An introduction to Bayesian modeling using R and JAGS

Instructors
Kent Holsinger
Xiaojing Wang

University of Connecticut

22 July 2015

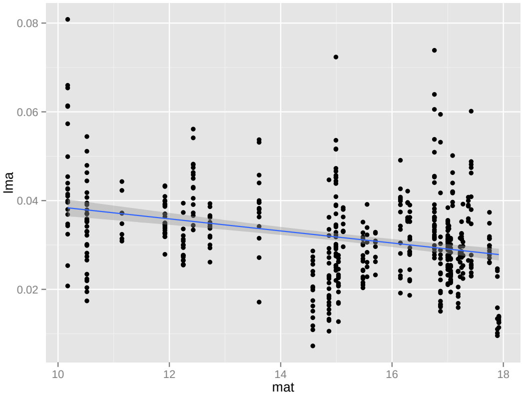
Overview of the Workshop

- Focus on using R and JAGS for Bayesian analysis
- Linear regression including mixed modeling – Kent Holsinger
  - Simple linear regression
  - Multiple regression (including random effects)
- Multicollinearity – Xiaojing Wang
  - Hierarchical independent prior distributions
  - Variable selection

Linear Regression

One of the most common statistical procedures in ecology and evolution. For example,

- Data on LMA from 535 individuals in the genus *Protea* (42 species, 48 sites, 142 unique site/species combinations)
- Data on mean annual temperature for each of those sites



Linear Regression

In R

```
> summary(lm(lma ~ mat, data=tmp))

Call:
lm(formula = lma ~ mat, data = tmp)

Residuals:
      Min        1Q    Median        3Q       Max
-0.025126 -0.005781 -0.000785  0.004647  0.044444

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0521895  0.0027116  19.246  < 2e-16 ***
mat         -0.0013587  0.0001785  -7.611 1.24e-13 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.009815 on 533 degrees of freedom
Multiple R-squared:  0.09803, Adjusted R-squared:  0.09634
F-statistic: 57.93 on 1 and 533 DF,  p-value: 1.241e-13
```

Linear Regression

Remember basic assumptions of simple linear regression

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\
\epsilon_i &\sim \mathrm{N}(0, \sigma^2)
\end{aligned}
$$

Here's another way to write that

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma^2) \\
\mu_i &= \beta_0 + \beta_1 x_i
\end{aligned}
$$

The second way of writing the model will be more convenient for us, so that's the approach we'll use.

Statistical Analysis

Statistical inference is the process of learning about the general characteristics of a population from a sample.

- Characteristics often expressed in terms of parameters θ.
- Measurements on the subset of members given by numerical values Y.
- Before the data are observed, both Y and θ are unknown.
- A probability model is assumed for observed data if we knew θ is the truth.
- What if we have prior information about θ?

Bayesian Inference

Bayesian inference allows us to update prior beliefs with the observed data to quantify uncertainty about θ.

- Prior Distribution: $p(\theta)$
- Sampling Model (likelihood): $p(y \mid \theta)$
- Posterior Distribution

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

- Calculating $p(y)$ is typically very challenging. Use MCMC (implemented in JAGS) to estimate $p(\theta \mid y)$.

Metropolis-Hastings Algorithm

For θ_j

- Propose a new $\theta_j^* \sim q(\theta_j^* | \theta_j^{(t)})$
- Calculate Metropolis-Hastings ratio

$$\alpha = \frac{p(Y|\theta_j^*)p(\theta_j^*)/q(\theta_j^*|\theta_j^{(t)})}{p(Y|\theta_j^{(t)})p(\theta_j^{(t)})/q(\theta_j^{(t)}|\theta_j^*)}$$

- if $\alpha < 1$ set

$$\theta_j^{(t+1)} = \begin{cases} \theta_j^* & \text{with probability } \alpha \\ \theta_j^{(t)} & \text{with probability } 1-\alpha \end{cases}$$

  If $\alpha > 1$ set $\theta_j^{(t+1)} = \theta_j^*$
- Repeat for $j = 1, \ldots, J$
- Repeat for $t = 1, \ldots, T$

