Running Bayesian Hierarchical Regression Models

JEFF GILL
Department of Government
Department of Mathematics and Statistics
Center for Behavioral Neurosciences
American University

Two General Approaches to Model Checking, G&H (2007)

- There are many ways to do model checking, some of which we've already seen.
- Two general procedures:
  - checking the fit of a single model to the data,
  - comparing the model of interest to alternatives.
- The methods discussed in Chapter 24 (plus some extra discussed today) are generally not unique to multilevel models, but increase in importance as model complexity increases.
- Most procedures generate some form of test statistics that help us make claims.
- Simulation tools are critical.

Four General Approaches to Assessing Model Quality, Gill (2008)

- Posterior Predictive Checks: formally or informally comparing model predictions with observed outcomes.
- Sensitivity Analysis/Predictive Checks: the *informal* process of altering assumptions according to researcher intuition with the objective of determining the extent to which these changes modify the posterior distribution (global and local).
- Global Robustness Evaluation: *systematic* analysis of the degree to which posterior inferences are affected by both potential misspecification of the prior and influential data points.
- Local Robustness Evaluation: using differential calculus to determine the volatility of specific reported results.

More On Posterior Predictive Checks

- We are after "systematic differences between the model and observed data."
- Perhaps the most powerful tool is posterior predictive checks, which generate replicate datasets from the posterior predictive distribution for the parameters, and then compares to the observed dataset.
- We may also be interested in the posterior predictive distribution as well.
- Some test statistics are produced from posterior predictive procedures.
- MCMC output actually makes this process *easier*.

Predictive Checks, Example

- Consider a model of support for abortion under different scenarios using survey data from Britain in consecutive years from 1983 to 1986.

- The panel data for 264 respondents is collected annually by McGrath and Waterton (1986) where seven scenarios are provided and these respondents have the option of expressing support or disagreement for abortion.

- The full collection of these seven queries do not fall into an obvious ordinal scale, so we will treat them here as nominal and judge total support for abortion as a binomial test for each respondent at each wave of the panel

- The scenarios are: (1) the woman decides on her own that she does not wish to have the child, (2) the couple agree that they do not wish to have the child, (3) the woman is not married and does not wish to marry the man, (4) the couple cannot afford any more children, (5) there a strong chance that the baby has a biological defect, (6) the woman's health is seriously endangered by the pregnancy, and (7) the woman became pregnant as a result of rape.

Predictive Checks

- The model is specified for $i = 1, \ldots, 264$ respondents across $j = 1, \ldots, 4$ panel waves:

$$y_{ij} \sim \mathcal{BN}(n_i, p_{ij})$$

$$\text{logit}(p_{ij}) = \boldsymbol{\beta}_{0,j} + \boldsymbol{\beta}_{1,i} X_{1,i}$$

$$\boldsymbol{\beta}_{0,j} \sim \mathcal{N}(\mu_0, \tau_0) \qquad \boldsymbol{\beta}_{1,i} \sim \mathcal{N}(\mu_1, \tau_1)$$

$$\mu_0 \sim \mathcal{N}(0, 100) \qquad \mu_1 \sim \mathcal{N}(0, 100)$$

$$\tau_0 \propto \mathcal{G}(1, 0.1) \qquad \tau_1 \propto \mathcal{G}(1, 0.1)$$

- Here $X_{1,i}$ is the ith person's self-identified religion: (1) Catholic, (2) Protestant, (3) Other, and (4) No Religion, plus $n_i = 7$ at each wave for each person.

- $X_{1,i}$ is indexed by 1 to infer that more explanatory variables can be specified, and $n_i = 7$ is indexed by i so we could account for dropouts in a more general setting.

Model, Abortion Attitudes in Britain, Running

```
lapply(c("rjags","R2jags","arm","coda","superdiag","R2WinBUGS"),library,
  character.only=TRUE)
source("http://jeffgill.org/files/jeffgill/files/abortion.panel-data.list_.txt")
names(attitudes.list) <- c("N","P","r","n","x1")
attitudes.mod <- function() {
  for (i in 1:N) {
    for (j in 1:P) {
      logit(p[i,j]) <- b0[j] + b1[i]*x1[i]
      r[i,j] ~ dbin(p[i,j], n[i])
    }
    b1[i] ~ dnorm(mu1, nu1)
  }
  for (j in 1:P) {
    b0[j] ~ dnorm(mu0, nu0)
  }
  mu0 ~ dnorm(0.0,0.1)
  mu1 ~ dnorm(0.0,0.1)
  nu0 ~ dgamma(1,0.1)
  nu1 ~ dgamma(1,0.1)
}
```

