Bayesian Hierarchical Modeling for the Social Sciences

Introduction: Critical Differences in Bayesian and Non-Bayesian Inference and Why the Former is Better

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

So What's All This *&#$@*$% Bayesian Stuff Anyway?

- Overt and clear model assumptions.
- A rigorous way to make *probability* statements about the real quantities of theoretical interest.
- An ability update these statements (i.e. learn) as new information is received.
- Systematic incorporation of *qualitative* knowledge on the subject.
- Recognition that population quantities are changing over time rather than fixed immemorial.
- Straightforward assessment of both model quality and sensitivity to assumptions.
- Freedom from the flawed NHST paradigm.

Typology of Statistics

- Frequentists: From the Neyman/Pearson/Wald setup. An orthodox view that sampling is infinite and decision rules can be sharp. Estimated quantities usually produced with closed-form statements.

- Bayesians: From Bayes/Laplace/de Finetti tradition. Unknown quantities are treated probabilistically and the state of the world can always be updated.

- Likelihoodists: From Fisher. Single sample inference based on finding the parameter value, $\hat{\theta}$, that maximizes the joint distribution of the observed data ($L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)$), with properties layed-out in Birnbaum (1962). Bayesians that don't know that they are.

- *So let's look at some critical differences between Frequentists and Bayesians...*

Critical Differences Between Bayesians and Non-Bayesians, What is Fixed?

Frequentist:

- Data are an IID random sample from a continuous stream.
- Parameters are fixed by nature.

Bayesian:

- Data are observed and therefore fixed by the sample generated.
- Parameters are unknown and described distributionally.

Critical Differences Between Bayesians and Non-Bayesians, Interpretation of Probability

Frequentist:

- Probability is observed from the long-run proportion of times that some event occurs in a replicated experiment.
- Probabilistic quantity of interest is $p(\text{data}|H_0)$.

Bayesian:

- Probability is the researcher/observer "degree of belief" before or after the data are observed.
- Probabilistic quantity of interest is $p(\theta|\text{data})$.

Critical Differences Between Bayesians and Non-Bayesians, General Inference

Frequentist:

- Point estimates and standard errors or 95% *confidence* intervals.
- Deduction from $p(\text{data}|H_0)$, by setting α in advance.
- Accept H_1 if $p(\text{data}|H_0) < \alpha$.
- Accept H_0 if $p(\text{data}|H_0) \geq \alpha$.

Bayesian:

- Induction from $p(\theta|\text{data})$, starting with $p(\theta)$.
- Broad descriptions of the posterior distribution such as means and quantiles.
- Highest posterior density intervals indicating region of highest posterior probability, regardless of contiguity.

Critical Differences Between Bayesians and Non-Bayesians, Post-hoc Quality Checks

Frequentist:

- Calculation of Type I and Type II errors, even if there is no setting α in advance.
- *Sometimes*: effect size and/or power.
- *Usually*: fixation with small differences in p-values despite large measurement error in the social sciences relative to other scientific disciplines.

Bayesian:

- Posterior predictive checks from integrating over posterior.
- Sensitivity checks to forms of the prior, and other assumptions.
- Bayes factors for model comparison, BIC, DIC.

Reasons *Not* to Use Bayesian Inference in the Social Sciences:

