1 An Overview of Bayesian Econometrics

1.1 BAYESIAN THEORY

Bayesian econometrics is based on a few simple rules of probability. This is one of the chief advantages of the Bayesian approach. All of the things that an econometrician would wish to do, such as estimate the parameters of a model, compare different models or obtain predictions from a model, involve the same rules of probability. Bayesian methods are, thus, universal and can be used any time a researcher is interested in using data to learn about a phenomenon.

To motivate the simplicity of the Bayesian approach, let us consider two random variables, A and B.[1] The rules of probability imply:

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

where $p(A, B)$ is the *joint probability*[2] of A and B occurring, $p(A|B)$ is the probability of A occurring conditional on B having occurred (i.e. the *conditional probability* of A given B), and $p(B)$ is the *marginal probability* of B. Alternatively, we can reverse the roles of A and B and find an expression for the joint probability of A and B:

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

Equating these two expressions for $p(A, B)$ and rearranging provides us with *Bayes' rule*, which lies at the heart of Bayesian econometrics:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} \tag{1.1}$$

[1]This chapter assumes the reader knows the basic rules of probability. Appendix B provides a brief introduction to probability for the reader who does not have such a background or would like a reminder of this material.

[2]We are being slightly sloppy with terminology here and in the following material in that we should always say 'probability density' if the random variable is continuous and 'probability function' if the random variable is discrete (see Appendix B). For simplicity, we simply drop the word 'density' or 'function'.

Econometrics is concerned with using data to learn about something the researcher is interested in. Just what the 'something' is depends upon the context. However, in economics we typically work with models which depend upon parameters. For the reader with some previous training in econometrics, it might be useful to have in mind the regression model. In this model interest often centers on the coefficients in the regression, and the researcher is interested in estimating these coefficients. In this case, the coefficients are the parameters under study. Let y be a vector or matrix of data and θ be a vector or matrix which contains the parameters for a model which seeks to explain y.[3] We are interested in learning about θ based on the data, y. Bayesian econometrics uses Bayes' rule to do so. In other words, the Bayesian would replace B by θ and A by y in (1.1) to obtain:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{1.2}$$

Bayesians treat $p(\theta|y)$ as being of fundamental interest. That is, it directly addresses the question "Given the data, what do we know about θ?". The treatment of θ as a random variable is controversial among some econometricians. The chief competitor to Bayesian econometrics, often called *frequentist econometrics*, says that θ is not a random variable. However, Bayesian econometrics is based on a subjective view of probability, which argues that our uncertainty about anything unknown can be expressed using the rules of probability. In this book, we will not discuss such methodological issues (see Poirier (1995) for more detail). Rather, we will take it as given that econometrics involves learning about something unknown (e.g. coefficients in a regression) given something known (e.g. data) and the conditional probability of the unknown given the known is the best way of summarizing what we have learned.

Having established that $p(\theta|y)$ is of fundamental interest for the econometrician interested in using data to learn about parameters in a model, let us now return to (1.2). Insofar as we are only interested in learning about θ, we can ignore the term $p(y)$, since it does not involve θ. We can then write:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{1.3}$$

The term $p(\theta|y)$ is referred to as the *posterior density*, the p.d.f. for the data given the parameters of the model, $p(y|\theta)$, as the *likelihood function* and $p(\theta)$ as the *prior density*. You often hear this relationship referred to as "posterior is proportional to likelihood times prior". At this stage, this may seem a little abstract, and the manner in which priors and likelihoods are developed to allow for the calculation of the posterior may be unclear. Things should become clearer to you in the following chapters, where we will develop likelihood functions and priors in specific contexts. Here we provide only a brief general discussion of what these are.

[3]Appendix A contains a brief introduction to matrix algebra.

