Robust Bayesian Regression

Readings: Hoff Chapter 9, West JRSSB 1984, Fúquene, Pérez & Pericchi 2015

STA 721 Duke University

Duke University

November 21, 2017
Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library (BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library (BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.
Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.
- Can also use BAS or other variable selection programs
Multiple Outliers

- Hoeting, Madigan and Raftery (in various permutations) consider the problem of simultaneous variable selection and outlier identification.
- This is implemented in the library(BMA) in the function MC3.REG. This has the advantage that more than 2 points may be considered as outliers at the same time.
- The function uses a Markov chain to identify both important variables and potential outliers, but is coded in Fortran so should run reasonably quickly.
- Can also use BAS or other variable selection programs
library(MASS)
data(stackloss)
n = nrow(stackloss)
stack.out = cbind(stackloss, diag(n))

library(BAS)
BAS.stack = bas.lm(stack.loss ~ ., data=stack.out,
prior="hyper-g-n", a=3,
modelprior=tr.beta.binomial(1, 1,15),
method="MCMC", MCMC.it=200000)
### Output

<table>
<thead>
<tr>
<th>P(B != 0</th>
<th>Y)</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Air. Flow</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Water. Temp</td>
<td>0.23</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Acid. Conc.</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'1'</td>
<td>0.22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>'2'</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'3'</td>
<td>0.24</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>'4'</td>
<td>0.75</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>'5'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'6'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'7'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'8'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'9'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'10'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'11'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'12'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'13'</td>
<td>0.16</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'14'</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'15'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'16'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'17'</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'18'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'19'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'20'</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'21'</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>BF</td>
<td>0.13</td>
<td>0.01</td>
<td>0.08</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PostProbs</td>
<td>0.24</td>
<td>0.11</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.96</td>
<td>0.93</td>
<td>0.97</td>
<td>0.97</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>dim</td>
<td>4.00</td>
<td>3.00</td>
<td>5.00</td>
<td>5.00</td>
<td>7.00</td>
<td></td>
</tr>
<tr>
<td>logmarg</td>
<td>22.17</td>
<td>19.43</td>
<td>21.68</td>
<td>21.57</td>
<td>24.18</td>
<td></td>
</tr>
</tbody>
</table>
Predictions under BMA

Residuals vs Fitted

Model Probabilities

Model Search Order

Model Complexity

Inclusion Probabilities
BAS

Model Rank
Log Posterior Odds

Intercept
Air. Flow
Water. Temp
Acid. Conc.
Body Fat Data: Intervals w/ All Data

Response % Body Fat and Predictor Waist Circumference

Which analysis do we use? with Case 39 or not – or something different?
Cook’s Distance

lm(Bodyfat ~ I(Abdomen - 2.54 * 34))

Cook’s distance

Residuals vs Leverage
Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?

\[ Y = X_\beta + I_\delta + \epsilon \]

- If \[ \gamma_i = 1 \] then case \[ i \] has a different mean "mean shift" outliers.
Options for Handling Influential Cases

▶ Are there scientific grounds for eliminating the case?
▶ Test if the case has a different mean than population
Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?
- Test if the case has a different mean than population
- Report results with and without the case
Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?
- Test if the case has a different mean than population
- Report results with and without the case
- Model Averaging to Account for Model Uncertainty?
Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?
- Test if the case has a different mean than population
- Report results with and without the case
- Model Averaging to Account for Model Uncertainty?
- Full model $\mathbf{Y} = \mathbf{X}\mathbf{\beta} + \mathbf{I}_n\mathbf{\delta} + \mathbf{\epsilon}$
Options for Handling Influential Cases

- Are there scientific grounds for eliminating the case?
- Test if the case has a different mean than population
- Report results with and without the case
- Model Averaging to Account for Model Uncertainty?
- Full model \( Y = X\beta + I_n\delta + \epsilon \)
- \( 2^n \) submodels \( \gamma_i = 0 \Leftrightarrow \delta_i = 0 \)
- If \( \gamma_i = 1 \) then case \( i \) has a different mean “mean shift” outliers.
Mean Shift = Variance Inflation

