

Bayesian Estimation – An Introduction to Theory and Practice: Lecture XXXII

Charles B. Moss ¹

¹University of Florida

April 9, 2012

1 Readings

2 Bayes Theorem

- Bayesian Estimation – An Overview
- Conjugate Prior – Beta/Bernoulli
- Estimation – Expectation
- Computations of Beta/Bernoulli

3 Bayesian Regression Analysis

- Uninformative Prior
- Informative Prior
- Bayesian Estimation by Simulation

4 Reading – Next Time

Readings – Next Time

- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition.
(Chapter 16 pp. 655-680)
- * Albert, J. 2009. *Bayesian Computation with R*.

Bayes Theorem

- Theorem 2.3 [Moss p.17] Let Events $A_1, A_2, \dots, A_n$ be mutually exclusive events such that $P(A_1 \cup A_2 \cup \dots \cup A_n) = 1$ and $P(A_i) > 0$ for each i . Let E be an arbitrary event such that $P(E) > 0$. Then

$$P(A_i|E) = \frac{P(E|A) P(A_i)}{\sum_{j=1}^n P(E|A_j) P(A_j)}. \quad (1)$$

Another manifestation of this theorem is the marginal probability of event E

$$P(E, A_i) = P(E \cap A_i) = P(E|A_i) P(A_i) \quad (2)$$

Bayes Theorem – Continued

- The bottom of Equation 1 reduces the probability of event E

$$P(E) = \sum_{i=1}^n P(E|A_i) P(A_i) \quad (3)$$

Bayesian Estimation – An Overview

- Setting up the model, consider the Likelihood function of a sample of normal variables

$$L(X_1, X_2, \dots, X_n | \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right] \quad (4)$$

- The updated probability (posterior distribution) given a sample is then computed as

$$\pi(\theta | X) = \frac{L(X|\theta) \pi(\theta)}{\int_{-\infty}^{\infty} L(X|\theta) \pi(\theta) d\theta} \quad (5)$$

Conjugate Prior – Beta/Bernoulli

- As a starting point suppose we want to estimate the parameter of the Bernoulli distribution (P) [Moss pp. 113-116]

$$\pi(P) = P^X (1 - P)^{(1-X)} \ni X \in \{0, 1\} \quad (6)$$

- Next, we assume that our likelihood function for this parameter is based on the Beta distribution

$$f(P|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} P^{\alpha-1} (1 - P)^{\beta-1} \quad (7)$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Conjugate Prior – Continued

- The posterior for this conjugate can be expressed as

$$\begin{aligned}\pi(P|X) &= \frac{P^X (1-P)^{1-X} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} P^{\alpha-1} (1-P)^{\beta-1}}{\int_0^1 P^X (1-P)^{1-X} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} P^{\alpha-1} (1-P)^{\beta-1} dP} \\ &= \frac{P^{X+\alpha-1} (1-P)^{\beta-X}}{\int_0^1 P^{X+\alpha-1} (1-P)^{\beta-X} dP}\end{aligned}\tag{8}$$

- Computing the proper integral (probability)

$$\pi(P|X) = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(X+\alpha)\Gamma(\beta-X+1)} P^{X+\alpha-1} (1-P)^{\beta-X}\tag{9}$$

Estimation – Expectation

- The Bayesian estimate of P is then the value that minimizes a loss function – such as the quadratic loss function

$$\begin{aligned}\min_P E \left[(\hat{P} - P)^2 \right] &\Rightarrow \frac{\partial E \left[(\hat{P} - P)^2 \right]}{\partial \hat{P}} \\ &= 2E \left[\hat{P} - P \right] = 0 \Rightarrow \hat{P} = E[P]\end{aligned}\tag{10}$$

Computations of Beta/Bernoulli

- Chapter 2 in Albert *Bayesian Computation with R* uses the **R** package **LearnBayes** to estimate the parameters of a Bernoulli distribution.

```
p <- seq(0.05, 0.95, by=0.1)
prior <- c(2,4,8,8,4,2,1,1,1,1)
prior <- prior/sum(prior)
plot(p,prior,type='h',ylab='Prior Probability')
data <- c(11,16)
post <- pdisc(p,prior,data)
plot(p,post,type='h',ylab='Posterior Probability')
```

