WZB

Social Science Research Center
Berlin

A Brief Introduction to Bayesian Statistics

Ryan Finnigan
ryan.finnigan@wzb.eu

Outline

- What is Bayesian Statistics, and What Are Its Advantages?
- A Little Bit of Math
- Estimation Methods
- Empirical Example

What’s the Difference?

- Frequentist statistics relies on repeated-sampling inference.

- Bayesian statistics estimates probability distributions for parameters.

- Bayesian methods also incorporate prior beliefs about the parameters of interest, and uncertainty about these beliefs.

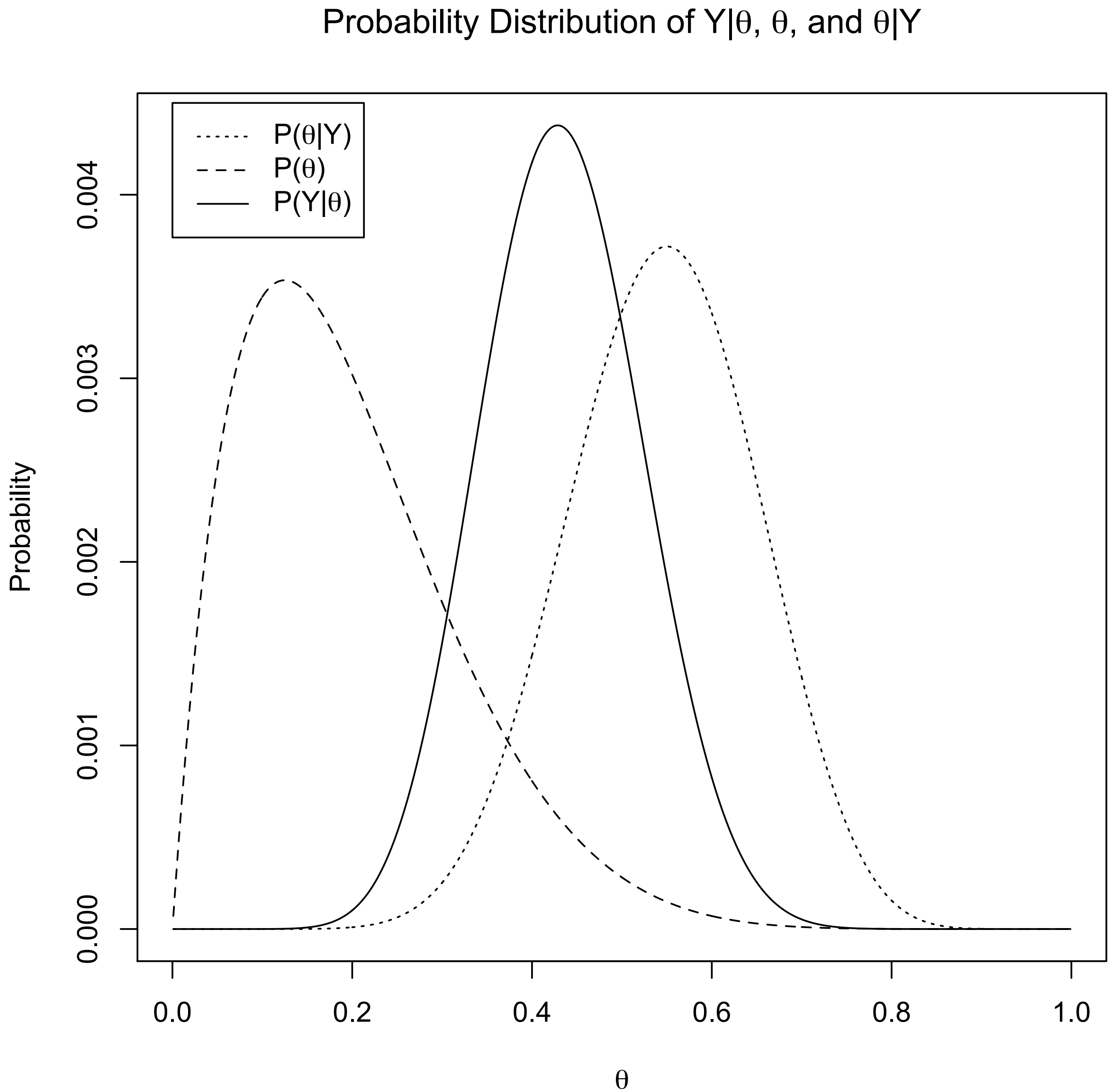
General Logic of Bayesian Statistics

- Researcher chooses a probability distribution to represent prior beliefs – $p(\theta)$

- Use data to evaluate the distribution – $p(Y|\theta)$

- Resulting posterior distribution estimates the parameters, and uncertainty – $p(\theta|Y)$

General Logic of Bayesian Statistics



Aren't the Priors Arbitrary?

