Part I

Fundamentals of Bayesian Inference

Chapter 1

Introduction

THIS CHAPTER INTRODUCES several important concepts, provides a guide to the rest of the book, and offers some historical perspective and suggestions for further reading.

1.1 Econometrics

Econometrics is largely concerned with quantifying the relationship between one or more variables y, called the response variables or the dependent variables, and one or more variables x, called regressors, independent variables, or covariates. The response variable or variables may be continuous or discrete; the latter case includes binary, multinomial, and count data. For example, y might represent the quantities demanded of a set of goods, and x could include income and the prices of the goods; or y might represent investment in capital equipment, and x could include measures of expected sales, cash flows, and borrowing costs; or y might represent a decision to travel by public transportation rather than private, and x could include income, fares, and travel time under various alternatives.

In addition to the covariates, it is assumed that unobservable random variables affect y, so that y itself is a random variable. It is characterized either by a probability density function (p.d.f.) for continuous y or a probability mass function (p.m.f.) for discrete y. The p.d.f. or p.m.f. depends on the values of unknown parameters, denoted by θ. The notation $y \sim f(y|\theta, x)$ means that y has the p.d.f. or p.m.f. $f(y|\theta, x)$, where the function depends on the parameters and covariates. It is customary to suppress dependence on the covariates when writing the p.d.f. of y, so write $y \sim f(y|\theta)$, unless it is necessary to mention the covariates explicitly.

The data may take the form of observations on a number of subjects at the same point in time – cross section data – or observations over a number of time periods – time series data. They may be a combination of cross section and time series observations: data over many subjects over a relatively short period of time – panel data – or data over a fairly small number of subjects over long periods of time – multivariate data. In some models, the researcher regards the covariates as fixed numbers, but in others they are regarded as random variables. If the latter, their distribution may be independent of the distribution of y, or there may be dependence. All of these possibilities are discussed in Part III.

An important feature of data analyzed by econometricians is that the data are almost always observational, in contrast to data arising from controlled experiments, where subjects are randomly assigned to treatments. Observational data are often generated for purposes other than research, for example, as by-products of data collected for governmental and administrative reasons. Observational data may also be collected from surveys, some of which may be specially designed for research purposes. No matter how data are collected, however, the analysis of observational data requires special care, especially in the analysis of causal effects – the attempt to determine the effect of a covariate on a response variable when the covariate is a variable whose value can be set by an investigator, such as the effect of participating in a training program on income and employment or the effect of exercise on health. When such data are collected from observing what people choose to do, rather than from a controlled experiment in which they are told what to do, there is a possibility that people who choose to take the training or to exercise are different in some systematic way from people who do not. If so, attempting to generalize the effect of training or exercise on people who do not freely choose those options may give misleading answers. The models discussed in Part III are designed to deal with observational data.

Depending on the nature of the data, models are constructed that relate response variables to covariates. A large number of models that can be applied to particular types of data have been developed, but, because new types of data sets may require new models, it is important to learn how to deal with models that have not been previously analyzed. Studying how Bayesian methodology has been applied to a variety of existing models is useful for developing techniques that can be applied to new models.

1.2 Overview of the Book

Part I of the book sets out the basic ideas of the Bayesian approach to statistical inference. It begins with an explanation of subjective probability to justify the application of probability theory to general situations of uncertainty. With this background, Bayes theorem is invoked to define the posterior distribution, the central concept in Bayesian statistical inference. I show how the posterior distribution is used to solve the standard problems of statistical inference: point and interval estimation, prediction, and model comparison. This material is illustrated with the

Bernoulli model of coin tossing. Because of its simplicity, all relevant calculations can be done analytically.

The remainder of Part I is devoted to general properties of posterior distributions and to the specification of prior distributions. These properties are illustrated with the normal distribution and linear regression models. For more complicated models, discussion turns to simulation as a way of studying posterior distributions because it is impossible to make the necessary computations analytically.

Part II is devoted to the explanation of simulation techniques. I start with the classical methods of simulation that yield independent samples, but are inadequate to deal with many common statistical models. The remainder of Part II describes Markov chain Monte Carlo (MCMC) simulation, a flexible simulation method that can deal with a wide variety of models.

Part III applies MCMC techniques to models commonly encountered in econometrics and statistics. Here, emphasis is on the design of algorithms to analyze these models as a way of preparing the student to devise algorithms for the new models that will arise in the course of his or her research.