Linear Regression - as a Bayesian

We start with the sampling model $p(y \mid \theta)$, where $\theta = (\beta_0, \beta_1, \sigma^2)'$,

$$\begin{aligned} y_i &\sim \mathrm{N}(\mu_i, \sigma^2) \ , \\ \mu_i &= \beta_0 + \beta_1 x_i \ , \end{aligned}$$

and x_i is the value of the covariate in individual i. Then we add prior distributions $p(\theta)$,

$$\begin{aligned} \beta_0 &\sim \mathrm{N}(0, \tau^{-1}), \\ \beta_1 &\sim \mathrm{N}(0, \tau^{-1}), \\ \sigma^2 &= \frac{1}{\tau_{resid}}, \\ \tau_{resid} &\sim \mathrm{Gamma}(1, \phi) \end{aligned}$$

Linear Regression - in R+JAGS[1]

- Rescale all variables to mean of 0, standard deviation of 1

```
Inference for Bugs model at "simple-linear-regression.jags", fit using jags,
 5 chains, each with 10000 iterations (first 5000 discarded), n.thin = 5
 n.sims = 5000 iterations saved
             mu.vect sd.vect   2.5%    25%    50%    75%  97.5% Rhat n.eff
beta.0         0.000   0.041 -0.082 -0.028  0.000  0.028  0.080 1.001  5000
beta.mat      -0.313   0.041 -0.393 -0.341 -0.314 -0.285 -0.231 1.002  1900
sigma.resid    0.953   0.029  0.897  0.933  0.952  0.972  1.012 1.001  3700
```

- Compare with lm() results from R

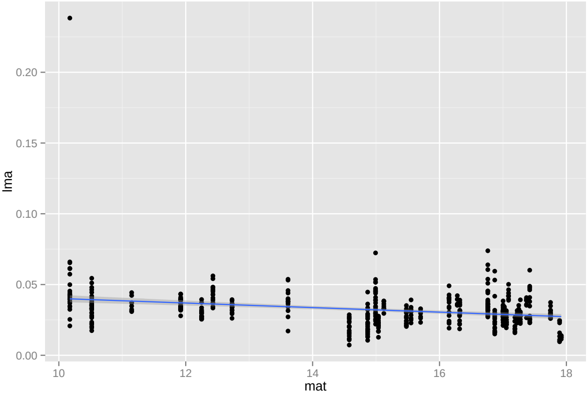
```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.614e-17  4.110e-02   0.000        1
mat         -3.131e-01  4.114e-02  -7.611 1.24e-13 ***

Residual standard error: 0.9506 on 533 degrees of freedom
```

[1] simple-linear-regression.jags

Robust Regression

- What do we do with outliers?



- Could delete the observation if we have good evidence that it's an error.
- Probably better to use an approach that leaves the data in and gives less weight to outliers.

Robust Regression[3]

- Use a Cauchy distribution ("fat tails") instead of normal for prior on regression coefficients and residual error.[2]

		Coefficient estimates	
Model	Data set	β_0	β_{MAT}
Simple	Original	0.000	-0.313
	Modified	0.028	-0.372
Robust	Original	-0.078	-0.294
	Modified	-0.079	-0.293

- Estimates not affected by extreme outlier when using Cauchy prior.

[2] t distribution with 1 degree of freedom in JAGS

[3] robust-linear-regression.jags

Multiple Linear Regression

Simple generalization of what we've already seen

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma^2) \ , \\
\mu_i &= \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} \ ,
\end{aligned}
$$

where x_{ik} is the value of the kth covariate in individual i. The priors are

$$
\begin{aligned}
\beta_k &\sim \mathrm{N}(0, \tau^{-1}), \quad k = 0, \ldots, K \ , \\
\sigma^2 &= \frac{1}{\tau_{resid}} \ , \\
\tau_{resid} &\sim \mathrm{Gamma}(1, \phi).
\end{aligned}
$$

Multiple Linear Regression[4]