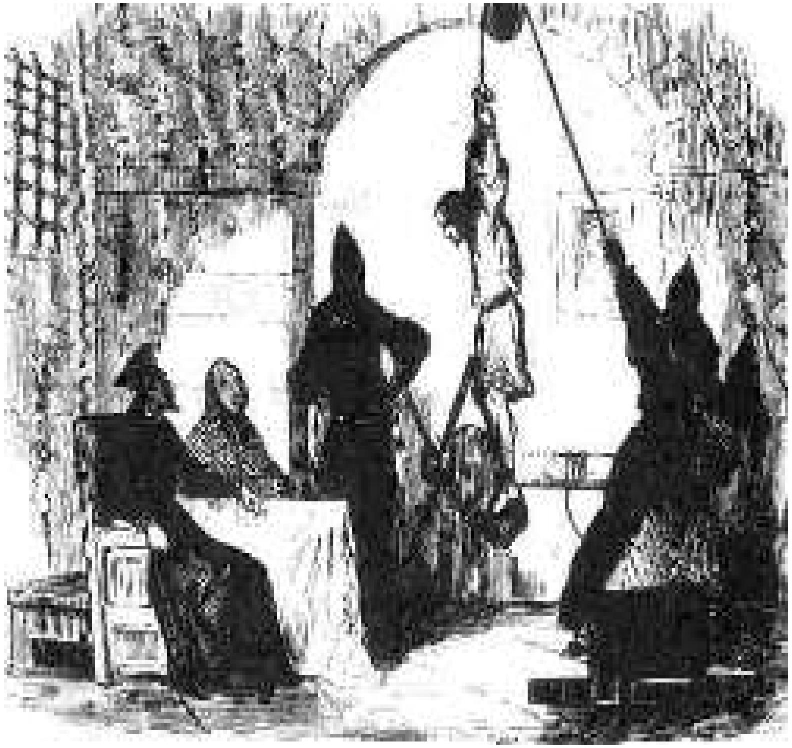
- The population parameters of interest truly fixed and unchanging under all realistic circumstances.
- We do *not* have any information prior to the model specification.
- It necessary to provide statistical results as if data were from a *controlled experiment*.
- We care more about "significance" than effect size.
- Computers are slow and relatively unavailable.
- We want very automated, "cookbook" type procedures.

Reasons to Use Bayesian Inference in the Social Sciences:

- We want to be very careful about stipulating assumptions and are willing to defend them.
- We view the world probabilistically, rather than as a set of fixed phenomena that are either known or unknown.
- Every statistical model ever created in the history of the human race is subjective; we are willing to admit it.
- Prior information abounds in the social sciences and it is important and helpful to use it.

Some Problems with Traditional Statistical Thinking in the Social Sciences

- Small-n inference.
- Significance through sample size.
- Confidence.
- Contrived ignorance and buried assumptions.
- Null Hypothesis Testing/Star-gazing.



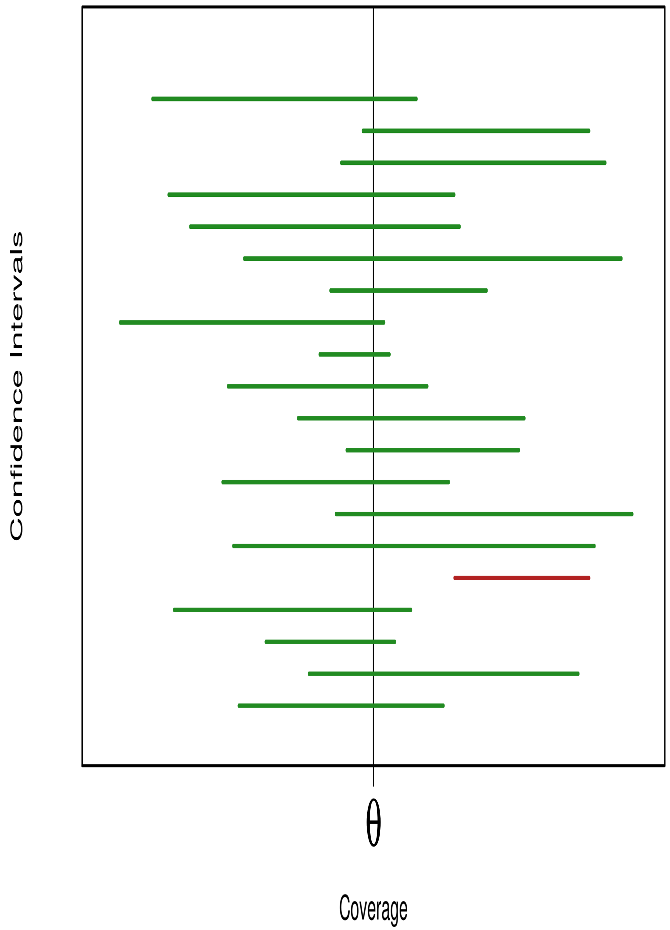
Large and Small Sample Inference

`http://setiathome.ssl.berkeley.edu/`

Marriage Rates per 1000 in Italy 1936 to 1951.

Confidence

- Which of these is the correct interpretation of a $(1-\alpha)$ confidence interval?
  - An interval that has a $1-\alpha\%$ chance of containing the true value of the parameter.
  - An interval that over $1-\alpha\%$ of replications contains the true value of the parameter, *on average*.
- What interpretation do people really *want*.