Appendix A contains definitions, properties, and notation for the standard probability distributions that are used throughout the book, a few important probability theorems, and several useful results from matrix algebra. Appendix B describes computer programs for implementing the methods discussed in the book.

1.3 Historical Note and Further Reading

Bayesian statistics is named for the Reverend Thomas Bayes (1702–61), but important contributions to the ideas, under the rubric of "inverse probability," were also made by Pierre-Simon Laplace (1749–1827). Stigler (1986) is an excellent introduction to the history of statistics up to the beginning of the twentieth century. Another important approach to inference, the frequentist approach, was largely developed in the second half of the nineteenth century. The leading advocates of the approach in the twentieth century were R. A. Fisher, J. Neyman, and E. Pearson, although Fisher's viewpoint differs in important respects from the others. Howie (2002) provides a concise summary of the development of probability and statistics up to the 1920s and then focuses on the debate between H. Jeffreys, who took the Bayesian position, and R. A. Fisher, who argued against it. McGrayne (2011) is a nontechnical account of the Bayesian approach with an emphasis on the people who figured prominently in its development and on its varied applications.

The application of the Bayesian viewpoint to econometric models was pioneered by A. Zellner starting in the early 1960s. His early work is summarized in his highly influential book, *Introduction to Bayesian Inference in Economics*, Zellner (1971), and he made many more contributions to the literature. Several recent textbooks cover Bayesian econometrics: Poirier (1995), Koop (2003), Lancaster (2004), and Geweke (2005). Poirier's book, unlike the present book and the others mentioned earlier, compares and contrasts Bayesian methods with other approaches to statistics and econometrics in great detail. Koop, Poirier, and Tobias (2007) covers Bayesian econometrics through a series of worked exercises.

Gill (2007) and Jackman (2009) take a Bayesian approach directed toward social and behavioral science researchers, and Rossi et al. (2005) focuses on the Bayesian approach to the analysis of marketing models and data. Two textbooks that emphasize the frequentist viewpoint – Mittelhammer, Judge, and Miller (2000) and Greene (2003) – also discuss Bayesian inference.

Several statistics books take a Bayesian viewpoint. Berry (1996) is an excellent introduction to Bayesian ideas. His discussion of differences between observational and experimental data is highly recommended. Another fine introductory book is Bolstad (2004). Albert (2009) and Robert and Casella (2010) expound computational aspects of Bayesian inference using the R system, which has become the preferred environment for this purpose. Excellent intermediate level books with many examples are Carlin and Louis (2009), Lee (2004), and Gelman et al. (2004). At a more advanced level, the following are especially recommended: O'Hagan and Forster (2004), Robert (2007), Marin and Robert (2007), Bernardo and Smith (1994), and Jaynes (2003). Robert and Casella (2004) is an excellent introduction to simulation techniques.

Although directed at a general statistical audience, three books by Congdon (2001, 2003, 2005) cover many common econometric models and utilize MCMC methods extensively. Schervish (1995) covers both Bayesian and frequentist ideas at an advanced level.

Cambridge University Press
978-1-107-01531-9 - Introduction to Bayesian Econometrics: Second Edition
Edward Greenberg
Excerpt
More information

Chapter 2

Basic Concepts of Probability and Inference

2.1 Probability

BECAUSE STATISTICAL INFERENCE is based on probability theory, the major difference between Bayesian and frequentist approaches to inference can be traced to the different views that each take about the interpretation and scope of probability theory. Therefore, this chapter begins by stating the basic axioms of probability and explaining the two views.

A probability is a number assigned to statements or events. I use the terms "statements" and "events" interchangeably. Examples of such statements are

- $A_1 =$ A coin tossed three times will come up heads either two or three times.
- $A_2 =$ A six-sided die rolled once shows an even number of spots.
- $A_3 =$ There will be measurable precipitation on January 1, 2013, at your local airport.

Before presenting the probability axioms, let us review some standard notation:

The *union* of A and B is the event that A or B (or both) occur; it is denoted by $A \cup B$.
The *intersection* of A and B is the event that both A and B occur; it is denoted by AB.
The *complement* of A is the event that A does not occur; it is denoted by A^c.