Western's (1999) response:

1. α-levels and null hypotheses are also arbitrary.
2. Can choose non-informative priors.
3. Influence of the priors shrinks with increasing sample size.
4. Test the influence of a variety of priors.

The Math…

$$p(A, B) = p(A|B)p(B)$$

$$p(A, B) = p(B|A)p(A)$$

$$p(A|B)p(B) = p(B|A)p(A)$$

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

Just a Little More Math…

$$p(A) = \sum_B p(A, B)$$

$$p(A) = \sum_B p(A|B)p(B)$$

$$p(A) = \int_B p(A|B)p(B)dB$$

$$p(B|A) = \frac{p(A|B)p(B)}{\int_B p(A|B)p(B)dB}$$

Bayes Rule…Finally

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{\int_{\Theta} p(Y|\theta)p(\theta)d\theta}$$

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

Estimation by Simulation

- Generally select prior distributions with desirable properties, called conjugate priors.

- The posterior distribution is approximated with MCMC estimation

- Repeated samples are drawn from the posterior distributions, with interdependent parameters.

Mean and Variance Estimation

- Prior distributions

$$y|\theta,\sigma^2 \sim Norm(\theta,\sigma^2)$$

$$\theta|\sigma^2 \sim Norm(\mu_0,\sigma^2)$$

$$\sigma^2 \sim gamma^{-1}(\nu_0,\sigma_0^2)$$

- Posterior distributions

$$p(\theta,\sigma^2|Y) \propto p(Y|\theta,\sigma^2)p(\theta,\sigma^2)$$

$$p(\theta|\sigma^2,Y) \propto p(Y|\theta,\sigma^2)p(\sigma^2|\theta)p(\theta)$$

$$p(\sigma^2|\theta,Y) \propto p(Y|\theta,\sigma^2)p(\theta|\sigma^2)p(\sigma^2)$$

Markhov-Chain Monte Carlo

$$p(\theta^{(1)}|\sigma^{2(0)}, Y) \longrightarrow p(\sigma^{2(1)}|\theta^{(1)}, Y)$$

$$p(\theta^{(2)}|\sigma^{2(1)}, Y) \longrightarrow p(\sigma^{2(2)}|\theta^{(2)}, Y)$$

$$\Downarrow \qquad \Downarrow$$

$$p(\theta^{(s+1)}|\sigma^{2(s)}, Y) \longrightarrow p(\sigma^{2(s+1)}|\theta^{(s+1)}, Y)$$

Empirical Example

- Test the hypothesis that immigration undermines public support for social policy.
- ISSP data measuring percent of respondents in a country support for unemployment policy.
- Measure immigration with percent foreign born.
- Condition on social welfare expenditures as percent of total state spending.

OLS Regression Results

```
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.6192      1.4786   2.448    0.0282 *
% For. Born  -0.2634      0.2446  -1.077    0.2998
Soc. Expend.  0.3913      0.2446   1.600    0.1320
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
Residual standard error: 0.8888 on 14 degrees of
freedom
Multiple R-squared: 0.3087,Adjusted R-squared:  0.21
F-statistic: 3.126 on 2 and 14 DF,  p-value: 0.07543
```

Bayesian Estimate

$$p(\beta) \propto N(\beta_0, \Sigma_0)$$

$$p(\sigma^2) \propto gamma^{-1}\left(\frac{\nu_0}{2}, \frac{\nu_0^2\sigma_0^2}{2}\right)$$

$$p(y|X, \beta, \sigma^2) \propto N(X\beta, \sigma^2)$$

$$p(\beta|\sigma^2, Y) \propto p(Y|\beta, \sigma^2)p(\sigma^2|\beta)p(\beta)$$

$$p(\sigma^2|\beta, Y) \propto p(Y|\beta, \sigma^2)p(\beta|\sigma^2)p(\sigma^2)$$

