Bayesian Computing in the Undergraduate Statistics Curriculum

Jim Albert
Emeritus Professor
Bowling Green State University

June 2022

Outline

- Review my efforts in Bayesian pedagogy
- Review some Bayesian computational methods
- Illustrate methods using a multilevel model
- What methods to use in a first course?

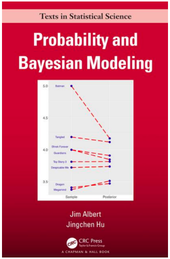
“Bayesian Inference”

- Masters-level Bayesian course taught to a wide audience
- Used a variety of different Bayesian texts over the years
- Computation component led to ”Bayesian Computation with R” text

"Baby Bayes"

- Frustrated with traditional intro-stats course
- Inspired by Don Berry and 1960's texts, I introduced a Bayes flavor of Introductory Statistics (workshop style)
- Text (with Allan Rossman) "Workshop Statistics: Discovery with Data, A Bayesian Approach"
- Out of print, but available on-line

Math-Stat Bayes (with Monika Hu)



- Alternative to traditional math-stat course
- Target audience is undergraduates with a calculus background

Learning Outcomes in Math/Stat Bayes

- How to think about and construct priors
- How are the prior and data information combined
- Simulation-based inference
- Applications of prediction
- Implement Bayes in popular methods (regression and multilevel modeling)

How to Compute in a First Course?

- Which Bayesian computational method to use?
- Which method will help in achieving the Bayesian learning goals?
- Is a "black-box" Bayesian tool desirable?

Computational Methods

- Grid approach
- Conjugate Priors
- Normal approximation
- MCMC - Metropolis & Hamiltonian Sampling

Example: A Bayesian Multilevel Model

Data: Collect number of hits (y) and number of at-bats (n) for a group of N baseball players

- $y_1, ..., y_N, y_i \sim Binomial(n_i, p_i)$
- $p_1, ..., p_N \sim Beta(K\eta, K(1-\eta))$
- $\eta \sim Beta(a, b)$, $\log K \sim Logistic(logn, 1)$

Focus on Second-Stage Parameters

- Have $N + 2$ parameters $p_1, \ldots, p_N, K, \eta$
- Interested in marginal posterior of (η, K) :

$$g(\eta, K|y) \propto g(\eta, K) \prod_{j=1}^{N} \frac{B(K\eta + y_j, K(1-\eta) + n_j - y_j)}{B(K\eta, K(1-\eta))}$$

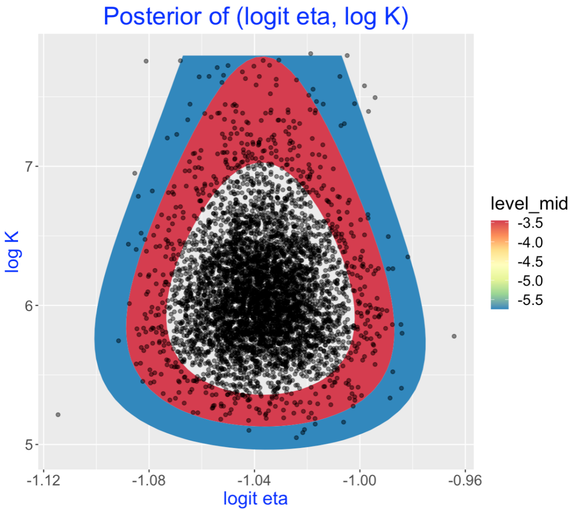
Grid Computation

- Set up a grid of values for each parameter
- Compute posterior on the grid

Grid Computation for Example

- Choose a 50 by 50 grid that covers posterior of (logit η, log K)
- Graph posterior by contour plot
- Simulate values of parameters from grid

Grid Computation & Simulation



Grid Computation - Pros and Cons

Pros:

- Easy to implement and visualize
- Introduce simulation of posterior

Cons:

- Only works for problems with a small number of parameters

Conjugate Priors

- Suppose have a sample from exponential family
- For each distribution, there exists a "conjugate" prior so that both prior and posterior have same functional form
- Posterior and predictive distributions are available

Conjugate Analyses - Pros and Cons

Pros:

- Simple expressions for posterior mean and variance
- Easy to see how prior information and data get combined
- Summarize posterior and predictive distributions by simulation

