A Quick Overview of Bayesian Inference

Jeff Dotson
Professor of Marketing
Marriott School of Business
Brigham Young University
www.occasionaldivergences.com
jpd@byu.edu

BYU
MARRIOTT SCHOOL

A brief tour of the wonderful world of Bayesian Inference!

1. A Quick Overview of Bayesian Inference

2. Learning About the Posterior: Modern Bayesian Computational Techniques

3. Priors, Likelihoods and Posteriors: A Simple Bayesian Workflow

4. Software Tools for Bayesian Inference

The posterior distribution:

$$p(\theta|\text{data}) \propto p(\text{data}|\theta)p(\theta)$$

$$\text{“posterior”} \propto \text{“likelihood”} \times \text{“prior”}$$

The posterior distribution reflects our revised beliefs about the model parameters after seeing the data.

The goal of Bayesian inference is to learn about the posterior distribution

$$\mathrm{p}(\theta|\mathrm{data}) \propto \mathrm{p}(\mathrm{data}|\theta)\mathrm{p}(\theta)$$

$$\text{“posterior”} \propto \text{“likelihood”} \times \text{“prior”}$$

If we specific a proper prior and have a well-defined likelihood, we are guaranteed that the <u>posterior distribution will exist</u>.

An example…

$$p(\theta|\text{data}) \propto p(\text{data}|\theta)p(\theta)$$

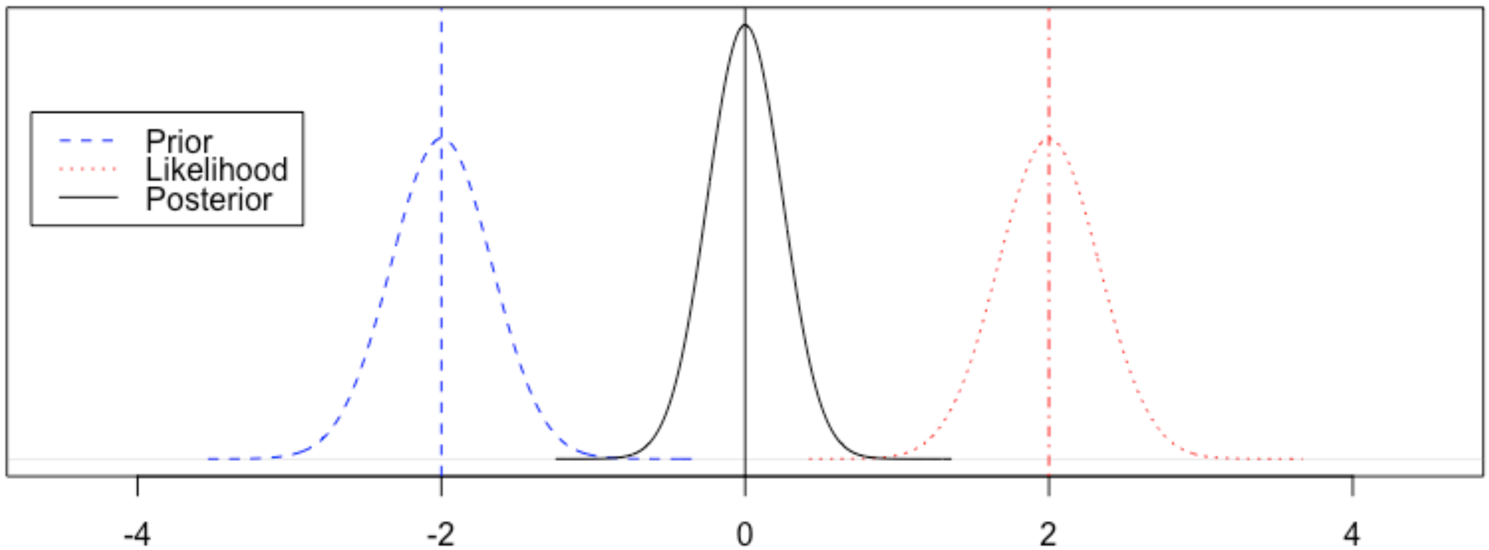
likelihood → $p(\text{data}|\theta) \sim N(\theta_x, \sigma_x^2)$