Contrived Ignorance, Buried Assumptions

- Models with uniform priors.
- Normality.
- Correlation coefficient.
- Only two models tested.
- No such thing as specification searches.

The pseudo-Frequentist NHST is wrong

- A few authors have noted this (just a sample): *Barnett 1973, Berger, Boukai, and Wang 1997, Berger Thomas Sellke 1987, Berkhardt and Schoenfeld 2003, Bernardo 1984, Brandstätter 1999, Carver 1978, 1993, Dar, Serlin and Omar 1994, Cohen 1988, 1994, 1992, 1977, 1962, Denis 2005, Falk and Greenbaum 1995, Gelman, Carlin, Stern, and Rubin 1995, Gigerenzer 1987, 1993, 1998, Gigerenzer and Murray 1987, Gill 1999, 2005, Gliner, Leech and Morgan 2002, Grayson 1998, Greenwald 1975, Greenwald, Gonzalez, Harris and Guthrie 1996, Hager 2000, Howson and Urbach 1993, Hunter 1997, Hunter and Schmidt 1990, Jeffreys 1961, Kirk 1996, Krueger 1999, 2001, Lindsay 1995, Loftus 1991, 1993a, 1993b, 1994, 1996, Loftus and Bamber 1990, Macdonald 1997, Meehl 1967, 1978, 1990, 1978, Nickerson 2000, Oakes 1986, Pollard 1993, Pollard and Richardson 1987, Robinson and Levin 1997, Rosnow and Rosenthal 1989, Rozeboom 1960, 1997, Schmidt 1996, Schmidt and Hunter 1977, Sedlmeier and Gigerenzer 1989, Thompson 2002, Wilkinson 1999.*

- Why?

1. Artificial Model Selection Criteria
2. The Arbitrariness of Alpha
3. Replication Fallacy
4. Asymmetry and Accepting the Null Hypothesis
5. Probabilistic Modus Tollens
6. Inverse Probability Problem

Regular Modus Tollens

If A then B	If H_0 is true then the data will follow an expected pattern
Not B observed	The data do not follow the expected pattern
Therefore not A	Therefore H_0 is false.

Probabilitistc Modus Tollens

If A then B is highly likely	If H_0 is true then the data are highly likely to follow an expected pattern
Not B observed	The data do not follow the expected pattern
Therefore A is highly unlikely	Therefore H_0 is highly unlikely.

Probabalistic Modus Tollens Example

If A then B is highly likely	If a person is an American, then it is highly unlikely she is a member of Congress.
Not B observed	The person is a member of Congress
Therefore A is highly unlikely	Therefore it is highly unlikely she is an American.

Misconceptions about Inverse Probability

- The inferential mechanism of the null hypothesis significance test is based on conditional probability.
- The test looks at: $p(\text{data}|H_0)$, "how likely is it to observe these data, given that the null hypothesis of no effect is *true*."
- It is commonly (mis)interpreted as: $p(H_0|\text{data})$, "how probable is the null hypothesis, given these observed data."
- These (the right and the wrong) statements are fundamentally different quantities and can only be related with Bayes' Law:

$$p(H_0|\text{data}) = \frac{p(H_0)}{p(\text{dat}a)} p(\text{data}|H_0).$$

- The problem comes from an unholy blending of Fisher and Neyman/Pearson.

Misconceptions about Inverse Probability

- The order of conditionality can be really important.

- suspected probability of AIDS in risk group: $P(A) = 0.02$

  probability of correct positive classification: $P(C|A) = 0.95$

  probability of correct negative classification: $P(C^c|A^c) = 0.97$

- Suppose we want $P(A|C)$, from:
$$P(A|C) = \frac{P(A)}{P(C)}P(C|A)$$

- Getting the unconditional:
$$\begin{aligned} P(C) &= P(C \cap A) + P(C \cap A^c) \\ &= P(C|A)P(A) + P(C|A^c)]P(A^c) \\ &= P(C|A)P(A) + [1 - P(C^c|A^c)]P(A^c) \\ &= (0.95)(0.02) + (1 - 0.97)(0.98) \cong 0.05 \end{aligned}$$

- So now we can calculate:
$$P(A|C) = \frac{P(A)}{P(C)}P(C|A) = \frac{0.02}{0.05}(0.95) = 0.38$$

The History of Bayesian Statistics–Milestones

- Reverend Thomas Bayes (1702-1761).
- Pierre Simon Laplace.
- Pearson (Karl), Fisher, Neyman and Pearson (Egon), Wald.
- Jeffreys, de Finetti, Good, Savage, Lindley, Zellner.
- A world divided.
- The revolution: Gelfand and Smith (1990).
- Today…

Two Primary Principles of Bayesian Inference

$\boxed{\text{Principle I.}}$

$\boxed{\text{Principle II.}}$

Two Primary Principles of Bayesian Inference

$\boxed{\text{Principle I.}}$ Explicit and direct use of probability for describing uncertainty:

- ▷ probability models (likelihood fn.) for data given parameters,
- ▷ probability distributions (PDF,PMF) for parameters.

