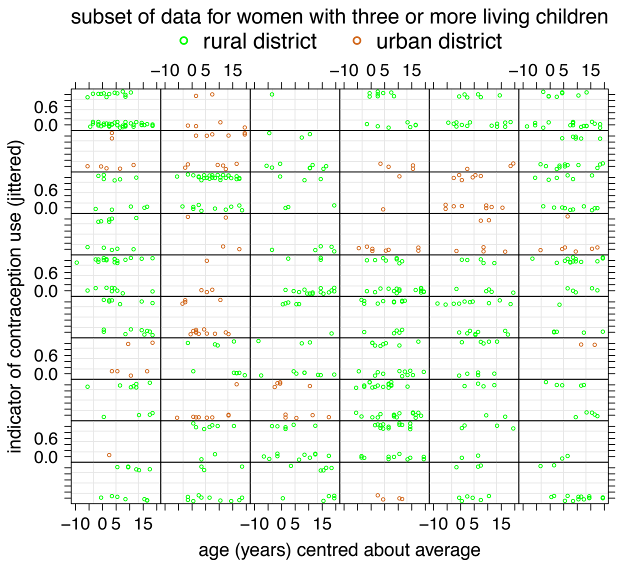
If standardised data used then best to transform β_1 to meaningful units (e.g. dollars per year).

See Exercise 1 of Assignment 6.

BINARY RESPONSE

GROUPED DATA





Logistic Mixed Models

An example is:

$$y_{ij}|\beta_0,\beta_1,\beta_2,\beta_3 \overset{\text{ind.}}{\sim} \text{Bernoulli}\left(\frac{1}{1+\exp[-\{\beta_0+u_{i0}+(\beta_1+u_{i1})x_{1ij}+\beta_2\,x_{2i}\}]}\right),$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix}\Bigg|\,\sigma^2_{u0},\sigma^2_{u1},\rho_u \overset{\text{ind.}}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\begin{bmatrix} \sigma^2_{u0} & \rho_u\sigma_{u0}\sigma_{u1} \\ \rho_u\sigma_{u0}\sigma_{u1} & \sigma^2_{u1} \end{bmatrix}\right)$$

Assignment 6, Question 4 has an example of this type.

Let's look at coding in Stan...

Markov Chain Monte Carlo Making a Difference

Without Markov chain Monte Carlo, fitting logistic mixed models with accurate inference is VERY HARD due to intractable integrals (even in the non-Bayesian case).

The availability of Bayesian inference engines makes such analyses routine – but this is quite a recent development for such an important model.

♦ ♦ ♦