$p(\theta) \sim N(\theta_0, \sigma_0^2)$ ← **prior**

posterior

$$p(\theta|\text{data}) \sim N\left(\frac{\sigma_0^2}{\sigma_x^2+\sigma_0^2}\theta_x + \frac{\sigma_x^2}{\sigma_x^2+\sigma_0^2}\theta_0, \left(\frac{1}{\sigma_x^2}+\frac{1}{\sigma_0^2}\right)^{-1}\right)$$

An example…



Learning About the Posterior: Modern Bayesian Computational Techniques

BYU

MARRIOTT SCHOOL

How can we learn about the posterior?
Approach 1 – Conjugate Distributions

- If we are clever in how we specify our models and choose our priors, we can compute the posterior in closed form (i.e., a formula for the posterior)
- Conjugate distributions: normal-normal, beta-binomial, etc.
- Prior to the mid-1990's, this was the limit of most Bayesian inference

How can we learn about the posterior? Approach 2 - MCMC

The posterior distribution exists, even if it doesn't come from a know parametric family!

$$\mathrm{p}(\theta|\mathrm{data}) \propto \mathrm{p}(\mathrm{data}|\theta)\mathrm{p}(\theta)$$

$$\text{“posterior”} \propto \text{“likelihood”} \times \text{“prior”}$$

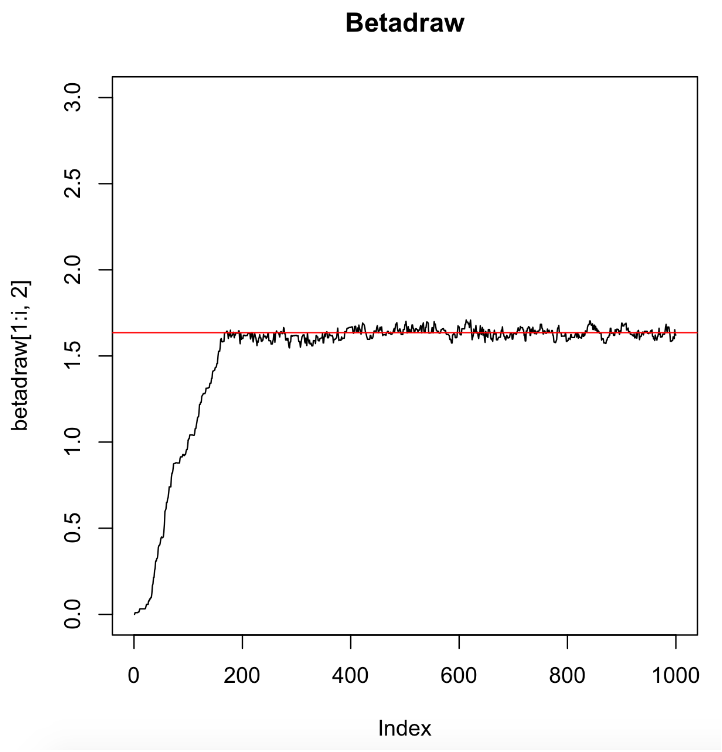
Key Principle: If we can find a way to **draw a random sample** from the posterior distribution, we can learn as much about it as we could if we knew the distribution in closed form.

How can we learn about the posterior? Approach 2 - MCMC

Modern computational Bayesian methods involve finding ways to draw random samples from probability distributions of known or unknown form.

If we can construct an empirical distribution of the parameters of our models, we can compute any quantities of interest, including point estimates and measures of uncertainty.

An Example: Quantile Regression



$$p(y|\beta,\sigma) \sim AL(X'\beta_p, p, \sigma)$$

$$p(\beta) \sim N(0,100I)$$

$$p(\sigma) \sim \text{Uniform}(0,100)$$

Key Points

1. We can learn as much about a distribution by drawing random samples as we could if we knew the distribution in closed form

2. The posterior distribution exists if we have a proper prior and likelihood

3. Many MCMC techniques that will allow us to sample from the posterior distribution

Priors, Likelihoods and Posteriors: A Simple Bayesian Workflow

A general workflow for Bayesian Modeling

1. Specify a model for your data
2. Set priors for all parameters in the model
3. Find a way to sample from the posterior distribution
4. Summarize the draws from the posterior distribution

Step 1: Specify a model for your data

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

Justify your choice of model!