$\boxed{\text{Principle II.}}$

Two Primary Principles of Bayesian Inference

Principle I. Explicit and direct use of probability for describing uncertainty:

- $\triangleright$ probability models (likelihood fn.) for data given parameters,
- $\triangleright$ probability distributions (PDF,PMF) for parameters.

Principle II. Inference for unknown values conditioned on observed data:

- $\triangleright$ use of inverse probability,
- $\triangleright$ Bayes theorem,
- $\triangleright$ description of full posterior.

The Three General Steps

$\boxed{\text{Step I.}}$

$\boxed{\text{Step II.}}$

$\boxed{\text{Step III.}}$

The Three General Steps

$\boxed{\text{Step I.}}$ Specify a probability model for unknown parameter values that includes some prior knowledge about the parameters if available.

$\boxed{\text{Step II.}}$

$\boxed{\text{Step III.}}$

The Three General Steps

$\boxed{\text{Step I.}}$ Specify a probability model for unknown parameter values that includes some prior knowledge about the parameters if available.

$\boxed{\text{Step II.}}$ Update knowledge about the unknown parameters by conditioning this probability model on observed data.

$\boxed{\text{Step III.}}$

The Three General Steps

Step I. Specify a probability model for unknown parameter values that includes some prior knowledge about the parameters if available.

Step II. Update knowledge about the unknown parameters by conditioning this probability model on observed data.

Step III. Evaluate the fit of the model to the data and the sensitivity of the conclusions to the assumptions.

Simple Mechanics

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

$$\propto p(\theta)L(\theta|\mathbf{x})$$

Posterior Probability $\propto$ Prior Probability $\times$ Likelihood Function

Views On Priors Determine Types of Bayesians

- Empirical Bayes: prior distributions are produced from other parts of the data, or possibly from the same data. Results are reported like Frequentists.

- Proper Bayes: prior distributions come from previously compiled evidence, such earlier studies or published work, researcher intuition, or substantive experts. Results are reported without utility or loss functions.

- Reference Bayes: prior distributions are created to influence the posterior as little as mathematically possible ("objective"). Results are reported without utility or loss functions.

- Decision-Theoretic Bayes: prior distributions are from either of the last two sources. Results are presented in a full decision-theoretic framework where utility functions determine decision losses, which are minimized according to different probabilistic criteria.

- Bayesians of Convenience: conjugate diffuse priors or uniform priors on all parameters. Results are reported without utility or loss functions.

George Box

- Sampling and Bayes' Inference in Scientific Modelling and Robustness. George E. P. Box, *Journal of the Royal Statistical Society. Series A (General)*, Volume 143, Number 4 (1980), 383-430.

- "In the past, the need for probabilities expressing prior belief has often been thought of, not as a necessity for all scientific inference, but rather as a feature peculiar to Bayesian inference. This seems to come from the curious idea that an outright assumption does not count as a prior belief.

- In my words: we all make prior assumptions all the time in the modeling process, Bayesians are just more overt about it.

Example: the Beta-Binomial

- $X_1, X_2, \ldots, X_n$ iid Bernoulli, $p \sim \text{beta}(A, B)$ prior.
- Standard trick: $Y = \sum_{i=1}^{n} X_i \sim \text{binomial}(n, p)$.