The prior, $p(\theta)$, does not depend upon the data. Accordingly, it contains any non-data information available about θ. In other words, it summarizes what you know about θ prior to seeing the data. As an example, suppose θ is a parameter which reflects returns to scale in a production process. In many cases, it is reasonable to assume that returns to scale are roughly constant. Thus, before you look at the data, you have prior information about θ, in that you would expect it to be approximately one. Prior information is a controversial aspect of Bayesian methods. In this book, we will discuss both informative and noninformative priors for various models. In addition, in later chapters, we will discuss empirical Bayes methods. These use data-based information to choose the prior and, hence, violate a basic premise of Bayesian methods. Nevertheless, empirical Bayes methods are becoming increasingly popular for the researcher who is interested in practical, objective, tools that seem to work well in practice.[4]

The likelihood function, $p(y|\theta)$, is the density of the data conditional on the parameters of the model. It is often referred to as the data generating process. For instance, in the linear regression model (which will be discussed in the next chapter), it is common to assume that the errors have a Normal distribution. This implies that $p(y|\theta)$ is a Normal density, which depends upon parameters (i.e. the regression coefficients and the error variance).

The posterior, $p(\theta|y)$, is the density which is of fundamental interest. It summarizes all we know about θ after (i.e. posterior to) seeing the data. Equation (1.3) can be thought of as an updating rule, where the data allows us to update our prior views about θ. The result is the posterior which combines both data and non-data information.

In addition to learning about parameters of a model, an econometrician might be interested in comparing different models. A model is formally defined by a likelihood function and a prior. Suppose we have m different models, M_i for $i = 1, \ldots, m$, which all seek to explain y. M_i depends upon parameters θ^i. In cases where many models are being entertained, it is important to be explicit about which model is under consideration. Hence, the posterior for the parameters calculated using M_i is written as

$$p(\theta^i|y, M_i) = \frac{p(y|\theta^i, M_i)p(\theta^i|M_i)}{p(y|M_i)} \tag{1.4}$$

and the notation makes clear that we now have a posterior, likelihood, and prior for each model.

The logic of Bayesian econometrics suggests that we use Bayes' rule to derive a probability statement about what we do not know (i.e. whether a model is a correct one or not) conditional on what we do know (i.e. the data). This means the *posterior model probability* can be used to assess the degree of support for

[4]Carlin and Louis (2000) is a good reference for the reader interested in developing a deeper understanding of empirical Bayes methods.

M_i. Using (1.1) with $B = M_i$ and $A = y$, we obtain

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)} \tag{1.5}$$

Of the terms in (1.5), $p(M_i)$ is referred to as the *prior model probability*. Since it does not involve the data, it measures how likely we believe M_i to be the correct one before seeing the data. $p(y|M_i)$ is called the *marginal likelihood*, and is calculated using (1.4) and a few simple manipulations. In particular, if we integrate both sides of (1.4) with respect to θ^i, use the fact that $\int p(\theta^i|y, M_i)d\theta^i = 1$ (since probability density functions integrate to one), and rearrange, we obtain:

$$p(y|M_i) = \int p(y|\theta^i, M_i)p(\theta^i|M_i)d\theta^i \tag{1.6}$$

Note that the marginal likelihood depends only upon the prior and the likelihood. In subsequent chapters, we discuss how (1.6) can be calculated in practice.

Since the denominator in (1.5) is often hard to calculate directly, it is common to compare two models, i and j, using the *posterior odds ratio*, which is simply the ratio of their posterior model probabilities:

$$PO_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)p(M_i)}{p(y|M_j)p(M_j)} \tag{1.7}$$

Note that, since $p(y)$ is common to both models, it cancels out when we take the ratio. As we will discuss in subsequent chapters, there are special techniques in many cases for calculating the posterior odds ratio directly. If we calculate the posterior odds ratio comparing every pair of models, and we assume that our set of models is exhaustive (in that $p(M_1|y) + p(M_2|y) + \cdots + p(M_m|y) = 1$), then we can use posterior odds ratios to calculate the posterior model probabilities given in (1.5). For instance, if we have $m = 2$ models then we can use the two equations

$$p(M_1|y) + p(M_2|y) = 1$$

and

$$PO_{12} = \frac{p(M_1|y)}{p(M_2|y)}$$

to work out

$$p(M_1|y) = \frac{PO_{12}}{1 + PO_{12}}$$

and

$$p(M_2|y) = 1 - p(M_1|y)$$

Thus, knowledge of the posterior odds ratio allows us to figure out the posterior model probabilities.