From JAGS

```
Inference for Bugs model at "multiple-linear-regression.jags", fit using jags,
 5 chains, each with 10000 iterations (first 5000 discarded), n.thin = 5
 n.sims = 5000 iterations saved
             mu.vect sd.vect   2.5%    25%    50%    75%  97.5%  Rhat n.eff
beta.0         0.000   0.040 -0.078 -0.027  0.000  0.028  0.078 1.001  4800
beta.cdd       0.106   0.073 -0.036  0.058  0.105  0.154  0.251 1.001  5000
beta.elev     -0.319   0.094 -0.500 -0.384 -0.319 -0.256 -0.136 1.001  5000
beta.inso      0.054   0.053 -0.048  0.018  0.054  0.091  0.157 1.001  5000
beta.map      -0.022   0.078 -0.178 -0.074 -0.020  0.031  0.129 1.001  5000
beta.mat      -0.463   0.114 -0.688 -0.541 -0.461 -0.385 -0.243 1.001  5000
beta.ratio    -0.016   0.079 -0.172 -0.069 -0.013  0.039  0.134 1.001  5000
sigma.resid    0.939   0.029  0.884  0.919  0.939  0.958  1.000 1.001  5000
```

Compare to lm() from R

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.520e-17  4.052e-02   0.000 1.000000
cdd          1.051e-01  7.315e-02   1.437 0.151291
elev        -3.184e-01  9.534e-02  -3.339 0.000899 ***
inso         5.288e-02  5.287e-02   1.000 0.317702
map         -2.309e-02  7.749e-02  -0.298 0.765836
mat         -4.629e-01  1.147e-01  -4.037  6.2e-05 ***
ratio       -1.481e-02  7.944e-02  -0.186 0.852218

Residual standard error: 0.9373 on 528 degrees of freedom
```

[4] multiple-linear-regression.jags

Multiple Linear Regression with Species Random Effect

$\gamma_i^{(s)}$ denotes the mean for species s to which inidividual i belongs

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma^2_{resid}) \ , \\
\mu_i &= \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \gamma_i^{(s)} \ , \\
\sigma^2_{resid} &= \frac{1}{\tau_{resid}} \ , \\
\tau_{resid} &\sim \mathrm{Gamma}(1, \phi) \ , \\
\gamma_i^{(s)} &\sim \mathrm{N}(0, \sigma^2_{species}) \ , \\
\sigma^2_{species} &= \frac{1}{\tau_{species}} \ , \\
\tau_{species} &\sim \mathrm{Gamma}(1, \phi), \\
\beta_k &\sim \mathrm{N}(0, \tau^{-1}), \quad k = 0, \ldots, K \ .
\end{aligned}
$$

Multiple Linear Regression with Species Random Effect

Alternatively

$$
\begin{aligned}
y_i &\sim \mathrm{N}(\mu_i, \sigma^2_{resid}) \ , \\
\mu_i &= \beta^{(s)}_{0i} + \sum_{k=1}^{K} \beta_k x_{ik} \ , \\
\sigma^2_{resid} &= \frac{1}{\tau_{resid}} \ , \\
\tau_{resid} &\sim \mathrm{Gamma}(1, \phi) \ , \\
\beta^{(s)}_{0i} &\sim \mathrm{N}(\beta_0, \sigma^2_{species}) \ , \\
\sigma^2_{species} &= \frac{1}{\tau_{species}} \ , \\
\tau_{species} &\sim \mathrm{Gamma}(1, \phi) \ , \\
\beta_k &\sim \mathrm{N}(0, \tau^{-1}), \quad k = 0, \ldots, K \ .
\end{aligned}
$$

Multiple Linear Regression with Species Random Effect[5]

From JAGS