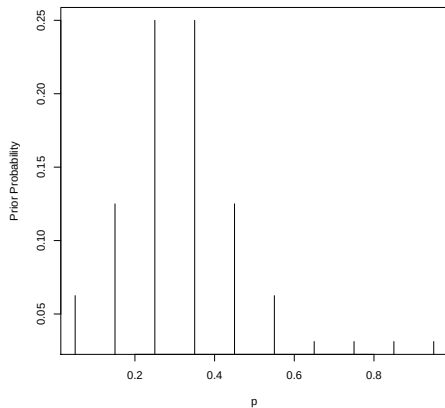


Figure: Prior Distribution

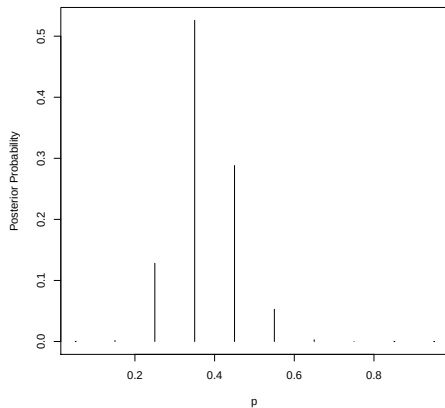


Figure: Posterior Distribution

Bayesian Regression Analysis

- Writing the standard normally distributed regression formulation as a Likelihood function

$$L(\beta, \sigma^2 | y, X) = [2\pi\sigma^2]^{-n/2} \exp \left[- (y - X\beta)' (y - X\beta) / 2\sigma^2 \right] \quad (11)$$

- Following Greene, let $d = n - K$ (the degrees of freedom) and substitute

$$y - X\beta = y - Xb - X(\beta - b) = e - X(\beta - b) \quad (12)$$

Bayesian Regression Analysis – Continued

- The term in the exponent then becomes

$$\begin{aligned} & \left(-\frac{1}{2\sigma^2}\right) (y - X\beta)' (y - X\beta) = \\ & \left(-\frac{1}{2}ds^2\right) \left(\frac{1}{\sigma^2}\right) - \frac{1}{2} (\beta - b)' \left(\frac{1}{\sigma^2} X'X\right) (\beta - b) \end{aligned} \quad (13)$$

- The Likelihood function can then be rewritten as

$$\begin{aligned} L(\beta, \sigma^2 | y, X) &= [2\pi]^{-d/2} [\sigma^2]^{-d/2} \exp \left[-\left(\frac{d}{2}\right) \left(\frac{s^2}{\sigma^2}\right) \right] \\ &\times [2\pi]^{-K/2} [\sigma^2]^{-K/2} \exp \left[-\frac{1}{2} (\beta - b)' [\sigma^2 (X'X)^{-1}]^{-1} (\beta - b) \right] \end{aligned} \quad (14)$$

Bayesian Regression Analysis – Continued

- Just as in Equation 9 we have to adjust the integral to become proper integral. In this case

$$A = \frac{\left(\frac{d}{2}s^2\right)^{(d/w)+1}}{\Gamma\left(\frac{d}{2} + 1\right)} [2\pi]^{d/2} |X'X|^{-1/2} \quad (15)$$

- Letting $v = d/2$ the proper Likelihood function becomes

$$L(\beta, \sigma^2 | y, X) \propto \frac{[vs^2]^{v+1}}{\Gamma(v+1)} \left(\frac{1}{\sigma^2}\right)^v \exp\left[-vs^2 \left(\frac{1}{\sigma^2}\right)\right] [2\pi]^{-K/2} \\ \times \left|\sigma^2 (X'X)^{-1}\right|^{-1/2} \exp\left[-\frac{1}{2}(\beta - b) \left[\sigma^2 (X'X)^{-1}\right]^{-1} (\beta - b)\right] \quad (16)$$

Bayesian Regression Analysis – Continued

- Looking at Equation 16 for any combination of β and σ^2 , $L(\beta, \sigma^2 | y, X)$ is the probability of drawing the sample of $\{y, X\}$.
- To complete the model, we need a prior distribution of β and/or σ^2 .
- There are two general forms of this prior information:
 - An uninformative prior $\Rightarrow g(\beta, \sigma^2) = \text{constant}$.
 - An informative prior (such as the Beta distribution above).