Example: the Beta-Binomial

- $X_1, X_2, \ldots, X_n$ iid Bernoulli, $p \sim \text{beta}(A, B)$ prior.
- Standard trick: $Y = \sum_{i=1}^{n} X_i \sim \text{binomial}(n, p)$.
- Joint Distribution:

$$\begin{aligned}
f(y,p) &= f(y|p)f(p) \\
&= \left[\binom{n}{y} p^y (1-p)^{n-y}\right] \quad \times \quad \left[\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} p^{A-1}(1-p)^{B-1}\right] \\
&= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} p^{y+A-1}(1-p)^{n-y+B-1}
\end{aligned}$$

Example: the Beta-Binomial

- $X_1, X_2, \ldots, X_n$ iid Bernoulli, $p \sim \text{beta}(A, B)$ prior.
- Standard trick: $Y = \sum_{i=1}^{n} X_i \sim \text{binomial}(n, p)$.
- Joint Distribution:

$$\begin{aligned} f(y,p) &= f(y|p)f(p) \\ &= \left[\binom{n}{y} p^y (1-p)^{n-y}\right] \quad \times \quad \left[\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} p^{A-1}(1-p)^{B-1}\right] \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} p^{y+A-1}(1-p)^{n-y+B-1} \end{aligned}$$

- Marginal Distribution for y:

$$\begin{aligned} f(y) &= \int_0^1 \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} p^{y+A-1}(1-p)^{n-y+B-1} dp \\ &= \frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \frac{\Gamma(y+A)\Gamma(n-y+B)}{\Gamma(n+A+B)} \end{aligned}$$

Example: the Beta-Binomial, Cont.

- Posterior Distribution for p:

$$f(p|y) = \frac{f(y,p)}{f(y)} = \frac{\frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} p^{y+A-1}(1-p)^{n-y+B-1}}{\frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \frac{\Gamma(y+A)\Gamma(n-y+B)}{\Gamma(n+A+B)}}$$

$$= \frac{\Gamma(n+A+B)}{\Gamma(y+A)\Gamma(n-y+B)} p^{(y+A)-1}(1-p)^{(n-y+B)-1}$$

$$p|y \sim \text{beta}(y+A, n-y+B)$$

Example: the Beta-Binomial, Cont.

- Posterior Distribution for p:

$$f(p|y) = \frac{f(y,p)}{f(y)} = \frac{\frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} p^{y+A-1}(1-p)^{n-y+B-1}}{\frac{\Gamma(n+1)\Gamma(A+B)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(A)\Gamma(B)} \frac{\Gamma(y+A)\Gamma(n-y+B)}{\Gamma(n+A+B)}}$$

$$= \frac{\Gamma(n+A+B)}{\Gamma(y+A)\Gamma(n-y+B)} p^{(y+A)-1}(1-p)^{(n-y+B)-1}$$

$$p|y \sim \text{beta}(y+A, n-y+B)$$

- An implication:

$$\bar{p} = \frac{(y+A)}{(y+A)+(n-y+B)} = \left[\frac{n}{A+B+n}\right]\left(\frac{y}{n}\right) + \left[\frac{A+B}{A+B+n}\right]\left(\frac{A}{A+B}\right)$$

Example: the Beta-Binomial, Cont.

- The Data (Romney 1999):