Step 2. Set priors for model parameters

$$\beta \sim N(0, 100)$$

$$\sigma^2 \sim \text{inv}\,\chi^2(3)$$

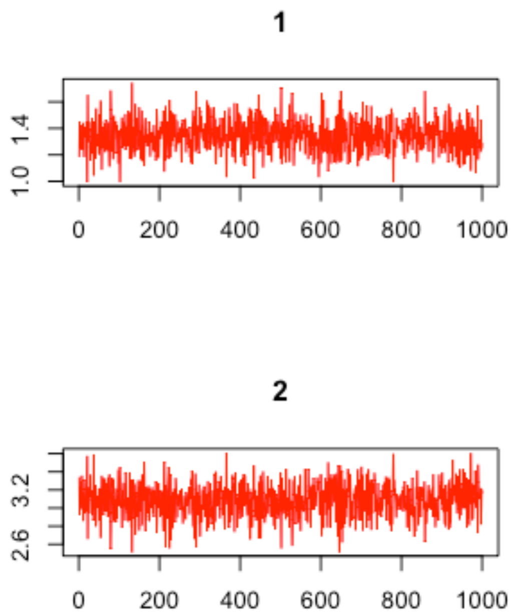
Justify your choice of prior!

How much influence does your prior have on inference?

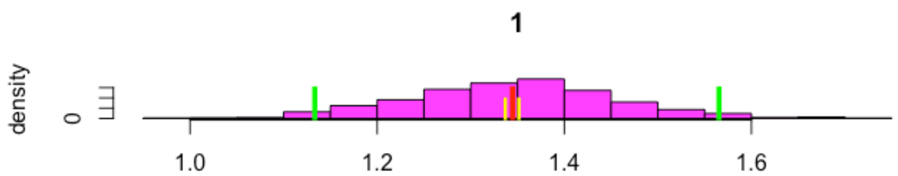
Use alternative priors to see if the results are robust to prior specification.

Step 3. Sample from the posterior distribution

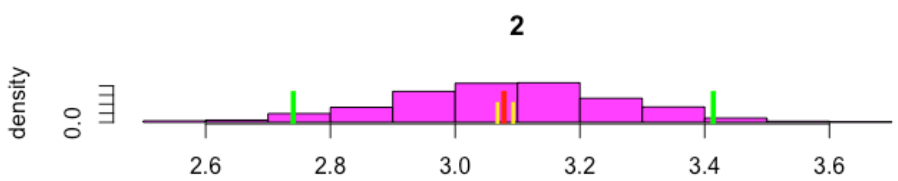
```r
library(bayesm)

n=100 #sample size
X=cbind(rep(1,n),runif(n)) #simulate X matrix
beta=c(1.5,3) #set regression coefficients
sigsq=.25 #set standard deviation

## Simulate data (y) given X, beta, sigma
y=X%*%beta+rnorm(n,sd=sqrt(sigsq))

## Create input objects required by bayesm
Data1=list(y=y,X=X)
Mcmc1=list(R=1000,keep=1)

## Run regression
out=runiregGibbs(Data=Data1,Mcmc=Mcmc1)

## Plot output
par(mfrow = c(1,2)) #plot beta and sigma side by side
matplot(out$betadraw,t="l")
abline(h=beta,col=c(1,2))
plot.ts(out$sigmasqdraw)
abline(h=sigsq)
```



Step 4. Summarize draws from the posterior



Mean: 1.3
Std Dev: 0.11
95% Interval: [1.1, 1.6]



Mean: 3.1
Std Dev: 0.18
95% Interval: [2.7, 3.4]

Summary

1. The Bayesian process: Model → Prior → Estimation → Inference

2. Justify the assumptions of your model

3. Justify your choice of prior

Software Tools for Bayesian Inference

Software Solutions: From Fixed to Flexible

Fixed ⟷ Flexible

Standard Models
(e.g., linear regression, logistic regression hierarchical linear models)

Custom Models
(e.g., hierarchical linear models with multinomial endogenous regressors)

Software Solutions: From Fixed to Flexible

Fixed ⟷ Flexible

Canned Routines

Code your Own

Bayes in STATA (great for beginners!)

Bayesian regression models using the **bayes** prefix

New in Stata release 15

Order Upgrade

What's this about?

Fitting Bayesian regression models can be just as intuitive as performing **Bayesian inference**—introducing the new **bayes** prefix in Stata. The **bayes** prefix combines Bayesian features with Stata's intuitive and elegant specification of regression models. It lets you fit Bayesian regression models more easily and fit more models.