```
beta.cdd        0.081   0.055  -0.024   0.045   0.082   0.118   0.188 1.001  3100
beta.elev      -0.201   0.108  -0.408  -0.274  -0.201  -0.128   0.010 1.002  2400
beta.inso      -0.073   0.058  -0.187  -0.112  -0.073  -0.034   0.040 1.001  3400
beta.map       -0.418   0.083  -0.581  -0.474  -0.419  -0.362  -0.257 1.001  5000
beta.mat        0.093   0.110  -0.117   0.020   0.092   0.169   0.307 1.001  5000
beta.ratio      0.427   0.078   0.278   0.374   0.427   0.478   0.578 1.001  4800
beta.zero      -0.064   0.154  -0.369  -0.167  -0.062   0.037   0.241 1.001  5000
sigma.resid     0.545   0.023   0.511   0.532   0.544   0.556   0.581 1.001  3100
sigma.species   0.966   0.117   0.767   0.884   0.960   1.036   1.218 1.001  5000
```

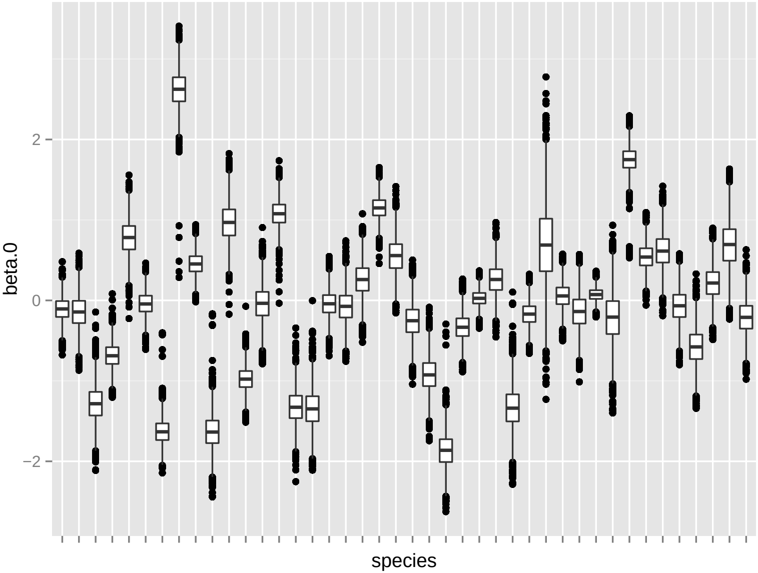
Compare to lmer() From R

```
Random effects:
 Groups   Name        Variance Std.Dev.
 species  (Intercept) 0.8951   0.9461
 Residual             0.2924   0.5408
Number of obs: 535, groups:  species, 42