Model, Abortion Attitudes in Britain, Running

```
attitudes.params <- c("b0","b1")
attitudes.out <- jags(data=attitudes.list, parameters.to.save=attitudes.params,
  n.iter=10000, model=attitudes.mod,n.burnin=2500, n.thin=1,n.chains=3)
|**************************************************| 100%
update(attitudes.out, n.iter=50000, n.burnin=0, n.thin=1,n.chains=3)
|**************************************************| 100%
    mu.vect sd.vect   2.5%    25%    50%    75%
b0[1]    0.060  0.208  -0.348  -0.081   0.060   0.201
b0[2]   -0.439  0.208  -0.847  -0.580  -0.440  -0.298
:
b1[263]    0.923 1.002  3700
b1[264]    1.149 1.002  1600
deviance 2957.939 1.003   930

attitudes.mcmc <- as.mcmc(attitudes.out)
attitudes.mat <- attitudes.mcmc[[1]]
means.vec <- apply(attitudes.mat,2,mean)
sd.vec <- apply(attitudes.mat,2,sd)
```

Model Summary, Abortion Attitudes in Britain

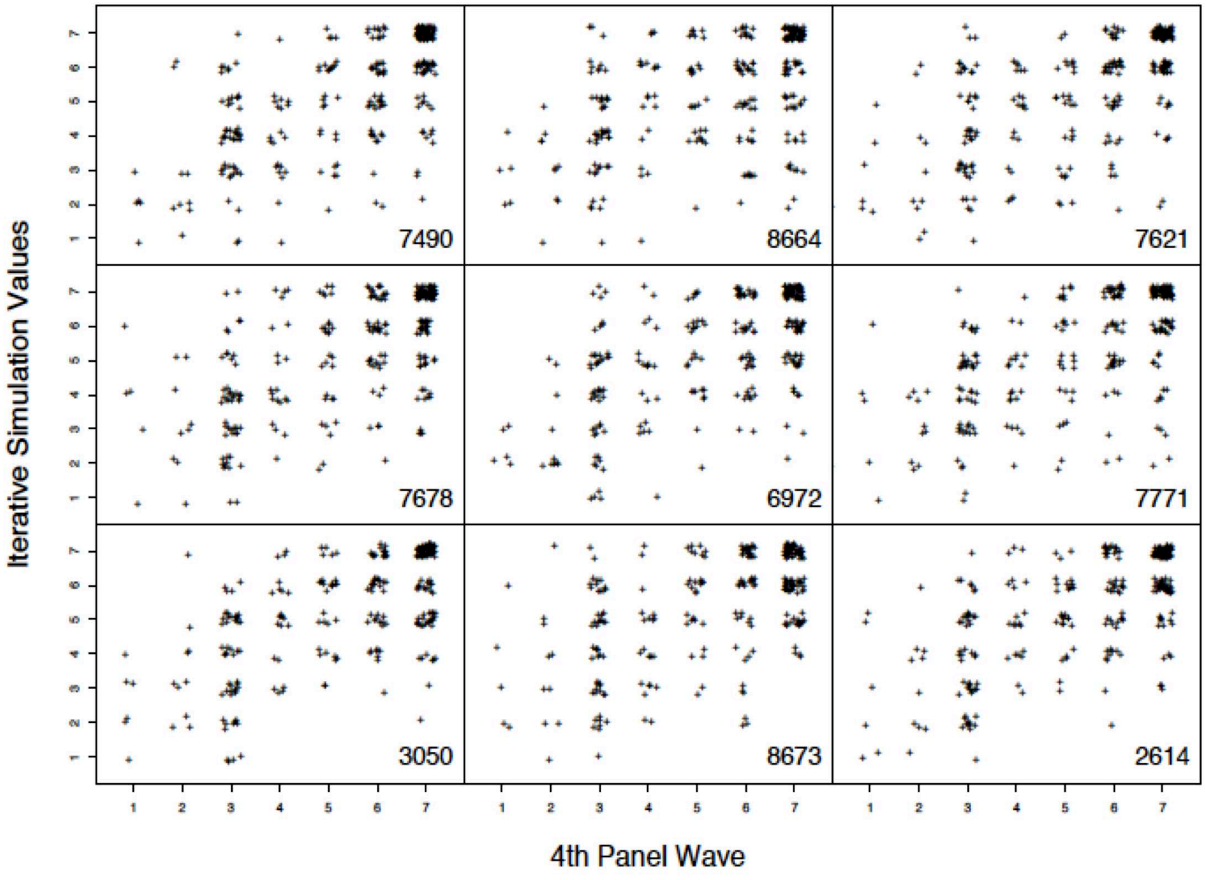
- Model Summary

	Posterior Quantiles				
	0.025	0.25	0.50	0.75	0.975
μ_0	-0.6708	-0.2563	-0.0294	0.1960	0.6245
τ_0	0.0283	0.0500	0.0641	0.0789	0.1113
μ_1	0.3029	0.4129	0.4728	0.5291	0.6336
τ_1	0.0303	0.3136	0.6726	1.1507	2.2453

- For posterior predictions, draw values from these posterior distribution MCMC samples, insert them into the model statement with data to produce $\hat{y}$ values.

- For example, pick the 4th (last) panel wave and 9 uniformly randomly chosen people for the 7 outcomes. . .

Outcome Comparison: Observed Versus Simulated



Outcome Comparison: Observed Versus Simulated

```
# SAVED HERE AS AN SPSS FILE, ALSO SEE THE TEACHING WEBPAGE FOR ASCII
library(foreign)
abortion <- read.spss("abortion_cluster.sav") # GET FROM DISTRIBUTED FILE
names(abortion)
[1] "religion"   "response1983" "response1984" "response1985" "response1986"

# THESE ARE CODA FILES, IF YOU RAN FROM R THEY ARE IN R AS WE'VE BEEN CREATING THEM
ab <- read.coda("CODAchain1.txt", "CODAindex.txt")
dim(ab)
[1] 10000  1060

N <- 264; P <- 4 # SUBJECTS AND WAVE

