A Pragmatic Justification for the Use of Bayesian Methods in the Social Sciences

Andrew D. Martin
Washington University

October 7, 2005
University of South Carolina
Political Science Research Workshop
Outline

• I. The Bayesian Approach
• II. Model Fitting via Simulation
• III. Pragmatic Justification
• IV. Practical Considerations and Software
• V. Conclusion
I. The Bayesian Approach

Bayesian inference is a means of making rational probability statements about quantities of interest (observables, model parameters, functions of model parameters). The central feature of Bayesian inference is the direct quantification of uncertainty (Gelman, et al., 1996, p. 4).

Inferences are to be made by combining the information provided by prior probabilities with that given by the sample data; this combination is achieved by the repeated use of Bayes' theorem (Lindley, 1965, p. xi), and the final inferences are expressed solely by the posterior probabilities (Barnett 1999, p. 208).
The Process of Bayesian Data Analysis

• Set up the full probability model
• Posit prior beliefs
• Calculate the posterior distribution
• Summarize the posterior distribution
• Check model adequacy
Notation

- **Data**: \( y \)
- **Covariates**: \( x \)
- **Parameters**: \( \theta = (\theta_1, \theta_2, \ldots, \theta_k) \)
- **Probability Model**: \( f(y|\theta, x) \) \( f(y|\theta) \)
- **Priors**: \( f(\theta) \)
Bayes Theorem

- Posterior: \[ f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \propto f(y|\theta)f(\theta) \]

- Prior predictive distribution: \[ f(y) = \int_\theta f(y|\theta)f(\theta)d\theta \]

- Posterior predictive distribution: \[ f(y_{new}|y) = \int_\theta f(y_{new}|\theta)f(\theta|y)d\theta \]
Example: Linear Regression

- **Observations:** \( i = 1, \ldots, n \)
- **DV:** \( y = \{y_1, y_2, \ldots, y_n\} \)
- **IV:** \( x = \{x_1, x_2, \ldots, x_n\} \)
- **Parameters:** \( \theta = \{\beta, \sigma^2\} \)
- **Probability Model:** \( y_i \mid \beta, \sigma^2, x_i \sim \mathcal{N}(x_i' \beta, \sigma^2) \)
Likelihood

\[ f(y|\beta, \sigma^2) = \prod_{i=1}^{n} \phi \left( \frac{y_i - x'_i \beta}{\sigma} \right) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_i - x'_i \beta)^2 \right] \]

- Estimate using OLS or ML and rely on asymptotic theory to get standard errors, perform hypothesis tests, etc.
- Regularity conditions
- Sometimes “integrate out” incidental parameters
Bayesian Inference for Linear Regression

- Priors:

\[ \beta \sim \mathcal{N}_K(b_0, B_0^{-1}) \]

\[ \sigma^{-2} \sim \text{Gamma}(c_0/2, d_0/2) \]

- Posterior:

\[
f(\beta, \sigma^2|y) \propto \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (y_i - x_i'\beta)^2 \right] \times f(\beta) f(\sigma^2)\]
II. Model Fitting Using Simulation

- Monte Carlo Method -- learn anything we want about a random variable if we can randomly sample from its density
- We want to learn about the posterior density
- Markov chain Monte Carlo
- CLARIFY
Learning About A Mixture of Normals

\[ f(y) = \frac{1}{2} \phi \left( \frac{y - 2}{2} \right) + \frac{1}{2} \phi \left( \frac{y + 3}{1} \right) \]
Bayesian Inference Using Simulation

• Posterior distribution is “target”

• We generate sequence of draws from the target $f_g^*(\theta|y)$

• Compute quantities of interest, such as the posterior mean, standard deviation, and credible intervals

• For example:

\[
E(\theta|y) = \int_\Theta \theta f(\theta|y) d\theta \approx \frac{1}{G} \sum_{g=1}^{G} f_g^*(\theta|y)
\]
Gibbs Sampling

- For many posteriors (especially with semi-conjugate priors) the conditionals are known distributions

- **Target:** \( f(\theta_1, \theta_2, \theta_3 | y) \)