Fixed effects:
            Estimate Std. Error t value
(Intercept) -0.06289    0.14898  -0.422
cdd          0.07941    0.05576   1.424
elev        -0.19835    0.10946  -1.812
inso        -0.07259    0.05791  -1.254
map         -0.42009    0.08176  -5.138
mat          0.09811    0.10893   0.901
ratio        0.43022    0.07776   5.533
```

[5] random-effect-multiple-regression.jags

Multiple Linear Regression with Species Random Effect



Problems with Multicollinearity

- Variables may appear to be unimportant (when they are).
- Coefficient estimates are unstable and hard to interpret (can estimate combinations of coefficients but not individual coefficients).

Alternative Bayesian solutions:

- Independent Prior Distributions
- Variable Selection

Hierarichal Model with Independent Priors

Hierarchical Model:

$$\begin{aligned}
\beta_j | \lambda_j, \sigma^2 &\sim \mathsf{N}(0, \sigma^2/\lambda_j) \\
\lambda_j | \sigma^2 &\sim \mathsf{Gamma}(1/2, 1/2) \\
1/\sigma^2 &\sim \mathsf{Gamma}(\nu_0/2, \nu_0 \sigma_0^2/2)
\end{aligned}$$

- Leads to nice conjugate updates for all full conditionals
- Easy to code in JAGS
- Allows each parameter to have own precision with mean 1

Cauchy Prior

First two equations imply that $\beta_j|\sigma^2 \sim \text{Cauchy}(0, \sigma^2)$

$$p(\beta) = \frac{1}{\pi\sigma}\left(1 + \frac{\beta^2}{\sigma^2}\right)^{-1},$$

leading to a collapsed model

$$\begin{aligned} Y|\boldsymbol{\beta}, \sigma^2 &\sim \text{N}(X\boldsymbol{\beta}, \sigma^2 I_n) \ , \\ \beta_j|\sigma^2 &\sim \text{Cauchy}(0, \sigma^2) \ , \\ 1/\sigma^2 &\sim \text{Gamma}(\nu_0/2, \nu_0\sigma_0^2/2) \ . \end{aligned}$$

No nice full conditional for β_j.

Cauchy Prior[6]

Independent $N(0,1)$

beta.cdd	0.081	0.055	-0.024	0.045	0.082	0.118	0.188	1.001	3100
beta.elev	-0.201	0.108	-0.408	-0.274	-0.201	-0.128	0.010	1.002	2400
beta.inso	-0.073	0.058	-0.187	-0.112	-0.073	-0.034	0.040	1.001	3400
beta.map	-0.418	0.083	-0.581	-0.474	-0.419	-0.362	-0.257	1.001	5000
beta.mat	0.093	0.110	-0.117	0.020	0.092	0.169	0.307	1.001	5000
beta.ratio	0.427	0.078	0.278	0.374	0.427	0.478	0.578	1.001	4800
beta.zero	-0.064	0.154	-0.369	-0.167	-0.062	0.037	0.241	1.001	5000
sigma.resid	0.545	0.023	0.511	0.532	0.544	0.556	0.581	1.001	3100
sigma.species	0.966	0.117	0.767	0.884	0.960	1.036	1.218	1.001	5000

Independent hierarchical

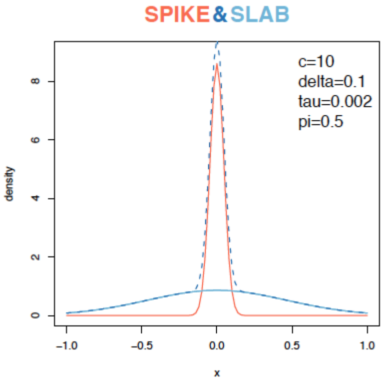
beta.cdd	0.084	0.055	-0.026	0.048	0.085	0.122	0.188	1.003	1300
beta.elev	-0.191	0.104	-0.394	-0.262	-0.190	-0.120	0.009	1.004	800
beta.inso	-0.077	0.059	-0.192	-0.117	-0.077	-0.038	0.039	1.001	5000
beta.map	-0.400	0.082	-0.562	-0.455	-0.401	-0.344	-0.240	1.004	870
beta.mat	0.084	0.107	-0.129	0.013	0.084	0.157	0.288	1.009	370
beta.ratio	0.416	0.079	0.262	0.363	0.417	0.469	0.570	1.007	510
beta.zero	-0.060	0.142	-0.338	-0.154	-0.061	0.035	0.224	1.002	2700
sigma.resid	0.545	0.018	0.512	0.533	0.544	0.556	0.580	1.001	5000
sigma.species	0.958	0.114	0.767	0.876	0.947	1.028	1.211	1.001	5000

[6] hierarchical-prior-multiple-regression.jags

Stochastic Search Variable Selection