To introduce some more jargon, econometricians may be interested in model comparison when equal prior weight is attached to each model. That is, $p(M_i) = p(M_j)$ or, equivalently, the *prior odds ratio* which is $\frac{p(M_i)}{p(M_j)}$ is set to one. In this case, the posterior odds ratio becomes simply the ratio of marginal likelihoods, and is given a special name, the *Bayes Factor,* defined as:

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)} \tag{1.8}$$

Finally, econometricians are often interested in prediction. That is, given the observed data, y, the econometrician may be interested in predicting some future unobserved data y^*. Our Bayesian reasoning says that we should summarize our uncertainty about what we do not know (i.e. y^*) through a conditional probability statement. That is, prediction should be based on the *predictive density* $p(y^*|y)$ (or, if we have many models, we would want to make explicit the dependence of a prediction on a particular model, and write $p(y^*|y, M_i)$). Using a few simple rules of probability, we can write $p(y|y^*)$ in a convenient form. In particular, since a marginal density can be obtained from a joint density through integration (see Appendix B), we can write:

$$p(y^*|y) = \int p(y^*, \theta|y)d\theta$$

However, the term inside the integral can be rewritten using another simple rule of probability:

$$p(y^*|y) = \int p(y^*|y, \theta)p(\theta|y)d\theta \tag{1.9}$$

As we shall see in future chapters, the form for the predictive in (1.9) is quite convenient, since it involves the posterior.

On one level, this book could end right here. These few pages have outlined all the basic theoretical concepts required for the Bayesian to learn about parameters, compare models and predict. We stress what an enormous advantage this is. Once you accept that unknown things (i.e. θ, M_i and y^*) are random variables, the rest of Bayesian approach is non-controversial. It simply uses the rules of probability, which are mathematically true, to carry out statistical inference. A benefit of this is that, if you keep these simple rules in mind, it is hard to lose sight of the big picture. When facing a new model (or reading a new chapter in the book), just remember that Bayesian econometrics requires selection of a prior and a likelihood. These can then be used to form the posterior, (1.3), which forms the basis for all inference about unknown parameters in a model. If you have many models and are interested in comparing them, you can use posterior model probabilities (1.5), posterior odds ratios (1.7), or Bayes Factors (1.8). To obtain any of these, we usually have to calculate the marginal likelihood (1.6). Prediction is done through the predictive density, $p(y^*|y)$, which is usually calculated using (1.9). These few equations can be used to carry out statistical inference in *any* application you may wish to consider.

The rest of this book can be thought of as simply examples of how (1.5)–(1.9) can be used to carry out Bayesian inference for various models which have been commonly-used by others. Nevertheless, we stress that Bayesian inference can be done with any model using the techniques outlined above and, when confronting an empirical problem, you should not necessarily feel constrained to work with one of the off-the-shelf models described in this book.

1.2 BAYESIAN COMPUTATION

The theoretical and conceptual elegance of the Bayesian approach has made it an attractive one for many decades. However, until recently, Bayesians have been in a distinct minority in the field of econometrics, which has been dominated by the frequentist approach. There are two main reasons for this: prior information and computation. With regards to the former, many researchers object to the use of ‘subjective’ prior information in the supposedly ‘objective’ science of economics. There is a long, at times philosophical, debate about the role of prior information in statistical science, and the present book is not the place to attempt to summarize this debate. The interested reader is referred to Poirier (1995), which provides a deeper discussion of this issue and includes an extensive bibliography. Briefly, most Bayesians would argue that the entire model building process can involve an enormous amount of non-data information (e.g. econometricians must decide which models to work with, which variables to include, what criteria to use to compare models or estimate parameters, which empirical results to report, etc.). The Bayesian approach is honest and rigorous about precisely how such non-data information is used. Furthermore, if prior information is available, it should be used on the grounds that more information is preferred to less. As a final line of defense, Bayesians have developed *noninformative priors* for many classes of model. That is, the Bayesian approach allows for the use of prior information if you wish to use it. However, if you do not wish to use it, you do not have to do so. Regardless of how a researcher feels about prior information, it should in no way be an obstacle to the adoption of Bayesian methods.