You fit linear regression by using

```
. regress y x1 x2
```

You can now fit Bayesian linear regression by simply using

```
. bayes: regress y x1 x2
```

That is convenient, but you could fit Bayesian linear regression before. What you could not

Highlights

- Simply prefix your command with **bayes:**
- Over 50 likelihood models supported
- Linear, binary, ordinal, ...
- Count, zero inflated
- GLM, survival, multivariate
- Censoring, truncation, sample selection
- **Panel-data models**
- **Multilevel models**
- **Time-series operators**
- Full Bayesian-features support

Fixed Models – Hierarchical Linear Model

Flexible Models

Strategic Management Journal
Strat. Mgmt. J., **38**: 322–341 (2017)
Published online EarlyView 9 February 2016 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2480
Received 9 January 2012; Final revision received 20 October 2015

CORPORATE DIVERSIFICATION AND THE VALUE OF INDIVIDUAL FIRMS: A BAYESIAN APPROACH

TYSON B. MACKEY,[1*] JAY B. BARNEY,[1] and JEFFREY P. DOTSON[2]
[1] *Department of Entrepreneurship and Strategy, University of Utah, Salt Lake City, Utah, U.S.A.*
[2] *Marketing, Brigham Young University, Provo, Utah, U.S.A.*

Research summary: *Prior theory suggests that the performance effects of a firm's diversification strategy depend on a firm's individual resources and capabilities and the setting within which it is operating. However, prior tests of this theory have examined the average diversification-performance relationship across all firms, instead of estimating the diversification-performance relationship at the individual firm level. Efforts to estimate this average relationship are inconsistent with a central assumption of much of strategic management theory—that firms maximize value by choosing strategies that exploit their heterogeneous resources and individual situation. By adopting an approach that allows an evaluation of the diversification-performance relationship for individual firms, this article shows that firms, both focused and diversified, tend to choose that diversification strategy—focus, related diversification, or unrelated diversification—that maximizes value.*

Bayes in Stan (intermediate – advanced)

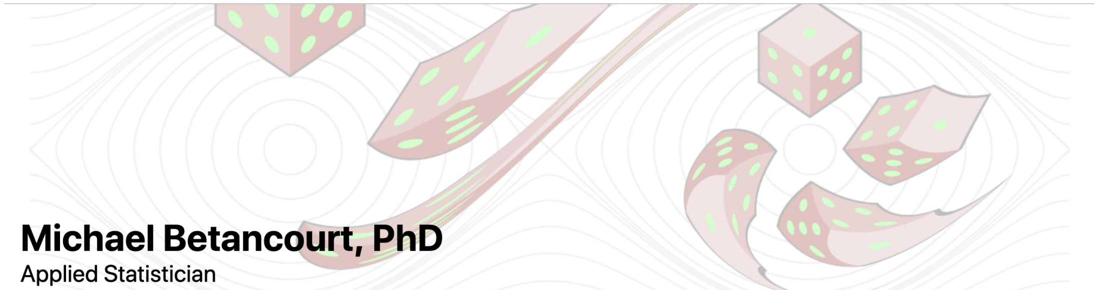
Stan

StanCon Cambridge UK, August 20-23, 2019

StanCon 2019 is coming up soon! There's still time to submit a poster: posters deadline is August 15. Register and more information here.

A Great Resource for Bayes and Stan

betanalpha.github.io Consulting Courses Speaking Writing



Long story short, I am a once and future physicist currently masquerading as a statistician in order to expose the secrets of inference that statisticians have long kept from scientists. More seriously, my research focuses on the development of robust statistical workflows, computational tools, and pedagogical resources that bridge statistical theory and practice and enable scientists to make the most out of their data.

The pursuit of general but scalable statistical computation has lead me to the intersection of differential geometry and probability theory where exploiting the inherent geometry of high-dimensional problems naturally leads to algorithms such Hamiltonian Monte Carlo and its generalizations. Along with some amazing colleagues I am developing both the theoretical foundations and the practical implementations of these algorithms, the latter specifically in the software ecosystem Stan.

If you would like to support this work consider

- hiring me to consult for your project,
- attending or commissioning a course,
- becoming a Patron.

For updates on courses, case studies, research, and more sign up for my mailing list.

Interested in doing similar research? Consider some of the open problems that I think would make impactful research projects.

Summary

1. There are lots of great software tools available for conducting Bayesian inference

2. For beginners, start by fitting simple models using the platform you are most familiar with (e.g., Stata, SAS, R)

3. When fitting more complicated models consider using a probabilistic programming system like Stan.