- **Starting Values:** \( \theta_2^{(0)} \) and \( \theta_3^{(0)} \)

- **Repeat:** \( g = 1, ..., G \)  
  Draw \( \theta_1^g \) from \( f(\theta_1 | \theta_2, \theta_3, y) \)  
  Draw \( \theta_2^g \) from \( f(\theta_2 | \theta_1, \theta_3, y) \)  
  Draw \( \theta_3^g \) from \( f(\theta_3 | \theta_1, \theta_2, y) \)
Gibbs Sampling from a Bivariate Normal Distribution
Metropolis-Hastings

- **Target:** \( f(\theta|y) \)

- **Starting Value:** \( \theta^{(0)} \quad g = 1, ..., G \)

- **Draw proposal:** \( \theta^* \) from \( p_g(\theta^*|\theta^{(g-1)}) \)

- **Set:** \( \theta^{(g)} = \begin{cases} \theta^* & \text{with probability } \alpha^* \\ \theta^{(g-1)} & \text{with probability } (1 - \alpha^*) \end{cases} \)

- **With:** \( \alpha^* = \min \left\{ \frac{f(\theta^*|y)}{f(\theta^{(g-1)}|y)} \frac{p_g(\theta^{(g-1)}|\theta^*)}{p_g(\theta^*|\theta^{(g-1)})}, 1 \right\} \)
Sampling from a Bivariate Normal Distribution using M-H
III. Pragmatic Justification (Or Six Reasons To Use Bayesian Methods)
A. Intuitive Interpretation of Results

• Confidence Interval vs. Credible Interval
• Hypothesis Testing
• Predicted Values and Probabilities
• No Asymptotic Theory Necessary
B. Quantities of Interest

- Using MCMC it is easy to directly estimate quantities of interest and assign probabilities to them.
- This is not unlike CLARIFY for predicted values, but is more powerful.
- Example -- Who is the median Senator? Is Senator Graham or DeMint more conservative? (Clinton, Jackman, and Rivers, *APSR*, 2004)
• Ideal point model with estimated ideal points:

\[ \theta_j \text{ for } j = 1, \ldots, J \]

• Draw from posterior:

\[ \{ \theta_1^{(g)}, \theta_2^{(g)}, \ldots, \theta_J^{(g)} \} \]

• Estimate of median:

\[ \text{median}^{(g)} = \text{med} \left( \theta_1^{(g)}, \theta_2^{(g)}, \ldots, \theta_J^{(g)} \right) \]

• Probability DeMint to right of Graham:

\[ \theta_{DeMint}^{(g)} > \theta_{Graham}^{(g)} \]
C. Incorporate Prior Information

- Much of the time “uninformative” priors are used.
- But other information, including that collected from qualitative sources, can formally be included in the analysis using priors.
- Must translate this knowledge into statements about parameters; prior elicitation.
- Example -- Priors on marginal effects of various covariates on attitudes toward the judiciary in Nicaragua Gill and Walker, *JOP*, 2005 (see also Western and Jackman, *APSR*, 1994).
D. Fit Otherwise Intractable Models

- Multinominal / Multivariate Probit Models (Quinn, Martin, and Whitford, AJPS, 1999)

\[ P_r(y_i = 3) = \int_{-\infty}^{0} \int_{0}^{\infty} \int_{-\infty}^{0} \phi_3(z_i, \mu_i, \Sigma) \, dz_1 \, dz_2 \, dz_3 \]


- Hierarchical / Multi-level Models (Western, AJPS, 1998, analyzing the Alvarez, Garrett, and Lange OECD data) / Complex Dependence

- Models with Discrete Parameters (change point models, mixture models with varying number of mixture components)
E. Model Comparison

• Compare models with different blocks of covariates
• Compare models with different functional forms (e.g., logit vs. probit)
• Compare any number of models
• Bayes factors (which use the prior predictive distribution or the marginal likelihood)
• Example -- Multinomial Probit or Multinomial Logit: Quinn, Martin, and Whitford, AJPS, 1999 (see also Kass and Raftery, JASA, 1995)
F. Missing Data

- It is easy to deal with missing data in the Bayesian context
- Think of multiple imputation on the fly
- Data augmentation
- This is done automatically by certain software packages (BUGS)
Concerns

✧ This seems really complicated. What I currently do seems much easier.