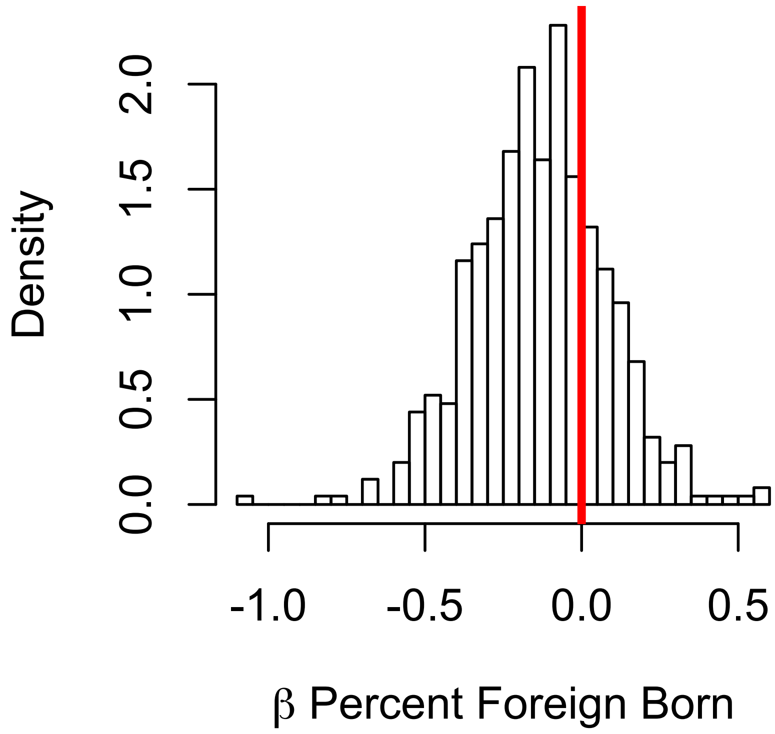
Bayesian Estimate with “Null” Prior

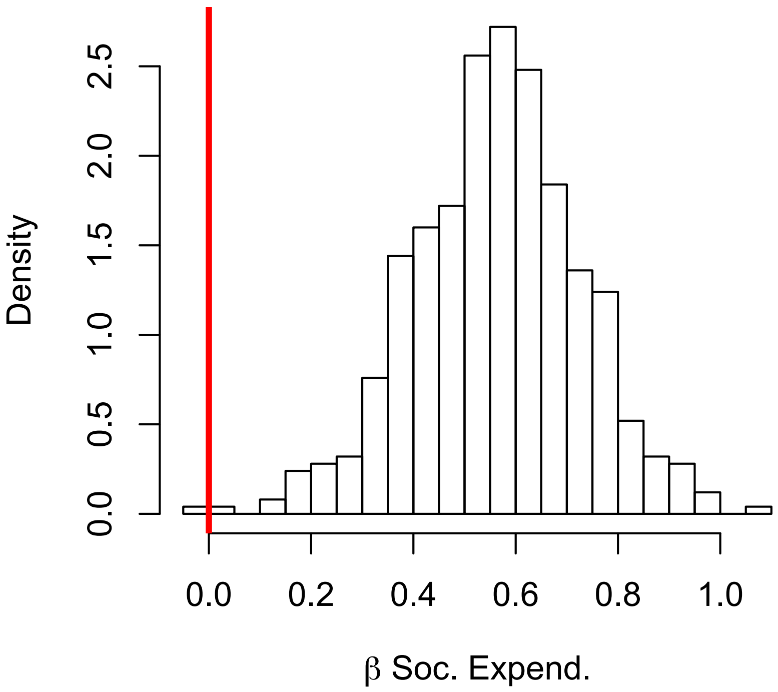
- Weakly assume the coefficients for immigration and social expenditures are 0.

- Assume the variances are relatively large.

- Assume wide residual variation.

