

Lecture I: Introduction to Bayesian Econometrics

Charles B. Moss

May 15, 2013

- 1 Overview
 - Normality
 - Conditioned on a Model – A
- 2 Conditioning and Updating
 - Share Equations
 - Loss Function – Bayesian Estimation

Overview

- Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, New Jersey: John Wiley & Sons – Chapter 1.
- Basic Model
 - An empirical or statistical model used for research purposes typically assumes that our observations are functions of unobservable parameters. Mathematically,

$$p(y|\theta) \tag{1}$$

where $p(\cdot)$ is the probability of y – a set of observable outcomes (i.e., crop yields) and θ is a vector of unobservable parameters (or latent variables).

Normality

- Under normality $p(y | \mu, \sigma^2)$ means that the probability an observed outcome ($y = 50.0$) is a function of unobserved parameters such as the mean of the normal (μ) and its variance (σ^2).
- The formulation in Equation 1 is for any general outcome (i.e., any potential level of y). It is important to distinguish between general outcome – y and a specific outcome that we observe – y^0 . In general, we typically refer to a specific outcome as "data".
- If we do not know what the value of θ is, we can depict the density function for θ as $p(\theta)$. We can then combine the density function for θ ($p(\theta)$) with our formulation of the probability of the observables ($p(y | \theta)$) to produce information about the observables that are not conditioned on knowing the value of θ

$$p(y) = \int p(\theta) p(y | \theta) d\theta \quad (2)$$

Conditioned on a Model – A

- Next, we index our relation between observables and unobservables as model A . Hence, $p(y|\theta)$ becomes $p(y|\theta_A, A)$, and $p(\theta)$ becomes $p(\theta_A|A)$.
- We shall denote the object of interest on which decision making depends, and which all models relevant to the decision have something to say, by the vector ω . We shall denote the implications of the model A for ω by $p(\omega|y, \theta_A, A)$.
- In summary, we have identified three components of a complete model, A , involving unobservables (often parameters) θ_A , observables y , and a vector of interest ω .

Conditioning and Updating

- Given this setup, we can sketch out the way we build on information from our prior beliefs about the unobservable parameters using sample information.
 - As a starting point, we denote the prior distribution as $p(\theta_A | A)$. This captures our initial intuition about the probability of the unobservable parameters (i.e., the mean and variance of the distribution). These expectations are conditioned on a model – A (i.e., normality).
 - In part, these prior beliefs specify a distribution for the observable variables $p(y | \theta_A, A)$.
 - Given these two pieces of information we can derive the probability of the unobservable parameters based on an observed sample (y^0)

$$p(\theta_A | y^0, A) = \frac{p(\theta_A, y^0 | A)}{p(y^0 | A)} = \frac{p(\theta_A | A) p(y^0 | \theta_A, A)}{p(y^0 | A)} \quad (3)$$

where $p(\theta_A | y^0, A)$ is referred to as the posterior distribution.

- As a simple example, assume that we are interested in estimating a simple mean for a normal distribution. As a starting point, assume that our prior distribution for the mean is a normal distribution with a mean of $\tilde{\mu}$ and a known variance of 1. Our prior distribution then becomes

$$f_1(\mu | \tilde{\mu}, 1) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(\mu - \tilde{\mu})^2}{2} \right]. \quad (4)$$

Next, assume that the outcome is normally distributed with a mean of μ from Equation 4

$$f_2(y | \mu, k) = \frac{1}{\sqrt{2\pi k}} \exp \left[-\frac{(y - \mu)^2}{2k} \right] \quad (5)$$

where k some fixed variance.

- The posterior distribution then becomes

$$\begin{aligned}
 f(\mu | y^0, k) &\propto f_1(\mu | \tilde{\mu}, 1) f_2(y^0 | \mu, k) \\
 &= \frac{1}{2\pi\sqrt{k}} \exp \left[-\frac{(\mu - \tilde{\mu})^2}{2} - \frac{(y^0 - \mu)^2}{2k} \right]
 \end{aligned} \tag{6}$$

(ignoring for the moment the denominator of Equation 4).

Share Equations

- One application of this formulation is to assume that we are interested in the share of cost associated with one input (say the constant in for a share equation in a Translog cost function). If there are four inputs, we could assume *a priori* (our prior guess) is that the average share for any one input would be 0.25. Hence, we set $\tilde{\mu} = 0.25$ in Equation 6. The prior distribution for any y^0 would then become

$$f(\mu | y^0, k) \propto \frac{1}{2\pi\sqrt{k}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(y^0 - \mu)^2}{2k} \right]. \quad (7)$$

Next, assume that we have a single draw from the sample $y^0 = 0.1532$ – the empirical posterior distribution would then become

$$f(\mu | y^0 = 0.1532, k) \propto \frac{1}{2\pi\sqrt{k}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(0.1532 - \mu)^2}{2k} \right]. \quad (8)$$

- Returning to the denominator from Equation 4, $p(y^0 | A)$ is the probability of drawing the observed sample unconditional on the value of θ . In this case, unconditioned on the value of μ . Mathematically,

$$f(0.1532) =$$

$$\int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sqrt{k}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(0.1532 - \mu)^2}{2k} \right] \right) d\mu = 0.28144. \quad (9)$$

(computed using Mathematica). The complete posterior distribution can then be written as

$$f(\mu | y^0 = 0.1532, k) = \frac{\frac{1}{2\pi\sqrt{k}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(0.1532 - \mu)^2}{2k} \right]}{0.28144}. \quad (10)$$

Loss Function – Bayesian Estimation

- One use of the posterior distribution in Equation 10 is to compute a Bayesian estimate of μ . This is typically accomplished by minimizing the loss function

$$\begin{aligned} \text{Min}_{\mu_B} L(\mu, \mu_B) &\equiv E[\mu - \mu_B]^2 \Rightarrow \mu_B = E[\mu] \\ \Rightarrow E[\mu] &= \int_{-\infty}^{\infty} f(\mu | y^0 = 0.1532, k) \mu d\mu = 0.2016 \end{aligned} \quad (11)$$

again relying on Mathematica for the numeric integral.

- Next, consider the scenario where we have two sample points $y^0 = \{0.1532, 0.1620\}$. In this case the sample distribution becomes

$$f_2(y^0 | \mu, k) = \frac{1}{\sqrt{2\pi k^2}} \exp \left[-\frac{(0.1532 - \mu)^2 + (0.1620 - \mu)^2}{2k} \right] \quad (12)$$

The posterior distribution can then be derived based on

$$f(\mu | y^0, k) \propto \frac{1}{2\pi\sqrt{k^2}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(0.1532 - \mu)^2 + (0.1620 - \mu)^2}{2k} \right]. \quad (13)$$

- Integrating the denominator numerically yields 0.2300. Hence, the posterior distribution becomes

$$f(\mu | y^0, k) = \frac{\frac{1}{2\pi\sqrt{k^2}} \exp \left[-\frac{(\mu - 0.25)^2}{2} - \frac{(0.1532 - \mu)^2 + (0.1620 - \mu)^2}{2k} \right]}{0.2300} \quad (14)$$

which yields a Bayesian estimate for μ of 0.1884. Note that this estimator is much higher than the standard mean of 0.1576 – the estimate is biased upward by prior.