ab.index <- sample(1:nrow(ab),9)   # SAMPLE 9 RESPONDENTS
ab.index
[1] 4580 9198 9936 4893  194 6387 1547 5891 4858
```

Outcome Comparison: Observed Versus Simulated

```
par(mfrow=c(3,3),mar=c(0,0,0,0),oma=c(8,8,1,1))
for (k in 1:length(ab.index)) {
 p <- matrix(ab[ab.index[k],1:(N*P)],nrow=N,ncol=P,byrow=FALSE)
 r <- matrix(NA,nrow=N,ncol=P) # BOTH 264*4 MATRICES
 for (i in 1:N) { for (j in 1:P) { r[i,j] <- rbinom(n=1,size=7,prob=p[i,j]) } }
 plot(jitter(abortion$response1986),jitter(r[,4]),pch="+",xaxt="n",yaxt="n",
    xlab="", ylab="", xlim=c(0.5,7.5),ylim=c(0.5,7.5))
 text(7,1,ab.index[k],cex=2.0)
 if(k %% 3 == 1) { axis(2,labels=1:7,at=1:7) }
 if(k > 6) { axis(1,labels=1:7,at=1:7) }
}
mtext(side=2,line=4.5,outer=TRUE,cex=1.5,"Iterative Simulation Values")
mtext(side=1,line=4.5,outer=TRUE,cex=1.5,"4th Panel Wave")
```

Posterior Predictive Distribution

- First consider the Prior predictive distribution of a new data value, x_{new} before observing the full dataset:

$$p(x_{new}) = \int_{\Theta} p(x_{new}, \theta)d\theta = \int_{\Theta} p(x_{new}|\theta)p(\theta)d\theta.$$

- Note that this is the PDF of a data value times the prior distribution for θ.

- This is the marginal distribution of an unobserved data value is the product of the prior for θ and the single variable PDF or PMF, integrating out this parameter.

Posterior Predictive Distribution (cont.)

- Now consider the distribution of a new data point, x_{new} *after* the full iid data set, $\mathbf{x}$, has been observed:

$$p(x_{new}|\mathbf{x}) = \int_\Theta p(x_{new}, \theta|\mathbf{x})d\theta$$
$$= \int_\Theta \frac{p(x_{new}, \theta|\mathbf{x})}{p(\theta|\mathbf{x})} p(\theta|\mathbf{x})d\theta$$
$$= \int_\Theta p(x_{new}|\theta, \mathbf{x})p(\theta|\mathbf{x})d\theta.$$

This can be simplified since x_{new} and $\mathbf{x}$ are assumed independent given θ:

$$= \int_\Theta p(x_{new}|\theta)p(\theta|\mathbf{x})d\theta.$$

Posterior Predictive Distribution (cont.)