Model $\mathbf{Y} = \mathbf{X}\beta + \mathbf{1}_n\delta + \epsilon$

Prior

$\delta_i \mid \gamma_i \sim N(0, V\sigma^2\gamma_i)$

$\gamma_i \sim \text{Ber}(\pi)$

Then $\epsilon_i$ given $\sigma^2$ is independent of $\delta_i$ and

$$\epsilon_i^* \equiv \epsilon_i + \delta_i \mid \sigma^2 \begin{cases} N(0, \sigma^2) & \text{wp} \quad (1 - \pi) \\ N(0, \sigma^2(1 + V)) & \text{wp} \quad \pi \end{cases}$$

Model $\mathbf{Y} = \mathbf{X}\beta + \epsilon^*$ “variance inflation”

$V + 1 = K = 7$ in the paper by Hoeting et al. package BMA
Simultaneous Outlier and Variable Selection

MC3.REG(all.y = bodyfat$Bodyfat, all.x = as.matrix(bodyfat$Abdomen),
         num.its = 10000, outliers = TRUE)

Model parameters: PI=0.02 K=7 nu=2.58 lambda=0.28 phi=2.85

15 models were selected
Best 5 models (cumulative posterior probability = 0.9939):

<table>
<thead>
<tr>
<th>variables</th>
<th>prob</th>
<th>model 1</th>
<th>model 2</th>
<th>model 3</th>
<th>model 4</th>
<th>model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>all.x</td>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>outliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>0.94932</td>
<td>x</td>
<td>x</td>
<td>.</td>
<td>x</td>
<td>.</td>
</tr>
<tr>
<td>204</td>
<td>0.04117</td>
<td>.</td>
<td></td>
<td>.</td>
<td>x</td>
<td>.</td>
</tr>
<tr>
<td>207</td>
<td>0.10427</td>
<td>.</td>
<td>x</td>
<td>.</td>
<td>.</td>
<td>x</td>
</tr>
<tr>
<td>post prob</td>
<td>0.815</td>
<td>0.095</td>
<td>0.044</td>
<td>0.035</td>
<td>0.004</td>
<td></td>
</tr>
</tbody>
</table>
Change Error Assumptions

\[ Y_i \overset{\text{ind}}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi) \]
Change Error Assumptions

\[ Y_i \overset{\text{ind}}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi) \]

\[ L(\alpha, \beta, \phi) \propto \prod_{i=1}^{n} \phi^{1/2} \left( 1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\frac{(\nu+1)}{2}} \]
Change Error Assumptions

\[ Y_i \overset{\text{ind}}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi) \]

\[ L(\alpha, \beta, \phi) \propto \prod_{i=1}^{n} \phi^{1/2} \left( 1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\left(\frac{\nu+1}{2}\right)} \]

Use Prior \( p(\alpha, \beta, \phi) \propto 1/\phi \)
Change Error Assumptions

\[ Y_i \overset{\text{ind}}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi) \]

\[ L(\alpha, \beta, \phi) \propto \prod_{i=1}^{n} \phi^{1/2} \left(1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\frac{(\nu+1)}{2}} \]

Use Prior \( p(\alpha, \beta, \phi) \propto 1/\phi \)

Posterior distribution

\[ p(\alpha, \beta, \phi \mid Y) \propto \phi^{n/2-1} \prod_{i=1}^{n} \left(1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\frac{(\nu+1)}{2}} \]
Change Error Assumptions

\[ Y_i \overset{\text{ind}}{\sim} t(\nu, \alpha + \beta x_i, 1/\phi) \]

\[ L(\alpha, \beta, \phi) \propto \prod_{i=1}^{n} \phi^{1/2} \left( 1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\frac{(\nu+1)}{2}} \]

Use Prior \( p(\alpha, \beta, \phi) \propto 1/\phi \)

Posterior distribution

\[ p(\alpha, \beta, \phi \mid Y) \propto \phi^{n/2-1} \prod_{i=1}^{n} \left( 1 + \frac{\phi(y_i - \alpha - \beta x_i)^2}{\nu} \right)^{-\frac{(\nu+1)}{2}} \]
Treat $\sigma^2$ as given, then influence of individual observations on the posterior distribution of $\beta$ in the model where $E[Y_i] = x_i^T \beta$ is investigated through the score function:
Bounded Influence - West 1984 (and references within)

Treat $\sigma^2$ as given, then influence of individual observations on the posterior distribution of $\beta$ in the model where $E[Y_i] = x_i^T \beta$ is investigated through the score function:

$$\frac{d}{d\beta} \log p(\beta | Y) = \frac{d}{d\beta} \log p(\beta) + \sum_{i=1}^{n} x g(y_i - x_i^T \beta)$$

where $g(\epsilon) = -\frac{d}{d\epsilon} \log p(\epsilon)$ is the influence function of the error distribution (unimodal, continuous, differentiable, symmetric)