The Spike-and-Slab prior:

$$\begin{aligned}
\beta_j \mid \gamma_j, c, \tau_j^{-1} &\sim (1-\gamma_j)\mathcal{N}(0, \tau_j^{-1}) + \gamma_j \mathcal{N}(0, \tau_j^{-1} c^2) \ , \\
\gamma_j \mid \pi_j &\sim \mathit{Bernoulli}(\pi_j) \ .
\end{aligned}$$



- $\gamma_j = 0$: Variable not in the model;
- $\gamma_j = 1$: Variable in the model;
- Calibration of hyper-parameters c, τ_j^{-1} needed.

Inference for Variable Selection

- Highest posterior model (HPM): Select a model that has been visited most often.
- Median probability model (MPM): Select variables that appear at least in 50% of visited models.

Alternative spike and slab models

- Popular approach in genomic research;
- Variants:
  - Conjugate version:

$$\beta_j \mid \gamma_j, c, \tau_j^{-1} \sim (1-\gamma_j)\mathcal{N}(0, \sigma^2\tau_j^{-1}) + \gamma_j\mathcal{N}(0, \sigma^2\tau_j^{-1}c^2).$$

  - Replace the spike normal in Spike-and-Slab prior by Dirac, i.e.,

$$\beta_j \mid \gamma_j, \tau_j^{-1} \sim (1-\gamma_j)\delta_0 + \gamma_j\mathcal{N}(0, \tau_j^{-1}).$$

Variable selection – Dirac + Normal[7]

Independent N(0, 1)

beta.cdd	0.081	0.055	-0.024	0.045	0.082	0.118	0.188	1.001	3100
beta.elev	-0.201	0.108	-0.408	-0.274	-0.201	-0.128	0.010	1.002	2400
beta.inso	-0.073	0.058	-0.187	-0.112	-0.073	-0.034	0.040	1.001	3400
beta.map	-0.418	0.083	-0.581	-0.474	-0.419	-0.362	-0.257	1.001	5000
beta.mat	0.093	0.110	-0.117	0.020	0.092	0.169	0.307	1.001	5000
beta.ratio	0.427	0.078	0.278	0.374	0.427	0.478	0.578	1.001	4800
beta.zero	-0.064	0.154	-0.369	-0.167	-0.062	0.037	0.241	1.001	5000
sigma.resid	0.545	0.023	0.511	0.532	0.544	0.556	0.581	1.001	3100
sigma.species	0.966	0.117	0.767	0.884	0.960	1.036	1.218	1.001	5000

Dirac + N(0, 1)

beta.cdd	0.002	0.015	0.000	0.000	0.000	0.000	0.039	1.029	1100
beta.elev	-0.277	0.112	-0.457	-0.349	-0.292	-0.224	0.000	1.016	310
beta.inso	-0.003	0.019	-0.052	0.000	0.000	0.000	0.000	1.054	460
beta.map	-0.473	0.075	-0.609	-0.523	-0.478	-0.429	-0.306	1.012	440
beta.mat	0.015	0.060	0.000	0.000	0.000	0.000	0.240	1.009	1100
beta.ratio	0.378	0.057	0.282	0.343	0.374	0.407	0.516	1.002	1800
beta.zero	-0.061	0.151	-0.355	-0.160	-0.060	0.039	0.240	1.001	5000
sigma.resid	0.546	0.023	0.512	0.533	0.545	0.557	0.581	1.001	5000
sigma.species	0.948	0.116	0.750	0.868	0.939	1.019	1.193	1.002	2100

[7] dirac-plus-normal-multiple-regression.jags

Variable selection – Dirac + Normal

Dirac + $N(0,1)$

beta.cdd	0.002	0.015	0.000	0.000	0.000	0.000	0.039	1.029	1100
beta.elev	-0.277	0.112	-0.457	-0.349	-0.292	-0.224	0.000	1.016	310
beta.inso	-0.003	0.019	-0.052	0.000	0.000	0.000	0.000	1.054	460
beta.map	-0.473	0.075	-0.609	-0.523	-0.478	-0.429	-0.306	1.012	440
beta.mat	0.015	0.060	0.000	0.000	0.000	0.000	0.240	1.009	1100
beta.ratio	0.378	0.057	0.282	0.343	0.374	0.407	0.516	1.002	1800
beta.zero	-0.061	0.151	-0.355	-0.160	-0.060	0.039	0.240	1.001	5000
sigma.resid	0.546	0.023	0.512	0.533	0.545	0.557	0.581	1.001	5000
sigma.species	0.948	0.116	0.750	0.868	0.939	1.019	1.193	1.002	2100

Posterior conditioned on $\gamma_i > 0$