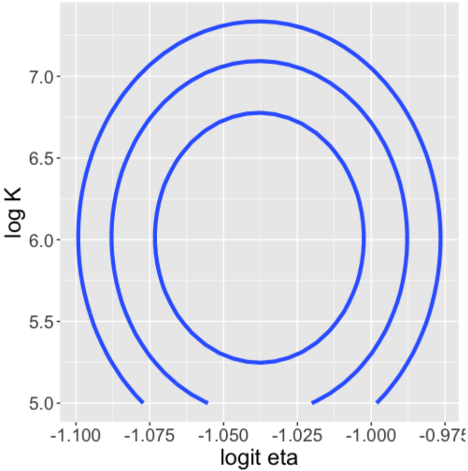
Cons:

- Limit to a small number of models

Normal Approximation

- Expand logarithm of posterior in Taylor series about mode $\hat{\theta}$
- Approximate posterior by a $N(\hat{\theta}, V)$ distribution
- Implement by Newton Raphson

Example: Normal Approximation



Normal Approximation - Pros and Cons

Pros:

- General approach – can be used for arbitrary prior and sampling density
- Computationally quick
- Use simulation methodology to do inference

Cons:

- Not "exact" method
- Limited to small number of parameters

MCMC - Metropolis Algorithm

- Simple random walk algorithm
- Easy to program
- Discuss MCMC diagnostics such as acceptance rates, trace plots and autocorrelation plots

Metropolis using JAGS

- User writes a model script
- Single R function command does the sampling

Limitations of Metropolis

- Efficient Metropolis may only accept 25% of the time.
- Can be slow in sampling of regions of high posterior content
- Metropolis can be ineffective for high-dimensional problems such as multilevel modeling

Hamiltonian Monte Carlo (HMC)

- Employs a guided proposal random walk
- Use gradient of log posterior to direct Markov chain towards regions of highest posterior density
- A well-tuned HMC chain will accept proposals at much higher rate
- Requires the log posterior and the gradient function

Stan

- Stan is well-documented software for implementing a version of HMC for a wide variety of Bayesian models
- There are R packages (such as brms) that provide high-level functions for popular Bayesian regression and multilevel models

MCMC - Pros and Cons

Pros:

- General approach – software is available
- Stan (HMC) is "state-of-the-art" for Bayesian computing

Cons:

- Does it achieve learning objectives, such as how to construct priors?
- For example, default priors are hidden in the commands in the brms package.

What computational methods are discussed in Bayesian texts?

Four modern Bayesian introductory texts:

- *A Student's Guide to Bayesian Statistics* by Lambert
- *Rethinking Statistics, 2nd edition* by McElreath
- *Bayes Rules!* by Johnson, Ott, and Dogucu
- *Probability and Bayesian Modeling* by Albert and Hu

A Student's Guide to Bayesian Statistics

Chapters focused on posterior computation:

- Chapter 5 – discrete approach to fish in bowl and proportion examples
- Chapter 9 – conjugate priors
- Chapter 12 – Monte Carlo with independent and dependent sampling
- Chapter 13 – Metropolis algorithm
- Chapter 14 – Gibbs sampling
- Chapters 15-16 – Hamiltonian MC and Stan

Rethinking Statistics

- Chapter 3 - sampling from grid-approximated posterior
- Chapter 4- use quadratic approximation (`quap()` function), and sampling from the approximation
- Chapter 9 - MCMC (Metropolis and HMC)
- Quadratic approximation and HMC used in multiple chapters

Bayes Rules!

- Chapters 3, 5 - conjugate models
- Chapter 7 - introduction to Metropolis-Hastings algorithms
- Emphasis on Stan for Bayesian computation
- No discussion of underlying HMC (black box?)

Probability and Bayesian Modeling

- What computational methods did we focus on in our book?
- Remember our students have had calculus.

Probability and Bayesian Modeling

- Illustrate Bayes for discrete models
- Conjugate priors (proportion and mean)
- Gibbs sampling and Metropolis algorithms
- JAGS for regression and multilevel models
- Don't mention HMC, but easy to learn Stan with this background

Closing Thoughts

- Think careful about learning objectives of course.
- Objectives include Bayesian foundational material.
- Choice of prior, Bayes sensitivity analysis, prediction, inference can be communicated using simple computational methods.
- Caution using “black-box” algorithms.

Reference

Albert , J. and Hu, J. (2020), "Bayesian Computing in the Undergraduate Statistics Curriculum," *Journal of Statistics and Data Science Education*.