An outlying observation $y_j$ is accommodated if the posterior distribution for $p(\beta | Y_i)$ converges to $p(\beta | Y)$ for all $\beta$ as $|Y_i| \rightarrow \infty$. Requires error models with influence functions that go to zero such as the Student $t$ (O'Hagan, 1979)
Treat $\sigma^2$ as given, then influence of individual observations on the posterior distribution of $\beta$ in the model where $E[Y_i] = x_i^T \beta$ is investigated through the score function:

$$\frac{d}{d\beta} \log p(\beta \mid Y) = \frac{d}{d\beta} \log p(\beta) + \sum_{i=1}^{n} x_i g(y_i - x_i^T \beta)$$

where

$$g(\epsilon) = -\frac{d}{d\epsilon} \log p(\epsilon)$$

is the influence function of the error distribution (unimodal, continuous, differentiable, symmetric)
Bounded Influence - West 1984 (and references within)

Treat $\sigma^2$ as given, then influence of individual observations on the posterior distribution of $\beta$ in the model where $E[Y_i] = x_i^T \beta$ is investigated through the score function:

$$
\frac{d}{d\beta} \log p(\beta \mid Y) = \frac{d}{d\beta} \log p(\beta) + \sum_{i=1}^{n} x_i g(y_i - x_i^T \beta)
$$

where

$$
g(\epsilon) = -\frac{d}{d\epsilon} \log p(\epsilon)
$$

is the influence function of the error distribution (unimodal, continuous, differentiable, symmetric)

An outlying observation $y_j$ is accommodated if the posterior distribution for $p(\beta \mid Y_{(i)})$ converges to $p(\beta \mid Y)$ for all $\beta$ as $|Y_i| \rightarrow \infty$. Requires error models with influence functions that go to zero such as the Student $t$ (O’Hagan, 1979)
Choice of df

- Score function for $t$ with $\alpha$ degrees of freedom has turning points at $\pm \sqrt{\alpha}$
Choice of df

- Score function for $t$ with $\alpha$ degrees of freedom has turning points at $\pm \sqrt{\alpha}$

- $g'(\epsilon)$ is negative when $\epsilon^2 > \alpha$ (standardized errors)
Choice of df

- Score function for $t$ with $\alpha$ degrees of freedom has turning points at $\pm \sqrt{\alpha}$

- $g'(\epsilon)$ is negative when $\epsilon^2 > \alpha$ (standardized errors)
- Contribution of observation to information matrix is negative and the observation is doubtful
Choice of df

- Score function for $t$ with $\alpha$ degrees of freedom has turning points at $\pm \sqrt{\alpha}$

- $g'(\epsilon)$ is negative when $\epsilon^2 > \alpha$ (standardized errors)
- Contribution of observation to information matrix is negative and the observation is doubtful
- Suggest taking $\alpha = 8$ or $\alpha = 9$ to reject errors larger than $\sqrt{8}$ or 3 sd.
Choice of df

- Score function for $t$ with $\alpha$ degrees of freedom has turning points at $\pm \sqrt{\alpha}$

$g'(\epsilon)$ is negative when $\epsilon^2 > \alpha$ (standardized errors)

- Contribution of observation to information matrix is negative and the observation is doubtful
- Suggest taking $\alpha = 8$ or $\alpha = 9$ to reject errors larger than $\sqrt{8}$ or 3 sd.
Scale-Mixtures of Normal Representation

\[ Z_i \overset{\text{iid}}{\sim} t(\nu, 0, \sigma^2) \iff \]

\[ \lambda_i \overset{\text{iid}}{\sim} \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \]
Scale-Mixtures of Normal Representation

\[ Z_i \overset{\text{iid}}{\sim} t(\nu, 0, \sigma^2) \iff \]

\[ Z_i \mid \lambda_i \overset{\text{ind}}{\sim} N(0, \sigma^2 / \lambda_i) \]
Scale-Mixtures of Normal Representation

\[ Z_i \overset{iid}{\sim} t(\nu, 0, \sigma^2) \iff \]

\[ Z_i \mid \lambda_i \overset{\text{ind}}{\sim} N(0, \sigma^2 / \lambda_i) \]
\[ \lambda_i \overset{iid}{\sim} G(\nu/2, \nu/2) \]
Scale-Mixtures of Normal Representation

\[ Z_i \overset{iid}{\sim} t(\nu, 0, \sigma^2) \iff \]

\[ Z_i | \lambda_i \overset{\text{iid}}{\sim} N(0, \sigma^2 / \lambda_i) \]

\[ \lambda_i \overset{iid}{\sim} G(\nu/2, \nu/2) \]