- Normal example: $X_1, X_2, \ldots, X_n$ are distributed iid $\mathcal{N}(\mu, \sigma_0^2)$, where σ_0^2 is known but μ is unknown. Place a normal prior on μ according to $\mu \sim \mathcal{N}(m, s^2)$.

- Posterior for μ:

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)\left(\mu - \frac{\left(\frac{m}{s^2}+\frac{n\bar{x}}{\sigma_0^2}\right)}{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}\right)^2\right].$$

- Re-expressed in terms of its mean and variance:

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2\right]$$

$$\text{where } \sigma_1^2 = \left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)^{-1} \qquad \mu_1 = \frac{\left(\frac{m}{s^2}+\frac{n\bar{x}}{\sigma_0^2}\right)}{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}.$$

Posterior Predictive Distribution (cont.)

- So the posterior predictive distribution is:

$$
\begin{aligned}
p(x_{new}|\mathbf{x}) &= \int_\mu p(x_{new}|\mu)p(\mu|\mathbf{x})d\mu \\
&\propto \int_\mu \exp\left[-\frac{1}{2}\left(\frac{(x_{new}-\mu)^2}{\sigma_0^2} + \frac{(\mu-\mu_1)^2}{\sigma_1^2}\right)\right] d\mu.
\end{aligned}
$$

- Summary statistics from this PPD:

$$
\begin{aligned}
E[x_{new}|\mathbf{x}] &= E[E(x_{new}|\mu,\mathbf{x})|\mathbf{x}] \\
&= E[\mu_1|\mathbf{x}] \\
&= E[\mu_1] = \mu. \\
\text{Var}[x_{new}|\mathbf{x}] &= E[\text{Var}(x_{new}|\mu,\mathbf{x})|\mathbf{x}] + \text{Var}[E(x_{new}|\mu,\mathbf{x})|\mathbf{x}] \\
&= E[\sigma_0^2|\mathbf{x}] + \text{Var}[\mu|\mathbf{x}] \\
&= \sigma_0^2 + \sigma_1^2/n.
\end{aligned}
$$

- For simple forms these analytical calculations lead to such easy closed-form solutions, but for more complex hierarchical regression forms we need to use MCMC output which contains the full distributional information required.

Setting-Up the Posterior Predictions

- Notation:
  - $\mathbf{y} = (y_1, \ldots, y_n)$ for discrete observed data,
  - $\mathbf{X}$ for the matrix of predictor variables,
  - and $\boldsymbol{\theta}$ for the vector of all parameters.
- Assume that a model has been fit and that we have a set of MCMC simulations, $\boldsymbol{\theta}^{(s)}, s = 1, \ldots, n_{\text{sims}}$.
- For each of these draws, a replicated dataset, $y^{\text{rep}(s)}$, has been simulated from the predictive distribution of the data $p(y^{\text{rep}}|\mathbf{X}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(s)})$.
- The abortions attitudes model is specified for $i = 1, \ldots, 264$ respondents across $j = 1, \ldots, 4$ panel waves:

$$y_{ij} \sim \mathcal{BN}(n_i, p_{ij})$$

$$\text{logit}(p_{ij}) = \boldsymbol{\beta}_{0,j} + \boldsymbol{\beta}_{1,i} X_{1,i}$$

Using Simulations

- From the core of the code:

```
logit(p[i,j]) <- b0[j] + b1[i]*x1[i]
r[i,j] ~ dbin(p[i,j], n[i])
```

- Therefore we generated 5,000 $\boldsymbol{\beta}_{0,j} \times 4$ and $\boldsymbol{\beta}_{1,i} \times 264$ simulated values.

- Using these we produce 5,000 predictions.

- The ensemble of simulated datasets $(y^{\text{rep}(s)}(1), \ldots, y^{\text{rep}}(n_{\text{sims}}))$ estimates the posterior predictive distribution, $p(y^{\text{rep}}|\mathbf{X}, \mathbf{y})$.

- Notationally suppress the conditioning on X until Section 24.2, where $\mathbf{X}$ is allowed to vary simulating X^{rep}.