Computation is the second, and historically more substantive, reason for the minority status of Bayesian econometrics. That is, Bayesian econometrics has historically been computationally difficult or impossible to do for all but a few specific classes of model. The computing revolution of the last 20 years has overcome this hurdle and has led to a blossoming of Bayesian methods in many fields. However, this has made Bayesian econometrics a field which makes heavy use of the computer, and a great deal of this book is devoted to a discussion of computation. In essence, the ideas of Bayesian econometrics are simple, since they only involve the rules of probability. However, to use Bayesian econometrics in practice often requires a lot of number crunching.

To see why computational issues are so important, let us return to the basic equations which underpin Bayesian econometrics. The equations relating to model comparison and prediction either directly or indirectly involve integrals (i.e. (1.6) and (1.9) involve integrals, and (1.6) is a building block for (1.7) and (1.8)). In some (rare) cases, analytical solutions for these integrals are available. That is, you can sit down with pen and paper and work out the integrals. However, we usually need the computer to evaluate the integrals for us, and many algorithms for doing so have been developed.

The equation defining the posterior does not involve any integrals, but presentation of information about the parameters can often involve substantial computation. This arises since, although $p(\theta|y)$ summarizes all we know about θ after seeing the data, it is rarely possible to present all the information about $p(\theta|y)$ when writing up a paper. In cases where $p(\theta|y)$ has a simple form or θ is one-dimensional, it is possible to do so, for instance, by graphing the posterior density. However, in general, econometricians choose to present various numerical summaries of the information contained in the posterior, and these can involve integration. For instance, it is common to present a point estimate, or best guess, of what θ is. Bayesians typically use decision theory to justify a particular choice of a point estimate. In this book, we will not discuss decision theory. The reader is referred to Poirier (1995) or Berger (1985) for excellent discussions of this topic (see also Exercise 1 below). Suffice it to note here that various intuitively plausible point estimates such as the mean, median, and mode of the posterior can be justified in a decision theoretical framework.

Let us suppose you want to use the mean of the posterior density (or *posterior mean*) as a point estimate, and suppose θ is a vector with k elements, $\theta = (\theta_1, \ldots, \theta_k)'$. The posterior mean of any element of θ is calculated as (see Appendix B)

$$E(\theta_i|y) = \int \theta_i p(\theta|y) d\theta \tag{1.10}$$

Apart from a few simple cases, it is not possible to evaluate this integral analytically, and once again we must turn to the computer.

In addition to a point estimate, it is usually desirable to present a measure of the degree of uncertainty associated with the point estimate. The most common such measure is the *posterior standard deviation*, which is the square root of the *posterior variance*. The latter is calculated as

$$var(\theta_i|y) = E(\theta_i^2|y) - \{E(\theta_i|y)\}^2$$

which requires evaluation of the integral in (1.10), as well as

$$E(\theta_i^2|y) = \int \theta_i^2 p(\theta|y) d\theta$$

Depending on the context, the econometrician may wish to present many other features of the posterior. For instance, interest may center on whether a particular parameter is positive. In this case, the econometrician would calculate

$$p(\theta_i \geq 0|y) = \int_0^\infty p(\theta|y) d\theta$$

and, once again, an integral is involved.

All of these posterior features which the Bayesian may wish to calculate have the form:

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta \tag{1.11}$$

where $g(\theta)$ is a *function of interest*. For instance, $g(\theta) = \theta_i$ when calculating the posterior mean of θ_i and $g(\theta) = 1(\theta_i \geq 0)$ when calculating the probability that θ_i is positive, where $1(A)$ is the indicator function which equals 1 if condition A holds and equals zero otherwise. Even the predictive density in (1.9) falls in this framework if we set $g(\theta) = p(y^*|y, \theta)$. Thus, most things a Bayesian would want to calculate can be put in the form (1.11). The chief exceptions which do not have this form are the marginal likelihood and quantiles of the posterior density (e.g. in some cases, one may wish to calculate the posterior median and posterior interquartile range, and these cannot be put in the form of (1.11)). These exceptions will be discussed in the context of particular models in subsequent chapters.