Bayesian Results with “Null” Prior





Bayesian Results with “Null” Prior

- % Foreign born: -0.125 [-0.505, 0.302]
- p(b<0) = 0.714

- Soc. Expenditure: 0.572 [0.231, 0.865]
- p(b>0) = 0.998

- p(b_forborn > b_socx) = 0.042

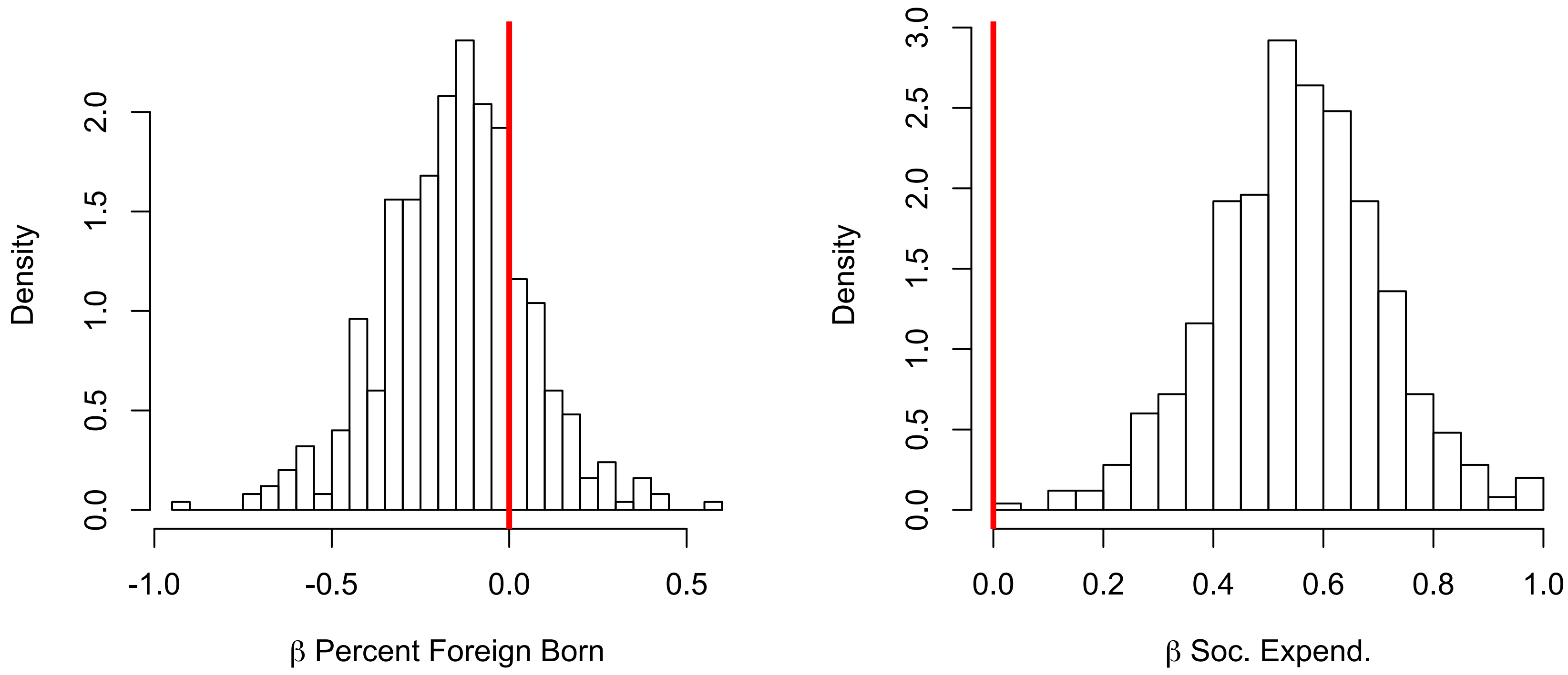
Bayesian Estimate with “Null” Prior

- Assume the coefficient for immigration is negative, with a small variance.

- Assume the coefficient for social expenditures is 0.

- Assume wide residual variation.

Bayesian Results with Negative Prior



Bayesian Results with Negative Prior

- % Foreign born: -0.141 [-0.591, 0.252]
- p(b<0) = 0.800

- Soc. Expenditure: 0.551 [0.246, 0.851]
- p(b>0) = 1.000

- p(b_forborn > b_socx) = 0.072

Further Applications

- Model Selection
- Multiple Imputation
- Non-Linear and Hierarchical Models

Useful References

- Western, Bruce. 1999. “Bayesian Analysis for Sociologists.” Sociological Methods and Research, 28:7-34.

- Hoff, Peter. 2009. A First Course in Bayesian Statistical Methods. New York: Springer.

- Gelman, Andrew, and Jennifer Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Social Science Research Center
Berlin

Thank you!

Ryan Finnigan

Gibbs Sampler R Code

```
# prior parameters
beta0 = c(2, 0, 0)
Sigma0 = matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 1), nrow=3,
ncol=3)
nu0 = 5
sigma20 = 1
sigma2 = sigma20
S = 1000
n = dim(X)[1]
sigma2_vector = c()
beta_matrix = c()
```

Gibbs Sampler R Code

```
for (s in 1:S) {
      # updating beta
      V = solve(solve(Sigma0) + (t(X)%*%X)/sigma2)
      m = solve(solve(Sigma0) + (t(X)%*%X)/sigma2)%*%
(t(Sigma0)%*%beta0 + (t(X)%*%y)/sigma2)
      beta = rmvnorm(1, m, V)

      # updating sigma2
      SSR = sum((y - X%*%t(beta))^2)
      sigma2 = 1/rgamma(1, .5*(nu0+n), .
5*(nu0*sigma20+SSR))

      # updating saved results
      sigma2_vector = c(sigma2_vector, sigma2)
      beta_matrix = rbind(beta_matrix, beta)
}
```