Posterior Predictive Distributions for the Abortion Attitudes Model

- Put the MCMC output into an array for the $p[i, j]$:

```
N.cases <- 264; N.panels <- 4; N.total <- 264*4
p.array <- array(NA,c(N.cases,N.panels,(stop-start)))
for (i in 1:(stop-start))
   p.array[,,i] <- matrix(full.out[i,1:N.total],ncol=N.panels)
dim(p.array)
[1]  264    4 5000
```

- Retrieve the data including r:

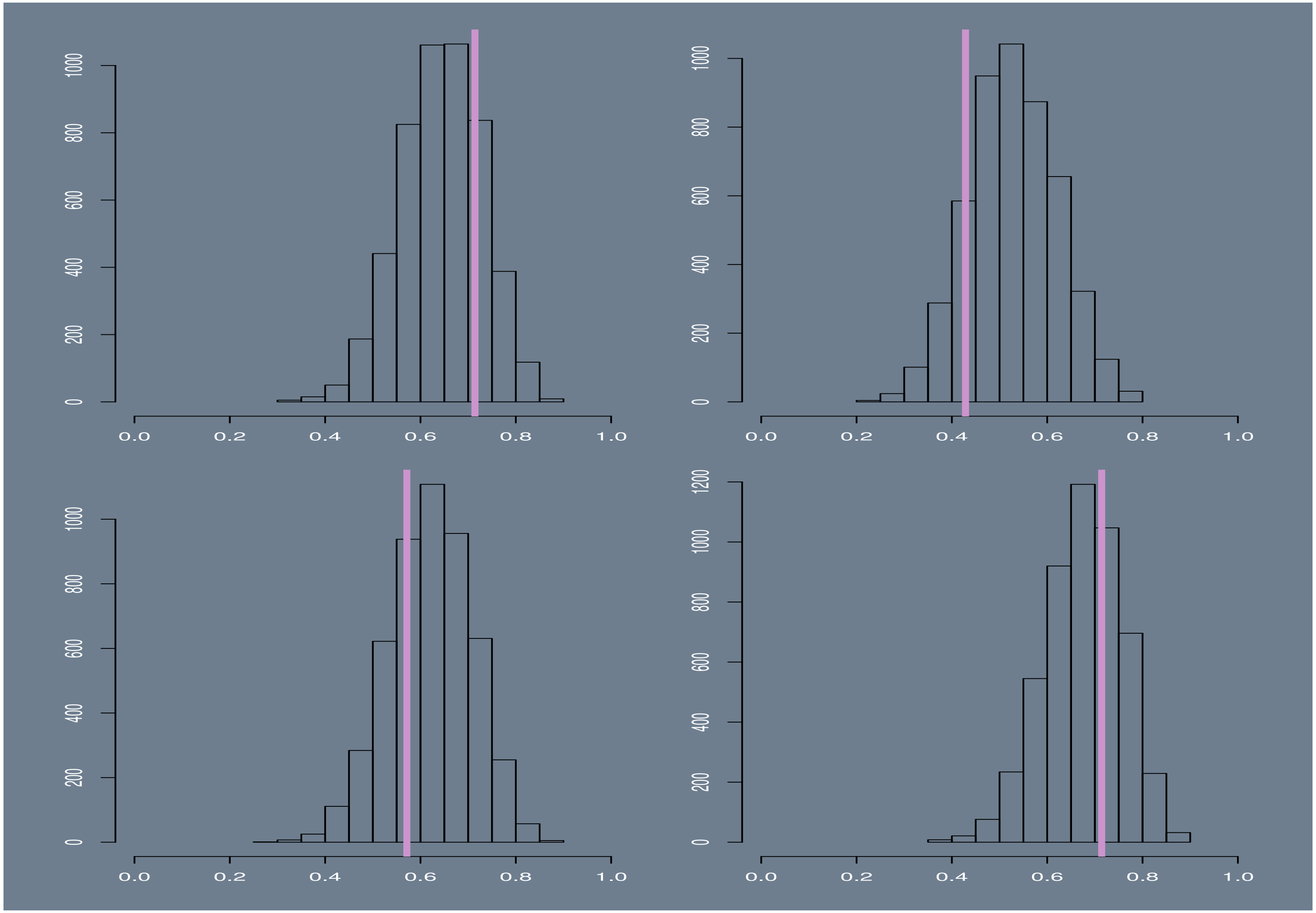
```
source("abortion.panel-data.R")
p <- r/7  # m_i=7
dim(p)
[1] 264  4
```