Integrate out “latent” \( \lambda \)'s to obtain marginal distribution.
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \overset{\text{ind}}{\sim} N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \text{ ind} \sim N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \text{ iid} \sim G(\nu/2, \nu/2) \]
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \quad \text{ind} \quad \sim \quad N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \quad \text{iid} \quad \sim \quad G(\nu/2, \nu/2) \]

\[ p(\alpha, \beta, \phi) \quad \propto \quad 1/\phi \]
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \stackrel{\text{ind}}{\sim} N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \sim \text{G}(\nu/2, \nu/2) \]

\[ p(\alpha, \beta, \phi) \propto 1/\phi \]

Joint Posterior Distribution:
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \overset{\text{ind}}{\sim} N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \overset{\text{iid}}{\sim} G(\nu/2, \nu/2) \]

\[ p(\alpha, \beta, \phi) \propto 1/\phi \]

Joint Posterior Distribution:

\[ p((\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \propto \phi^{n/2} \exp \left\{ -\frac{\phi}{2} \sum \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \]

\[ \ldots \]
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \quad \text{ind} \sim N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \quad \text{iid} \sim G(\nu/2, \nu/2) \]

\[ p(\alpha, \beta, \phi) \propto 1/\phi \]

Joint Posterior Distribution:

\[ p((\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \propto \phi^{n/2} \exp \left\{ -\frac{\phi}{2} \sum \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \phi^{-1} \]
Latent Variable Model

\[ Y_i \mid \alpha, \beta, \phi, \lambda \overset{\text{ind}}{\sim} N(\alpha + \beta x_i, \frac{1}{\phi \lambda_i}) \]

\[ \lambda_i \overset{\text{iid}}{\sim} G(\nu/2, \nu/2) \]

\[ p(\alpha, \beta, \phi) \propto 1/\phi \]

Joint Posterior Distribution:

\[ p((\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \propto \phi^{n/2} \exp \left\{ -\frac{\phi}{2} \sum \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \phi^{-1} \prod_{i=1}^{n} \lambda_i^{\nu/2-1} \exp(-\lambda_i \nu/2) \]

JAGS

Just Another Gibbs Sampler (and more)

- Model
Just Another Gibbs Sampler (and more)

- Model
- Data
Just Another Gibbs Sampler (and more)

- Model
- Data
- Initial values (optional)
Just Another Gibbs Sampler (and more)

- Model
- Data
- Initial values (optional)

May do this through ordinary text files or use the functions in R2jags to specify model, data, and initial values then call jags.
Model Specification via R2jags

```
rr.model = function() {
    for (i in 1:n) {
        mu[i] <- alpha0 + alpha1*(X[i] - Xbar)
        lambda[i] ~ dgamma(9/2, 9/2)
        prec[i] <- phi*lambda[i]
        Y[i] ~ dnorm(mu[i], prec[i])
    }
    phi ~ dgamma(1.0E-6, 1.0E-6)
    alpha0 ~ dnorm(0, 1.0E-6)
    alpha1 ~ dnorm(0, 1.0E-6)
}
```
Notes on Models

- Distributions of stochastic “nodes” are specified using ∼
Notes on Models

- Distributions of stochastic “nodes” are specified using $\sim$
- Assignment of deterministic “nodes” uses <- (NOT =)

JAGS allows expressions as arguments in distributions
Normal distributions are parameterized using precisions, so $\text{dnorm}(0, 1.0E-6)$ is a $N(0, 1.0 \times 10^{-6})$

uses for loop structure as in R for model description but coded in C++ so is fast!
Notes on Models

- Distributions of stochastic “nodes” are specified using $\sim$
- Assignment of deterministic “nodes” uses $\leftarrow$ (NOT $\equiv$)
- JAGS allows expressions as arguments in distributions
Notes on Models