Response:	1	1	1	1	0	1	1	0	1	0	1	1
	1	0	1	1	1	1	1	1	0	0	0	1

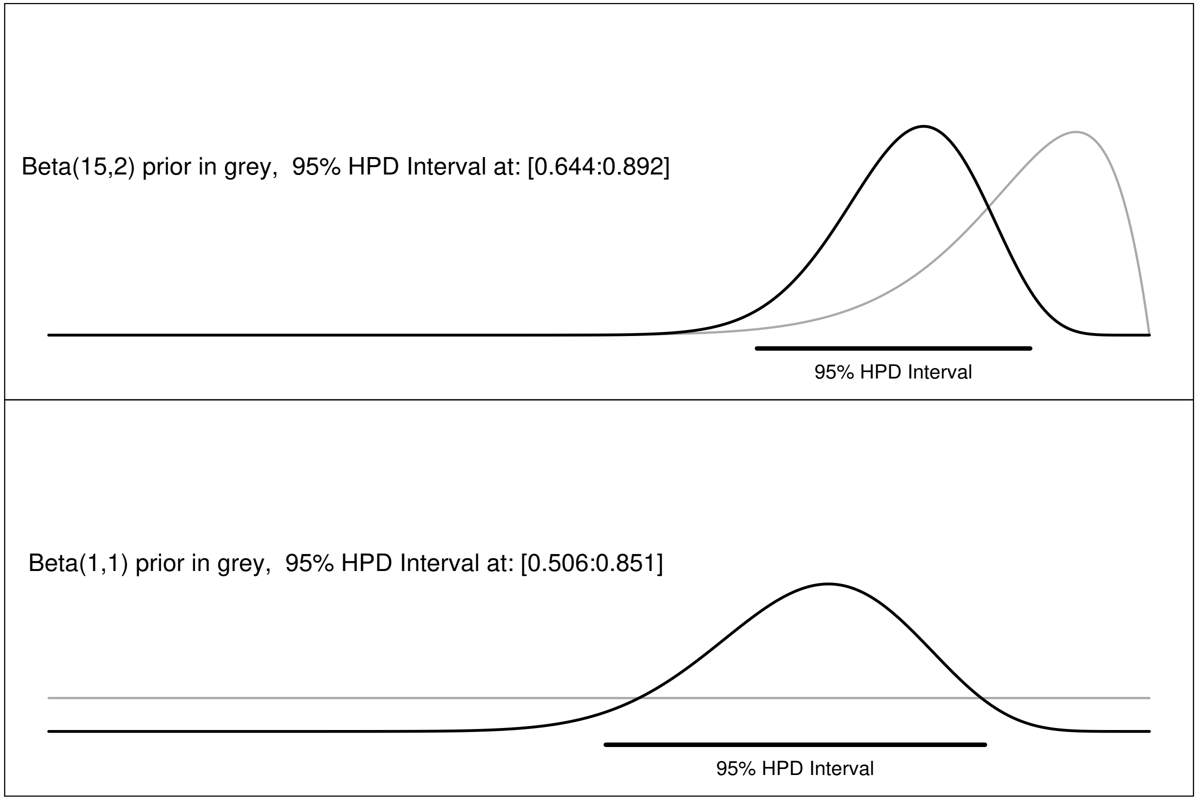
- Two Priors: $\mathcal{BE}(p|15,2)$, $\mathcal{BE}(p|1,1)$

- Resulting Posteriors:

$$\mathcal{BE}\left(\sum x_i + 15, n - \sum x_i + 2\right) = \mathcal{BE}(32,9),$$

and $$\mathcal{BE}\left(\sum x_i + 1, n - \sum x_i + 1\right) = \mathcal{BE}(18,8)$$

Example: the Beta-Binomial, Cont.



Bayesian Tobit Model for Death Penalty Support

- Use the Tobit model (Tobin 1958) to look at social and political influences on U.S. state decisions to impose the death penalty since the Supreme Court ruled the practice constitutional in *Furman v. Georgia* 1972.

- Does the ideological, racial and religious makeup, political culture, and urbanization are causal effects for state-level death sentences from 1993 to 1995.

- The Tobit model is necessary to account for censoring here because 15 states did not have capital punishment provisions on the books in the studied period.

Bayesian Tobit Model for Death Penalty Support

- If $\mathbf{z}$ is a latent outcome variable in this context with the assumptions

$$\mathbf{z} = \mathbf{x}\boldsymbol{\beta} + \mathbf{E}$$

and

$$z_i \sim \mathcal{N}(\mathbf{x}\boldsymbol{\beta}, \sigma^2),$$

then the observed outcome variable is produced according to:

$$y_i = z_i \qquad \text{if} \qquad z_i > 0,$$

and

$$y_i = 0 \qquad \text{if} \qquad z_i \leq 0.$$

- The likelihood function is then:

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = \prod_{y_i=0} \left[1 - \Phi\left(\frac{x_i\boldsymbol{\beta}}{\sigma} \right) \right] \prod_{y_i>0} (\sigma^{-1}) \exp\left[-\frac{1}{2\sigma^2}(y_i - x_i\boldsymbol{\beta})^2 \right].$$

Bayesian Tobit Model for Death Penalty Support

- A flexible parameterization for the priors is given by

$$\boldsymbol{\beta}|\sigma^2 \sim \mathcal{N}(\boldsymbol{\beta}_0, \mathbf{I}\sigma^2 B_0^{-1}) \qquad\qquad \sigma^2 \sim \mathcal{IG}\left(\frac{\gamma_0}{2}, \frac{\gamma_1}{2}\right)$$

with vector hyperparameter $\boldsymbol{\beta}_0$, scalar hyperparameters B_0, $\gamma_0 > 2$, $\gamma_1 > 0$, and appropriately

sized identity matrix $\mathbf{I}$.