Posterior Predictive Distributions for the Abortion Attitudes Model

- Select and compare for a single person/case:

```
person <- 200
par(mfrow=c(2,2),mar=c(3,3,1,1),oma=c(1,1,1,1),col.axis="white",col.lab="white",
  col.sub="white",col="white",bg="slategray")
for (i in 1:N.panels) {
   hist(p.array[person,i,],main="",xlab=paste("Panel",i),xlim=c(0,1))
   abline(v=p[person,i],col="plum3",lwd=4)
}
```

Prediction and Actual For Case 200



Bayes Factor

- 2 competing models, M_1: $f_1(\mathbf{x}|\boldsymbol{\theta_1})$ $\qquad M_2$: $f_2(\mathbf{x}|\boldsymbol{\theta_2})$
- $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2} \in \boldsymbol{\Theta}$ or $\boldsymbol{\Theta_1}$ and $\boldsymbol{\Theta_2}$
- specify parameter priors: $\pi_1(\boldsymbol{\theta_1})$ and $\pi_2(\boldsymbol{\theta_2})$ and model priors: $p(M_1)$ and $p(M_2)$.
- Note that $p(\mathbf{x}|M_i) = \int_{\theta_i} f_i(\mathbf{x}|\boldsymbol{\theta_i})\pi_i(\boldsymbol{\theta_i})d\theta_i$
- Thus:

$$\underbrace{\frac{p(M_1|\mathbf{x})}{p(M_2|\mathbf{x})}}_{\text{posterior odds}} = \underbrace{\frac{p(M_1)/p(\mathbf{x})}{p(M_2)/p(\mathbf{x})}}_{\text{prior odds/data}} \times \underbrace{\frac{\int_{\theta_1} f_1(\mathbf{x}|\boldsymbol{\theta_1})\pi_1(\boldsymbol{\theta_1})d\theta_1}{\int_{\theta_2} f_2(\mathbf{x}|\boldsymbol{\theta_2})\pi_2(\boldsymbol{\theta_2})d\theta_2}}_{\text{Bayes factor}}.$$

$$\text{posterior odds ratio} = \text{prior odds ratio} \times \text{integrated likelihood ratio}$$

- Rearranging this and canceling out $p(\mathbf{x})$ gives:

$$B(\mathbf{x}) = \frac{p(M_1|\mathbf{x})/p(M_1)}{p(M_2|\mathbf{x})/p(M_2)}$$

$$= \text{“posterior to prior odds ratio”}$$

Jeffreys' Typology

$(\mathbf{x}) \geq 1$	model 1 supported
$1 > B(\mathbf{x}) \geq 10^{-\frac{1}{2}}$	minimal evidence against model 1
$10^{-\frac{1}{2}} > B(\mathbf{x}) \geq 10^{-1}$	substantial evidence against model 1
$10^{-1} > B(\mathbf{x}) \geq 10^{-2}$	strong evidence against model 1
$10^{-2} > B(\mathbf{x})$	decisive evidence against model 1

Bayes Factor Example

- $H_0: \theta = 0.5$, vs. $H_1: \theta = 0.7$
- priors: $p(H_0) = p(H_1) = \frac{1}{2}$ (so prior odds ratio just one)
- data: $n = 10, x = 7$
- Bayes Factor between "model 0 supported" and "minimal evidence against model 0":

$$B(\mathbf{x}) = \frac{p(M_1|\mathbf{x})/p(M_1)}{p(M_2|\mathbf{x})/p(M_2)} = \frac{p(x=7|\theta=0.5)}{p(x=7|\theta=0.7)} = \frac{\binom{10}{7}(0.5)^7(0.5)^3}{\binom{10}{7}(0.7)^7(0.3)^3} = 0.44$$

- Now calculate the corresponding p-value:

$$p = p(\text{observed data or more extreme}|H_0) = p(x = \{7, 8, 9, 10\}|\theta = 0.5)$$
$$= \sum_{x=7}^{10} \binom{10}{x} (0.5)^x (0.5)^{10-x} = 0.172, \text{fail to reject } H_0$$

Problems (challenges) with Bayes Factors

- Sensitivity to priors
- No directional hypotheses
- Improper priors...
  $\pi(\theta) \propto h(\theta)$, set $\pi(\theta) = ch(\theta)\ldots$ Okay for posteriors:

$$p(\theta|\mathbf{x}) = \frac{\pi(\theta)p(\mathbf{x}|\theta)}{\int_\theta \pi(\theta)p(\mathbf{x}|\theta)d\theta}$$

$$= \frac{cg(\theta)p(\mathbf{x}|\theta)}{c\int_\theta g(\theta)p(\mathbf{x}|\theta)d\theta}$$

$$= \frac{g(\theta)p(\mathbf{x}|\theta)}{\int_\theta g(\theta)p(\mathbf{x}|\theta)d\theta}$$