- Distributions of stochastic “nodes” are specified using $\sim$
- Assignment of deterministic “nodes” uses <- (NOT =)
- JAGS allows expressions as arguments in distributions
- Normal distributions are parameterized using *precisions*, so `dnorm(0, 1.0E-6)` is a $N(0, 1.0 \times 10^6)$
Notes on Models

- Distributions of stochastic “nodes” are specified using $\sim$
- Assignment of deterministic “nodes” uses $\leftarrow$ (NOT $\equiv$)
- JAGS allows expressions as arguments in distributions
- Normal distributions are parameterized using *precisions*, so
  \[ \text{dnorm}(0, \ 1.0E-6) \] is a $N(0, 1.0 \times 10^6)$
- Uses for loop structure as in R for model description but coded in C++ so is fast!
Data

A list or rectangular data structure for all data and summaries of data used in the model

```r
bf.data = list(Y = bodyfat$Bodyfat,
               X=bodyfat$Abdomen)
bf.data$n = length(bf.data$Y)
bf.data$Xbar = mean(bf.data$X)
```
Specifying which Parameters to Save

The parameters to be monitored and returned to R are specified with the variable `parameters`

```r
parameters = c("beta0", "beta1", "sigma",
               "mu34", "y34", "lambda[39]"
)
```

- All of the above (except lambda) are calculated from the other parameters. (See R-code for definitions of these parameters.)
Specifying which Parameters to Save

The parameters to be monitored and returned to R are specified with the variable `parameters`:

```r
parameters = c("beta0", "beta1", "sigma", "mu34", "y34", "lambda[39]")
```

- All of the above (except lambda) are calculated from the other parameters. (See R-code for definitions of these parameters.)
- `lambda[39]` saves only the 39th case of $\lambda$.
Specifying which Parameters to Save

The parameters to be monitored and returned to R are specified with the variable `parameters`

```
parameters = c("beta0", "beta1", "sigma",
               "mu34", "y34", "lambda[39]"
)
```

- All of the above (except lambda) are calculated from the other parameters. (See R-code for definitions of these parameters.)
- `lambda[39]` saves only the 39th case of $\lambda$
- To save a whole vector (for example all lambdas, just give the vector name)
Specifying which Parameters to Save

The parameters to be monitored and returned to R are specified with the variable `parameters`

```r
parameters = c("beta0", "beta1", "sigma", "mu34", "y34", "lambda[39]")
```

- All of the above (except lambda) are calculated from the other parameters. (See R-code for definitions of these parameters.)
- `lambda[39]` saves only the 39th case of \( \lambda \)
- To save a whole vector (for example all lambdas, just give the vector name)
bf.sim = jags(bf.data, inits=NULL, par=parameters,
model=rr.model,
n.chains=2, n.iter=10000,
)
<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>50%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta0</td>
<td>-41.70</td>
<td>2.75</td>
<td>-46.91</td>
<td>-41.67</td>
<td>-36.40</td>
</tr>
<tr>
<td>beta1</td>
<td>0.66</td>
<td>0.03</td>
<td>0.60</td>
<td>0.66</td>
<td>0.71</td>
</tr>
<tr>
<td>sigma</td>
<td>4.48</td>
<td>0.23</td>
<td>4.05</td>
<td>4.46</td>
<td>4.96</td>
</tr>
<tr>
<td>mu34</td>
<td>15.10</td>
<td>0.35</td>
<td>14.43</td>
<td>15.10</td>
<td>15.82</td>
</tr>
<tr>
<td>y34</td>
<td>14.94</td>
<td>5.15</td>
<td>4.37</td>
<td>15.21</td>
<td>24.65</td>
</tr>
<tr>
<td>lambda[39]</td>
<td>0.33</td>
<td>0.16</td>
<td>0.11</td>
<td>0.30</td>
<td>0.72</td>
</tr>
</tbody>
</table>

95% HPD interval for expected bodyfat (14.5, 15.8)
95% HPD interval for bodyfat (5.1, 25.3)
Comparison

- $95\%$ Probability Interval for $\beta$ is $(0.60, 0.71)$ with $t_9$ errors
Comparison

- 95% Probability Interval for $\beta$ is $(0.60, 0.71)$ with $t_9$ errors
- 95% Confidence Interval for $\beta$ is $(0.58, 0.69)$ (all data normal model)
Comparison

- 95% Probability Interval for $\beta$ is $(0.60, 0.71)$ with $t_9$ errors
- 95% Confidence Interval for $\beta$ is $(0.58, 0.69)$ (all data normal model)
- 95% Confidence Interval for $\beta$ is $(0.61, 0.73)$ (normal model without case 39)
Comparison