At this point, a word of warning is called for. Throughout this book, we focus on evaluating $E[g(\theta)|y]$ for various choices of $g(.)$. Unless otherwise noted, for every model and $g(.)$ discussed in this book, $E[g(\theta)|y]$ exists. However, for some models it is possible that $E[g(\theta)|y]$ does not exist. For instance, for the Cauchy distribution, which is the t distribution with one degree of freedom (see Appendix B, Definition B.26), the mean does not exist. Hence, if we had a model which had a Cauchy posterior distribution, $E[\theta|y]$ would not exist. When developing methods for Bayesian inference in a new model, it is thus important to prove that $E[g(\theta)|y]$ does exist. Provided that $p(\theta|y)$ is a valid probability density function, quantiles will exist. So, if you are unsure that $E[g(\theta)|y]$ exists, you can always present quantile-based information (e.g. the median and interquartile range).

In rare cases, (1.11) can be worked out analytically. However, in general, we must use the computer to calculate (1.11). There are many methods for doing this, but the predominant approach in modern Bayesian econometrics is *posterior simulation*. There are a myriad of posterior simulators which are commonly used in Bayesian econometrics, and many of these will be discussed in future chapters in the context of particular models. However, all these are applications or extensions of *laws of large numbers* or *central limit theorems*. In this book, we do not discuss these concepts of *asymptotic distribution theory* in any detail. The interested reader is referred to Poirier (1995) or Greene (2000). Appendix B provides some simple cases, and these can serve to illustrate the basic ideas of posterior simulation.

A straightforward implication of the law of large numbers given in Appendix B (see Definition B.31 and Theorem B.19) is:

Theorem 1.1: Monte Carlo integration

Let $\theta^{(s)}$ for $s = 1, \ldots, S$ be a random sample from $p(\theta|y)$, and define

$$\widehat{g}_S = \frac{1}{S}\sum_{s=1}^{S} g(\theta^{(s)}) \tag{1.12}$$

then $\widehat{g}_S$ converges to $E[g(\theta)|y]$ as S goes to infinity.

In practice, this means that, if we can get the computer to take a random sample from the posterior, (1.12) allows us to approximate $E[g(\theta)|y]$ by simply averaging the function of interest evaluated at the random sample. To introduce some jargon, this sampling from the posterior is referred to as *posterior simulation*, and $\theta^{(s)}$ is referred to as a *draw* or *replication*. Theorem 1.1 describes the simplest posterior simulator, and use of this theorem to approximate $E[g(\theta)|y]$ is referred to as *Monte Carlo integration*.

Monte Carlo integration can be used to approximate $E[g(\theta)|y]$, but only if S were infinite would the approximation error go to zero. The econometrician can, of course, choose any value for S (although larger values of S will increase the computational burden). There are many ways of gauging the approximation error associated with a particular value of S. Some of these will be discussed in subsequent chapters. However, many are based on extensions of the central limit theorem given in Appendix B, Definition B.33 and Theorem B.20. For the case of Monte Carlo integration, this central limit theorem implies:

Theorem 1.2: A numerical standard error
Using the setup and definitions of Theorem 1.1,

$$\sqrt{S}\{\widehat{g}_S - E[g(\theta)|y]\} \to N(0, \sigma_g^2) \tag{1.13}$$

as S goes to infinity, where $\sigma_g^2 = var[g(\theta)|y]$.

Theorem 1.2 can be used to obtain an estimate of the approximation error in a Monte Carlo integration exercise by using properties of the Normal distribution. For instance, using the fact that the standard Normal has 95% of its probability located within 1.96 standard deviations from its mean yields the approximate result that:

$$\Pr\left[-1.96\frac{\sigma_g}{\sqrt{S}} \le \widehat{g}_S - E[g(\theta)|y] \le 1.96\frac{\sigma_g}{\sqrt{S}}\right] = 0.95$$

By controlling S, the econometrician can ensure that $\widehat{g}_S - E[g(\theta)|y]$ is sufficiently small with a high degree of probability. In practice, σ_g is unknown, but the Monte Carlo integration procedure allows us to approximate it. The term $\frac{\sigma_g}{\sqrt{S}}$ is known as the *numerical standard error*, and the econometrician can simply report it as a measure of approximation error. Theorem 1.2 also implies, for example, that if $S = 10\,000$ then the numerical standard error is 1%, as big as the posterior standard deviation. In many empirical contexts, this may be a nice way of expressing the approximation error implicit in Monte Carlo integration.