Bad for Bayes factors:

$$B(\mathbf{x}) = \frac{c_1}{c_2}\frac{\int_{\theta_1} g_1(\theta_1)p(\mathbf{x}|\theta_1)d\theta_1}{\int_{\theta_2} g_2(\theta_2)p(\mathbf{x}|\theta_2)d\theta_2}$$

Bayes Factors for the Linear Model

- We want to compare two, not necessarily nested, different right-hand-side specifications in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an $n \times k$, rank k matrix of explanatory variables with a leading vector of ones, $\boldsymbol{\beta}$ is a $k \times 1$ unknown vector of coefficients, $\mathbf{y}$ is an $n \times 1$ vector of outcomes, and $\boldsymbol{\epsilon}$ is a $n \times 1$ vector of residuals with $\mathcal{N}(0, \sigma^2 I)$ for a constant σ^2 (homoscedasticity).

- The likelihood function for model j is:

$$L_j(\boldsymbol{\beta}_j, \sigma_j^2 | \mathbf{X}_j, \mathbf{y}) = (2\pi\sigma_j^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_j^2} (\mathbf{y} - \mathbf{X}_j\boldsymbol{\beta}_j)'(\mathbf{y} - \mathbf{X}_j\boldsymbol{\beta}_j) \right] \tag{1}$$

  where $j = 0, 1$ providing models M_0 and M_1.

- $\mathbf{y}$ is not indexed here since both models intend to explain the structure of the same outcome.

- Make the definitions $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})/(n - k)$.

Bayes Factors for the Linear Model

- Now specify conjugate priors for each of these models with k_j columns of $\mathbf{X}$ according to:

$$p(\boldsymbol{\beta}_j|\sigma^2) = (2\pi)^{-\frac{k_j}{2}}|\boldsymbol{\Sigma}_j|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_j - \mathbb{B}_j)'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\beta}_j - \mathbb{B}_j)\right],$$

$$p(\sigma_j^2) \propto \sigma_j^{-(a_j-k_j)} \exp\left[-\frac{b_j}{\sigma_j^2}\right]$$

- We add a multiplier h_j on the variance term in the normal prior for $\boldsymbol{\beta}_j$: $\boldsymbol{\Sigma}_j = h_j\sigma_j^2\mathbf{I}$.

- Make the common choice of prior mean for $\boldsymbol{\beta}$ to be $\mathbb{B} = 0$ in both models.

- The marginal likelihood for model j from this setup is:

$$p_j(\mathbf{y}|\mathbf{X}_j, M_j) = \frac{|\mathbf{X}_j'\mathbf{X}_j + h|^{-\frac{1}{2}}|h_j|^{\frac{1}{2}}b_j^{a_j}\Gamma(a_j + \frac{a_j}{2})}{\pi^{\frac{n}{2}}\Gamma(a_j)}\left(2b_j + (n-k_j)\hat{\sigma_j}^2\right). \tag{2}$$

Bayes Factors for the Linear Model

- This means that the Bayes Factor for Model 1 over Model 0 is given by:

$$BF_{(1,0)} = \frac{p_1(\mathbf{y}|\mathbf{X}_1, M_1)}{p_0(\mathbf{y}|\mathbf{X}_0, M_0)} = \frac{\frac{|\mathbf{X}_1'\mathbf{X}_1+h|^{-\frac{1}{2}}|h_1|^{\frac{1}{2}}b_1^{a_1}\Gamma(a_1+\frac{a_1}{2})}{\pi^{\frac{n}{2}}\Gamma(a_1)}\left(2b_1+(n-k_1)\hat{\sigma_1}^2\right)}{\frac{|\mathbf{X}_0'\mathbf{X}_0+h|^{-\frac{1}{2}}|h_0|^{\frac{1}{2}}b_0^{a_0}\Gamma(a_0+\frac{a_0}{2})}{\pi^{\frac{n}{2}}\Gamma(a_0)}\left(2b_0+(n-k_0)\hat{\sigma_0}^2\right)}. \tag{3}$$

- This is a long expression but a relatively simple form due to the elegance of the linear model.