- Substantial prior flexibility can be achieved with varied levels of these parameters, although values far from those implied by the data will make the Gibbs sampler algorithm run very slowly.

Bayesian Tobit Model for Death Penalty Support

- The resulting joint posterior, $\pi(\boldsymbol{\beta}, \sigma^2, z|\mathbf{y}, \mathbf{X})$, is now analytically *intractible*, even with this basic model:

$$\pi(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) = \prod_{y_i=0}\left[1 - \Phi\left(\frac{x_i\boldsymbol{\beta}}{\sigma}\right)\right]\prod_{y_i>0}(\sigma^{-1})\exp\left[-\frac{1}{2\sigma^2}(y_i - x_i\boldsymbol{\beta})^2\right]$$

$$\times\,(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_0)\right]\frac{\left(\frac{\gamma_1}{2}\right)^{\frac{\gamma_0}{2}}}{\Gamma\left(\frac{\gamma_0}{2}\right)}(\sigma^2)^{-\left(\frac{\gamma_0}{2}+1\right)}\exp[-\left(\frac{\gamma_1}{2}\right)/\sigma^2]$$

- To produce a regression table we now need to seven six-dimensional integrals (6 parameters in $\boldsymbol{\beta}$ plus σ^2) to get the marginal posteriors, then seven times two more one-dimensional integrals to get the first two moments for each parameter.

- Better solution: Gibbs sampling (MCMC) which cycles through iterative draws of the full conditional distributions for each model parameter.

Bayesian Tobit Model for Death Penalty Support

- The full conditional distributions for Gibbs sampling are given for the $\boldsymbol{\beta}$ block, σ^2, and the individual $z_i|y_i = 0$ as:

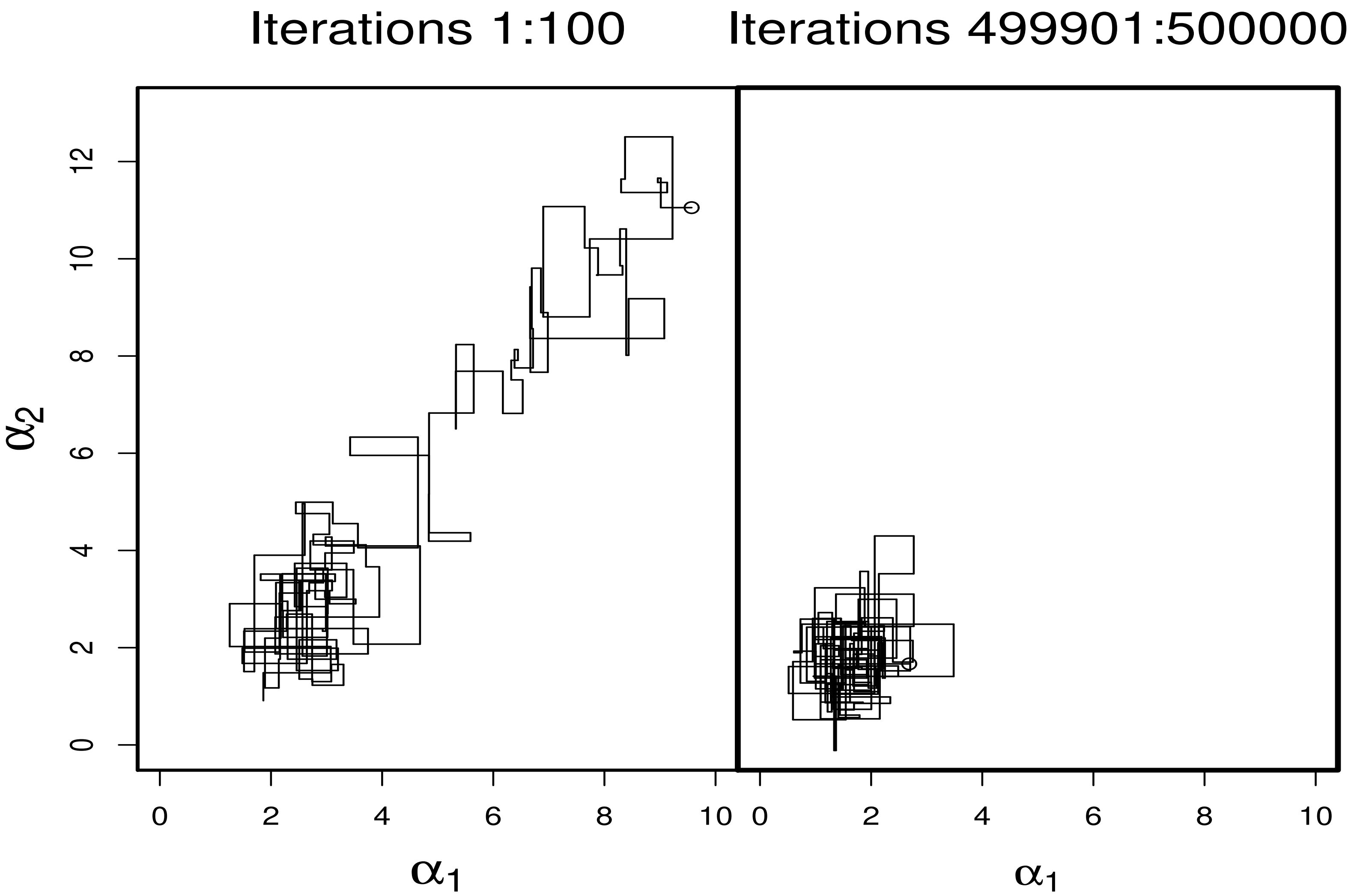
$$\boldsymbol{\beta}|\sigma^2, \mathbf{z}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}\Big((B_0 + \mathbf{X}'\mathbf{X})^{-1})(\boldsymbol{\beta}_0 B_0 + \mathbf{X}'\mathbf{z}), (\sigma^{-2}B_0 + \sigma^{-2}\mathbf{X}'\mathbf{X})^{-1})\Big)$$