- 95% Probability Interval for $\beta$ is $(0.60, 0.71)$ with $t_9$ errors
- 95% Confidence Interval for $\beta$ is $(0.58, 0.69)$ (all data normal model)
- 95% Confidence Interval for $\beta$ is $(0.61, 0.73)$ (normal model without case 39)

Results intermediate without having to remove any observations
Comparison

- 95% Probability Interval for $\beta$ is $(0.60, 0.71)$ with $t_9$ errors
- 95% Confidence Interval for $\beta$ is $(0.58, 0.69)$ (all data normal model)
- 95% Confidence Interval for $\beta$ is $(0.61, 0.73)$ (normal model without case 39)

Results intermediate without having to remove any observations
Case 39 down weighted by $\lambda_{39}$
Full Conditional for $\lambda_j$

$$p(\lambda_j \mid \text{rest}, Y) \propto p(\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y)$$
Full Conditional for $\lambda_j$

\[
p(\lambda_j \mid \text{rest}, Y) \propto p(\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \\
\propto \phi^{n/2-1} \prod_{i=1}^{n} \exp \left\{ -\frac{\phi}{2} \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \\
\prod_{i=1}^{n} \lambda_i^{\frac{\nu+1}{2}-1} \exp(-\lambda_i \frac{\nu}{2})
\]
Full Conditional for $\lambda_j$

\[ p(\lambda_j \mid \text{rest}, Y) \propto p(\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \]

\[ \propto \phi^{n/2-1} \prod_{i=1}^{n} \exp \left\{ -\frac{\phi}{2} \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \]

\[ \prod_{i=1}^{n} \lambda_i^{\nu+1-1} \exp(-\lambda_i \nu) \]

Ignore all terms except those that involve $\lambda_j$
Full Conditional for $\lambda_j$

\[
p(\lambda_j \mid \text{rest}, Y) \propto p(\alpha, \beta, \phi, \lambda_1, \ldots, \lambda_n \mid Y) \\
\propto \phi^{n/2-1} \prod_{i=1}^{n} \exp \left\{ -\frac{\phi}{2} \lambda_i (y_i - \alpha - \beta x_i)^2 \right\} \times \\
\prod_{i=1}^{n} \lambda_i^{\nu+1 - \frac{1}{2}} \exp(-\lambda_i \frac{\nu}{2})
\]

Ignore all terms except those that involve $\lambda_j$

\[
\lambda_j \mid \text{rest}, Y \sim G \left( \frac{\nu + 1}{2}, \frac{\phi (y_j - \alpha - \beta x_j)^2 + \nu}{2} \right)
\]
Weights

Under prior $E[\lambda_i] = 1$
Weights

Under prior $E[\lambda_i] = 1$
Under posterior, large residuals are down-weighted (approximately those bigger than $\sqrt{\nu}$)
As a general recommendation, the prior distribution should have “heavier” tails than the likelihood.
Prior Distributions on Parameter

As a general recommendation, the prior distribution should have “heavier” tails than the likelihood

- with \( t_9 \) errors use a \( t_\alpha \) with \( \alpha < 9 \)
Prior Distributions on Parameter

As a general recommendation, the prior distribution should have “heavier” tails than the likelihood

- with $t_9$ errors use a $t_\alpha$ with $\alpha < 9$
- also represent via scale mixture of normals
Prior Distributions on Parameter

As a general recommendation, the prior distribution should have “heavier” tails than the likelihood

- with $t_9$ errors use a $t_\alpha$ with $\alpha < 9$
- also represent via scale mixture of normals
- Horseshoe, Double Pareto, Cauchy all have heavier tails
Prior Distributions on Parameter

As a general recommendation, the prior distribution should have “heavier” tails than the likelihood

- with $t_9$ errors use a $t_\alpha$ with $\alpha < 9$
- also represent via scale mixture of normals
- Horseshoe, Double Pareto, Cauchy all have heavier tails
- See Stack-loss code
Prior Distributions on Parameter

As a general recommendation, the prior distribution should have “heavier” tails than the likelihood

- with $t_9$ errors use a $t_\alpha$ with $\alpha < 9$
- also represent via scale mixture of normals
- Horseshoe, Double Pareto, Cauchy all have heavier tails
- See Stack-loss code