Unfortunately, it is not always possible to do Monte Carlo integration. Algorithms exist for taking random draws from many common densities (e.g. the

Normal, the Chi-squared).[5] However, for many models, the posteriors do not have one of these common forms. In such cases, development of posterior simulators is a more challenging task. In subsequent chapters, we describe many types of posterior simulators. However, we introduce Monte Carlo integration here so as to present the basic ideas behind posterior simulation in a simple case.

1.3 BAYESIAN COMPUTER SOFTWARE

There are several computer software packages that are useful for doing Bayesian analysis in certain classes of model. However, Bayesian econometrics still tends to require a bit more computing effort than frequentist econometrics. For the latter, there are many canned packages that allow the user to simply click on an icon in order to carry out a particular econometric procedure. Many would argue that this apparent advantage is actually a disadvantage, in that it encourages the econometrician to simply use whatever set of techniques is available in the computer package. This can lead to the researcher simply presenting whatever estimates, test statistics, and diagnostics that are produced, regardless of whether they are appropriate for the application at hand. Bayesian inference forces the researcher to think in terms of the models (i.e. likelihoods and priors), which are appropriate for the empirical question under consideration. The myriad of possible priors and likelihoods make it difficult to construct a Bayesian computer package that can be used widely. For this reason, many Bayesian econometricians create their own programs in matrix programming languages such as MATLAB, Gauss, or Ox. This is not that difficult to do. It is also well worth the effort, since writing a program is a very good way of forcing yourself to fully understand an econometric procedure. In this book, the empirical illustrations are carried out using MATLAB, which is probably the most commonly-used computer language for Bayesian econometrics and statistics. The website associated with this book contains copies of the programs used in the empirical illustrations, and the reader is encouraged to experiment with these programs as a way of learning Bayesian programming. Furthermore, some of the questions at the end of each chapter require the use of the computer, and provide another route for the reader to develop some basic programming skills.

For readers who do not wish to develop programming skills, there are some Bayesian computer packages that allow for simple analysis of standard classes of models. BUGS, an acronym for Bayesian Inference Using Gibbs Sampling (see Best *et al.*, 1995), handles a fairly wide class of models using a common posterior simulation technique called Gibbs sampling. More directly relevant for econometricians is Bayesian Analysis, Computation and Communication (BACC), which handles a wide range of common models (see McCausland and Stevens, 2001).

[5]Draws made by the computer follow a particular algorithm and, hence, are not formally random. It is more technically correct to call draws generated by the computer *pseudo-random*. Devroye (1986) provides a detailed discussion of pseudo-random number generation.

The easiest way to use BACC is as a dynamically linked library to another popular language such as MATLAB. In other words, BACC can be treated as a set of MATLAB commands. For instance, instead of programming up a posterior simulator for analysis of the regression model discussed in Chapter 4, BACC allows for Bayesian inference to be done using one simple MATLAB command. Jim LeSage's Econometrics Toolbox (see LeSage, 1999) also contains many MATLAB functions that can be used for aspects of Bayesian inference. The empirical illustrations in this book which involve posterior simulation use his random number generators. At the time of writing, BUGS, BACC, and the Econometrics Toolbox were available on the web for free for educational purposes. Many other Bayesian software packages exist, although most are more oriented towards the statistician than the econometrician. Appendix C of Carlin and Louis (2000) provides much more information about relevant software.

1.4 SUMMARY

In this chapter, we have covered all the basic issues in Bayesian econometrics at a high level of abstraction. We have stressed that the ability to put all the general theory in one chapter, involving only basic concepts in probability, is an enormous advantage of the Bayesian approach. The basic building blocks of the Bayesian approach are the likelihood function and the prior, the product of these defines the posterior (see (1.3)), which forms the basis for inference about the unknown parameters in a model. Different models can be compared using *posterior model probabilities* (see (1.5)), which require the calculation of *marginal likelihoods* (1.6). Prediction is based on the *predictive density* (1.9). In most cases, it is not possible to work with all these building blocks analytically. Hence, Bayesian computation is an important topic. *Posterior simulation* is the predominant method of Bayesian computation.

Future chapters go through particular models, and show precisely how these abstract concepts become concrete in practical contexts. The logic of Bayesian econometrics set out in this chapter provides a template for the organization of following chapters. Chapters will usually begin with a likelihood function and a prior. Then a posterior is derived along with computational methods for posterior inference and model comparison. The reader is encouraged to think in terms of this likelihood/prior/posterior/computation organizational structure both when reading this book and when beginning a new empirical project.