$$\sigma^2|\boldsymbol{\beta}, \mathbf{z}, \mathbf{y}, \mathbf{X} \sim \mathcal{IG}\left(\frac{\gamma_0 + n}{2}, \frac{\gamma_1 + (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})}{2}\right)$$

$$z_i|y_i = 0, \boldsymbol{\beta}, \sigma, \mathbf{X} \sim \mathcal{TN}(\mathbf{X}\boldsymbol{\beta}, \sigma^2)I_{(-\infty,0)},$$

where $\mathcal{TN}()$ denotes the truncated normal and the indicator function $I_{(-\infty,0)}$ provides the bounds of truncation.

Gibbs Sampling Illustration



Posterior Summary, Tobit Model for Death Penalty Support

	$\bar{\boldsymbol{\beta}}$	$\sigma_{\boldsymbol{\beta}}$
Constant	-6.7600	3.5630
Past Rates	25.5586	8.0697
Political Culture	0.7919	0.1398
Current Opinion	5.9499	1.0805
Ideology	0.2638	1.0961
Murder Rate	0.1800	0.0764

Bayes Factors

- 2 competing models, M_1: $f_1(\mathbf{x}|\theta_{\mathbf{1}})$ $\qquad M_2$: $f_2(\mathbf{x}|\theta_{\mathbf{2}})$
- $\theta_{\mathbf{1}}$ and $\theta_{\mathbf{2}} \in \boldsymbol{\Theta}$ or $\boldsymbol{\Theta_1}$ and $\boldsymbol{\Theta_2}$
- specify parameter priors: $\pi_1(\theta_{\mathbf{1}})$ and $\pi_2(\theta_{\mathbf{2}})$ and model priors: $p(M_1)$ and $p(M_2)$.
- Note that $p(\mathbf{x}|M_i) = \int_{\theta_i} f_i(\mathbf{x}|\theta_{\mathbf{i}})\pi_i(\theta_{\mathbf{i}})d\theta_i$
- Thus:

$$\underbrace{\frac{p(M_1|\mathbf{x})}{p(M_2|\mathbf{x})}}_{\text{posterior odds}} = \underbrace{\frac{p(M_1)/p(\mathbf{x})}{p(M_2)/p(\mathbf{x})}}_{\text{prior odds/data}} \times \underbrace{\frac{\int_{\theta_1} f_1(\mathbf{x}|\theta_{\mathbf{1}})\pi_1(\theta_{\mathbf{1}})d\theta_1}{\int_{\theta_2} f_2(\mathbf{x}|\theta_{\mathbf{2}})\pi_2(\theta_{\mathbf{2}})d\theta_2}}_{\text{Bayes factor}}.$$

posterior odds ratio = prior odds ratio × integrated likelihood ratio

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

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

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