(Final) Exercise 15: BAAD Data

- *Big Allied and Dangerous* (BAAD) Database 1 (Asal, Rethemeyer & Anderson 2008).
- Assembled from several established databases: Memorial Institute for the Prevention of Terrorism's (MIPT) Terrorism Knowledge Base (TKB), Correlates of War (COW), Polity, and Polity2.
- This aggregates 395 worldwide lethal attacks from 1998-2005 by terrorist organizations.
- We use the version of their dataset that excludes Al Qaeda since its scope, profile, and effectiveness place it in a unique category during this period.
- The variable `fatalities` (total number) is used as the outcome variable to focus on the primary purpose of these attacks.

The Explanatory Variables Used

- `statespond` indicates whether the group is financially or logistically supported by one or more recognized governments (coded 1, $n_1 = 32$), or not (coded 0, $n_0 = 363$).

- `masterccode` denotes the COW `CCODE` value: where (country/region) attack took place.

- `ordsize` is size according to 0 for less than 100 members ($n_0 = 261$), 1 for 101-1,000 members ($n_1 = 77$)), 2 for 1,001-10,000 members ($n_2 = 45$), and 3 for more than 10,000 members ($n = 12$).

- `terrStrong` is coded 1 ($n_1 = 43$) if they possess territory and 0 if they do not ($n_0 = 352$).

- `degree` gives a count of alliance connections in the network sense.

The Explanatory Variables Used

- `LeftNoReligEthno`, where a 1 indicates that the group's ideology is leftist and it is not compounded with another ideological orientation ($n_1 = 94$), and a 0 indicates that group's ideology is either not leftist or is a mix of leftist and at other ideological dimensions ($n_0 = 301$).

- `PureRelig` indicates with a 1 whether the group's ideology is purely religious and not associated with other political or social factors ($n_1 = 50$), and 0 otherwise ($n_0 = 345$).

- `PureEthno` indicates with a 1 whether the group is ethnonationalist (nationalist causes tied to ethnic identity) and not associated with other ideological factors ($n_1 = 26$), and 0 otherwise ($n_0 = 369$).

- `Islam` where a 1 is assigned to groups inspired by some form of Islam ($n_1 = 287$) and 0 otherwise ($n_0 = 108$).

(Final) Exercise 15: BAAD Data

- Due to `JAGS` restrictions the variables are shortened in the dataset to. . .

```
N
fataliti
statespo
mastercc
ordsize
terrStro
degree
LeftNoRe
PureReli
PureEthn
Islam
cluster
```

(Final) Exercise 15: BAAD Data

- Your job is to code a model from scratch using these data.
- In my version `alpha[cluster[i]]` and `fataliti[i]` were important model components.
- Get the data according to:

```
baad <- source("http://jeffgill.org/files/jeffgill/files/baad.jags_.dat_.txt")
```

- The next slide has some models I ran a while ago to get you started.

	Standard Linear Model				Multilevel Linear Model		
	Mean	Std.Err.	95% HPD		Mean	Std.Err.	95% HPD
α	-0.290	1.287	[-2.811:2.232]	α_1	-3.835	0.843	[-5.486:-2.184]
				α_2	0.383	1.480	[-2.517: 3.283]
				α_3	-1.905	1.040	[-3.942: 0.133]
				α_4	19.235	1.139	[17.002:21.468]
`statespond`	0.514	1.193	[-1.824:2.851]		3.590	0.840	[1.945: 5.235]
`masterccode`	0.006	0.032	[-0.057:0.069]		-0.054	0.019	[-0.092:-0.016]
`ordsize`	4.749	0.719	[3.339:6.159]		3.163	0.452	[2.277: 4.049]
`terrStrong`	3.849	1.355	[1.193:6.504]		1.886	0.974	[-0.022: 3.795]
`degree`	2.307	0.298	[1.723:2.890]		1.169	0.179	[0.818: 1.520]
`LeftNoreligEthno`	0.290	1.070	[-1.808:2.388]		0.838	0.707	[-0.548: 2.224]
`PureRelig`	1.131	1.307	[-1.431:3.694]		1.669	0.955	[-0.202: 3.540]
`PureEthno`	-0.948	1.410	[-3.713:1.816]		-1.378	1.045	[-3.427: 0.670]
`Islam`	2.851	1.203	[0.492:5.210]		3.179	0.857	[1.499: 4.858]
τ	0.009	0.001	[0.007:0.020]		0.027	0.002	[0.023: 0.031]
	Summed Deviance 3002				Summed Deviance 2553		
						Variance	Std.Dev.
					σ_α	113.44	10.65
					σ_y	1.31	1.15