```
  beta.cdd:  0.04      0.057 (-0.042,  0.168)
 beta.elev:  0.93     -0.298 (-0.459, -0.134)*
 beta.inso:  0.05     -0.061 (-0.188,  0.059)
  beta.map:  1.00     -0.473 (-0.609, -0.306)*
  beta.mat:  0.11      0.142 (-0.111,  0.367)
beta.ratio:  1.00      0.379 ( 0.283,  0.516)*
```

Model Selection

Selection of a single model has the following problems

- When the criteria suggest that several models are equally good, what should we report? Still pick only one model?
- What do we report for our uncertainty after selecting a model?

Typical analysis ignores model uncertainty!

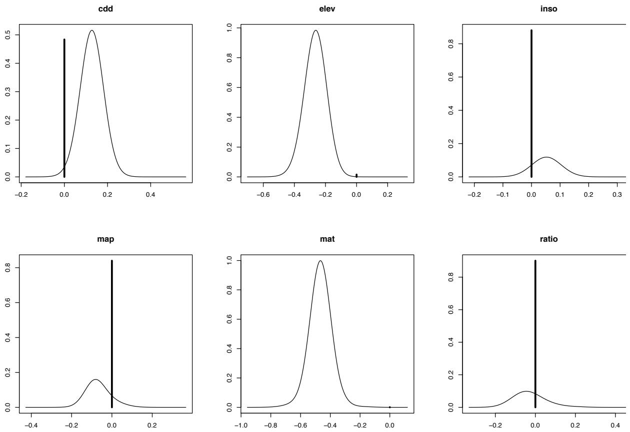
Bayesian Model Choice

- Models for the variable selection problem are based on a subset of the $x_1, \cdots, x_p$ variables.
- Encode models with a vector $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p)'$ where $\gamma_j \in \{0, 1\}$ is an indicator for whether variable x_j should be included in the model $M_{\boldsymbol{\gamma}}$. Notice $\gamma_j = 0 \Leftrightarrow \beta_j = 0$.
- Each value of $\boldsymbol{\gamma}$ represents one of the 2^p models.
- Under model $M_{\boldsymbol{\gamma}}$:

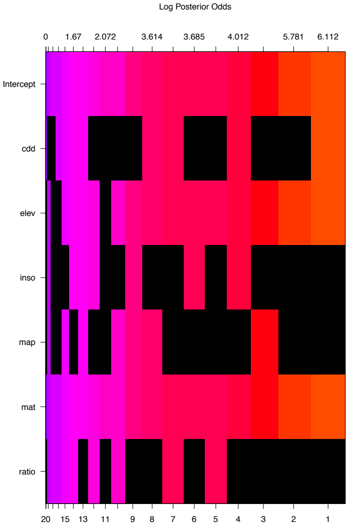
$$Y \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau \sim \mathcal{N}(X_{\boldsymbol{\gamma}} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \tau^{-1} \mathbf{I})$$

  where $X_{\boldsymbol{\gamma}}$ is the design matrix using the columns in X where $\gamma_j = 1$ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is the subset of $\boldsymbol{\beta}$ that are non-zero.

Bayesian Model Model Choice (*Protea* data)



Bayesian Model Model Choice (*Protea* data)



Bayesian Model Averaging

Rather than use a single model, BMA uses all (or potentially a lot) models, but weights model predictions by their posterior probabilities (measure of how much each model is supported by the data).

- Posterior model probabilities

$$\mathrm{P}(M_j \mid Y) = \frac{\mathrm{P}(Y \mid M_j)\mathrm{P}(M_j)}{\sum_j \mathrm{P}(Y \mid M_j)\mathrm{P}(M_j)},$$

  Marginal likelihod of a model is

$$\mathrm{P}(Y \mid M_\gamma) = \int\int \mathrm{P}(Y \mid \boldsymbol{\beta}_\gamma, \tau)\mathrm{P}(\boldsymbol{\beta}_\gamma \mid \boldsymbol{\gamma}, \tau)\mathrm{P}(\tau \mid \boldsymbol{\gamma}) d\boldsymbol{\beta}_\gamma d\tau.$$

- Probability $\beta_j \neq 0$: $\sum_{M_j : \beta_j \neq 0} \mathrm{P}(M_j \mid Y)$.

Bayesian Model Averaging (Continued)

- Predictions

$$P(Y^{new} \mid Y) = \sum_j P(Y^{new} \mid Y, M_j) P(M_j \mid Y),$$

where

$$P(Y^{new} \mid Y, M_\gamma) = \int P(Y^{new} \mid Y, \boldsymbol{\beta}_\gamma, \tau) P(\boldsymbol{\beta}_\gamma, \tau \mid Y) d\boldsymbol{\beta}_\gamma d\tau.$$