1.5 EXERCISES

1.5.1 Theoretical Exercises

Remember that Appendix B describes basic concepts in probability, including definitions of common probability distributions.

1. *Decision Theory*. In this book, we usually use the posterior mean as a point estimate. However, in a formal decision theoretic context, the choice of a point estimate of θ is made by defining a loss function and choosing the point estimate which minimizes expected loss. Thus, if $C(\widetilde{\theta}, \theta)$ is the loss (or cost) associated with choosing $\widetilde{\theta}$ as a point estimate of θ, then we would choose $\widetilde{\theta}$ which minimizes $E[C(\widetilde{\theta}, \theta)|y]$ (where the expectation is taken with respect to the posterior of θ). For the case where θ is a scalar, show the following:
   (a) *Squared error loss function*. If $C(\widetilde{\theta}, \theta) = (\widetilde{\theta} - \theta)^2$ then $\widetilde{\theta} = E(\theta|y)$.
   (b) *Asymmetric linear loss function*. If

$$C(\widetilde{\theta}, \theta) = \begin{cases} c_1|\widetilde{\theta} - \theta| \text{ if } \widetilde{\theta} \le \theta \\ c_2|\widetilde{\theta} - \theta| \text{ if } \widetilde{\theta} > \theta \end{cases}$$

   where $c_1 > 0$ and $c_2 > 0$ are constants, then $\widetilde{\theta}$ is the $\frac{c_1}{c_1+c_2}$th quantile of $p(\theta|y)$.
   (c) *All-or-nothing loss function*. If

$$C(\widetilde{\theta}, \theta) = \begin{cases} c \text{ if } \widetilde{\theta} \ne \theta \\ 0 \text{ if } \widetilde{\theta} = \theta \end{cases}$$

   where $c > 0$ is a constant, then $\widetilde{\theta}$ is the mode of $p(\theta|y)$.
2. Let $y = (y_1, \ldots, y_N)'$ be a random sample where $p(y_i|\theta) = f_G(y_i|\theta, 2)$. Assume a Gamma prior for θ: $p(\theta) = f_G(\theta|\underline{\theta}, \underline{\nu})$:
   (a) Derive $p(\theta|y)$ and $E(\theta|y)$.
   (b) What happens to $E(\theta|y)$ as $\underline{\nu} \to 0$? In what sense is such a prior 'noninformative'?
3. Let $y = (y_1, \ldots, y_N)'$ be a random sample, where

$$p(y_i|\theta) = \begin{cases} \theta^{y_i}(1-\theta)^{y_i} \text{ if } 0 \le y_i \le 1 \\ 0 \quad \text{otherwise} \end{cases}$$

   (a) Derive the posterior for θ assuming a prior $\theta \sim U(0, 1)$. Derive $E(\theta|y)$.
   (b) Repeat part (a) assuming a prior of the form:

$$p(\theta) = \begin{cases} \dfrac{\Gamma(\underline{\alpha} + \underline{\beta})}{\Gamma(\underline{\alpha})\Gamma(\underline{\beta})}\theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\beta}-1} \text{ if } 0 < \theta < 1 \\ 0 \quad \text{otherwise} \end{cases}$$

   where $\underline{\alpha}$ and $\underline{\beta}$ are prior hyperparameters.

1.5.2 Computer-Based Exercises

4. Suppose that the posterior for a parameter, θ, is $N(0, 1)$:
   (a) Create a computer program which carries out Monte Carlo integration (see (1.12)) to estimate the posterior mean and variance of θ. (Note: Virtually any relevant computer package such as MATLAB or Gauss will have a function which takes random draws from the standard Normal.)
   (b) How may replications are necessary to ensure that the Monte Carlo estimates of the posterior mean and variance are equal to their true values of 0 and 1 to three decimal places?
   (c) To your computer program, add code which calculates numerical standard errors (see (1.13)). Experiment with calculating posterior means, standard deviations, and numerical standard errors for various values of S. Do the numerical standard errors give a reliable indication of the accuracy of approximation in the Monte Carlo integration estimates?