Uninformative Prior

- Greene derives the distribution of an uninformative prior given that the variance is known

$$g(\beta, \sigma^2) = g_\beta(\beta) g_{\sigma^2}(\sigma^2) \propto \frac{1}{\sigma^2} \quad (17)$$

- The posterior distribution then becomes

$$f(\beta|y, X) \propto \frac{\frac{[vs^2]^{\nu+2} \Gamma(\nu + K/2)}{\Gamma(\nu + 2)} [2\pi]^{K/2} |X'X|^{-1/2}}{\left[vs^2 + \frac{1}{2} (\beta - b)' X'X (\beta - b) \right]^{\nu+K/2}} \quad (18)$$

Informative Prior

- As a starting point in our development of an informative prior, consider a conjugate of three distributions in a linear regression:
 - The “sampling distribution” of the observed variables – Our standard normal distribution

$$f([y, X] | b, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{-T/2} \exp \left[\frac{1}{2\sigma^2} (y - Xb)' (y - Xb) \right] \quad (19)$$

- A distribution where the b (or regression coefficients) are normally distributed

$$g(b | \beta, \Omega) = \frac{1}{\sqrt{2\pi}} |\Omega|^{-1/2} \exp \left(-\frac{1}{2} (b - \beta)' \Omega^{-1} (b - \beta) \right) \quad (20)$$

Informative Prior – Continued

- Continued
 - Finally, assume that the variance is distributed Gamma

$$h(\sigma^2 | \alpha, \lambda) = \frac{1}{\Gamma(\alpha) \lambda^\alpha} (\sigma^2)^{\alpha-1} \exp\left(-\frac{\sigma^2}{\lambda}\right) \quad (21)$$

- Focusing on the exponential terms

$$f(.) g(.) h(.) \propto \exp \left[-\frac{1}{2} \left([y - Xb]' [I_{T \times T} \sigma^2]^{-1} [y - Xb] + [b - \beta]' \Omega^{-1} [b - \beta] + \frac{2\sigma^2}{\lambda} \right) \right] \quad (22)$$

Informative Prior – Continued

- Focusing on the mean (or the b and β terms), consider that

$$\min_b E [(b - \beta)' (b - \beta)] \Rightarrow \quad (23)$$

$$\min_b [y - Xb]' [I_{T \times T} \sigma^2] [y - Xb] + [b - \beta]' \Omega^{-1} [b - \beta]$$

- The solution of this minimization problem is similar to the standard Ordinary Least Squares formulation

$$\left[\begin{pmatrix} y \\ \beta \end{pmatrix} - \begin{pmatrix} X \\ I_{k \times k} \end{pmatrix} b \right] \begin{bmatrix} \sigma^2 I_{T \times T} & 0 \\ 0 & \Omega \end{bmatrix}^{-1} \left[\begin{pmatrix} y \\ \beta \end{pmatrix} - \begin{pmatrix} X \\ I_{k \times k} \end{pmatrix} b \right] \quad (24)$$

Informative Prior – Continued

- Solving the minimization problem

$$b_{Bayes} = [\psi X'X + (1 - \psi) I]^{-1} [\psi X'y + (1 - \psi) \beta] \quad (25)$$

- This is the Theil-Goldberger Mixed Estimator.
 - Note if $\psi \rightarrow 0$ the estimator is simply the prior.
 - Note if $\psi \rightarrow 1$ the estimator is simply ordinary least squares.

Bayesian Estimation by Simulation

- Recalling from last lecture, if

$$L(b|\beta, \Omega, \alpha, \lambda) = \frac{f([y, X]|b, \sigma^2) g(b|\beta, \Omega) h(\sigma^2|\alpha\lambda)}{\int_{-\infty}^{\infty} \cdots \int_0^{\infty} f([y, X]|b, \sigma^2) g(b|\beta, \Omega) h(\sigma^2|\alpha\lambda) d\beta_0 \cdots d\lambda} \quad (26)$$

$$\begin{bmatrix} \hat{\beta} \\ \text{vec}(\hat{\Omega}) \\ \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} = \int_{-\infty}^{\infty} \cdots \int_0^{\infty} \begin{bmatrix} \beta \\ \text{vec}(\Omega) \\ \alpha \\ \lambda \end{bmatrix} L(b|\beta, \Omega, \alpha, \lambda) d\beta_0 \cdots d\lambda \quad (27)$$

- The system can be solved by a double integration by simulation.

Reading – Next Time

- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 14).
- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition (Chapter 17 [Sections 17.1 17.3 pp. 681 714])
- * Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press. (Chapters 1 and 2 pp. 1-56)