✧ Priors seem subjective, and should not be part of a scientific analysis.

✧ Priors will dictate results.

✧ There is no way to know whether an MCMC algorithm has converged.

✧ There is not good software to do Bayesian inference.
IV. Practical Considerations and Software

- Roll your own
- The BUGS language
- WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs)
- JAGS (http://calvin.iarc.fr/~martyn/software/jags/)
- OpenBUGS (http://mathstat.helsinki.fi/openbugs/)
- R (http://www.r-project.org/)
BUGS Code to Fit Linear Regression

model {

    # likelihood
    for(i in 1:N) {
        Y[i] ~ dnorm(mu[i],tau)
        mu[i] <- inprod(X[i,], beta)
    }

    for(j in 1:M) {

        # priors
        for(k in 1:K) {
            beta[k] ~ dnorm(0,0.001)
        }

        tau ~ dgamma(0.001, 0.001)
    }
}
Fitting Models in R

- **MCMCpack** -- general model-fitting for linear regression, a general linear panel model, ecological inference models, probit model, logit, a item response theory models, factor analysis models, Poisson regression, tobit regression, multinomial logit, and an ordered probit model ([http://mcmcpack.wustl.edu](http://mcmcpack.wustl.edu))

- **bayesm** -- more model fitting for a variety of models in econometrics

- **coda** -- post-estimation analysis for summarization and convergence diagnostics / mcmc objects

- See the Bayesian Task View on the Comprehensive R Archive Network ([http://cran.r-project.org/](http://cran.r-project.org/))
library(MCMCpack)
data(birthwt)
posterior <- MCMCprobit(low~age+
        as.factor(race) + smoke, data=birthwt)
summary(posterior)
plot(posterior)
adm@ichiro R> summary(posterior)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.61453</td>
<td>0.52113</td>
<td>0.0052113</td>
<td>0.0089813</td>
</tr>
<tr>
<td>age</td>
<td>-0.02233</td>
<td>0.02026</td>
<td>0.0002026</td>
<td>0.0003908</td>
</tr>
<tr>
<td>as.factor(race)2</td>
<td>0.62144</td>
<td>0.30495</td>
<td>0.0030495</td>
<td>0.0051266</td>
</tr>
<tr>
<td>as.factor(race)3</td>
<td>0.65385</td>
<td>0.24595</td>
<td>0.0024595</td>
<td>0.0047243</td>
</tr>
<tr>
<td>smoke</td>
<td>0.69552</td>
<td>0.22075</td>
<td>0.0022075</td>
<td>0.0036388</td>
</tr>
</tbody>
</table>

2. Quantiles for each variable:

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.64237</td>
<td>-0.95598</td>
<td>-0.62114</td>
<td>-0.264527</td>
<td>0.42319</td>
</tr>
<tr>
<td>age</td>
<td>-0.06312</td>
<td>-0.03586</td>
<td>-0.02195</td>
<td>-0.008426</td>
<td>0.01626</td>
</tr>
<tr>
<td>as.factor(race)2</td>
<td>0.02928</td>
<td>0.41279</td>
<td>0.62681</td>
<td>0.824368</td>
<td>1.21178</td>
</tr>
<tr>
<td>as.factor(race)3</td>
<td>0.17655</td>
<td>0.48533</td>
<td>0.65218</td>
<td>0.818054</td>
<td>1.13375</td>
</tr>
<tr>
<td>smoke</td>
<td>0.26961</td>
<td>0.54749</td>
<td>0.69319</td>
<td>0.845839</td>
<td>1.13360</td>
</tr>
</tbody>
</table>

adm@ichiro R>
V. Conclusion

• Bayesian methods are important components of the political methodology toolkit.

• The ability to creatively and flexibly model data and quantify all quantities of interest makes the approach very powerful.

• Better software is needed before Bayesian methods become mainstream.

• Perhaps it is time to re-think the political methodology canon as it is taught in nearly every Ph.D. program.