The probability of event A is denoted by $P(A)$. Probabilities are assumed to satisfy the following axioms:

Probability Axioms

1. $0 \leq P(A) \leq 1$.
2. $P(A) = 1$ if A represents a logical truth, that is, a statement that must be true; for example, A coin comes up either heads or tails.

3. If A and B describe disjoint events (events that cannot both occur), then $P(A \cup B) = P(A) + P(B)$.
4. Let $P(A|B)$ denote the probability of A, given (or conditioned on the assumption) that B is true. Then
$$P(A|B) = \frac{P(AB)}{P(B)}.$$

All of the theorems of probability theory can be deduced from these axioms, and probabilities that are assigned to statements will be consistent if these rules are observed. Consistent means that it is not possible to assign two or more different values to the probability of a particular event, if probabilities are assigned by following these rules. As an example, if $P(A)$ has been assigned a value, then Axioms 1 and 2 imply that $P(A^c) = 1 - P(A)$, and $P(A^c)$ can take no other value. Assigning some probabilities may put bounds on others. For example, if A and B are disjoint and $P(A)$ is given, then by Axioms 1 and 3, $P(B) \leq 1 - P(A)$.

2.1.1 Frequentist Probabilities

A major controversy in probability theory is over the types of statements to which probabilities can be assigned. One school of thought is that of the "frequentists." Frequentists restrict the assignment of probabilities to statements that describe the outcome of an experiment that can be repeated. Consider A_1: imagine repeating the experiment of tossing a coin three times and recording the number of times that two or three heads were reported. If you define

$$P(A_1) = \lim_{n \to \infty} \frac{\text{number of times two or three heads occurs}}{n},$$

you find that the definition is consistent with the axioms of probability.

Axiom 1 is satisfied because a ratio of a subset of outcomes to all possible outcomes is between zero and one. Axiom 2 is satisfied if the probability of a logically true statement such as $A_4 =$ "either 0, 1, 2, or 3 heads appear" is computed by following the rule because the numerator is then equal to n. Axiom 3 implies that you can compute $P(A \cup B)$ as $P(A) + P(B)$ because, for disjoint events, the number of times A or B occurs is equal to the number of times A occurs plus the number of times B occurs. Axiom 4 is satisfied because to compute $P(A|B)$ you can confine your attention to the outcomes of the experiment for which B is true;

suppose there are n_B of these. Then,

$$
\begin{aligned}
P(A|B) &= \lim_{n_B \to \infty} \frac{\text{number of times } A \text{ and } B \text{ are true}}{n_B} \\
&= \lim_{n, n_B \to \infty} \frac{\text{number of times } A \text{ and } B \text{ are true}}{n} \div \frac{n_B}{n} \\
&= \frac{p(AB)}{p(B)}.
\end{aligned}
$$

This method of assigning probabilities, even to experiments that can be repeated in principle, suffers from the problem that its definition requires repeating the experiment an infinite number of times, which is impossible. But to those who believe in a subjective interpretation of probability, an even greater problem is its inability to assign probabilities to such statements as A_3, which cannot be considered an outcome of a repeated experiment. Next, the subjective view is considered.

2.1.2 Subjective Probabilities

Those who take the subjective view of probability believe that probability theory is applicable to any situation in which there is uncertainty. Outcomes of repeated experiments fall in that category, but so do statements about tomorrow's weather, which are not outcomes of repeated experiments. Calling the probabilities "subjective" does not imply that they may be assigned without regard to the axioms of probability. Such assignments would lead to inconsistencies. De Finetti (1990) provides a principle for assigning probabilities that does not rely on outcomes of repeated experiments but is consistent with the probability axioms.

De Finetti develops his approach in the context of setting odds on a bet that are fair in the sense that, in your opinion, neither you nor your opponent has an advantage. In particular, when the odds are fair, you will not find yourself in the position that you will lose money no matter what outcome obtained. De Finetti calls your behavior *coherent* when you set odds in this way. I now show that coherent behavior implies that probabilities satisfy the axioms.

First, let us review the standard betting setup: in a standard bet, on the event A, you buy or sell betting tickets at a price of 1 per ticket, and the money you receive or pay out depends on the betting odds, k. (I omit the currency unit in this discussion.) In this setup, the price of the ticket is fixed and the payout depends on the odds. Denote the number of tickets by S and make the convention that $S > 0$ means that you are betting that A occurs (i.e., you have bought S tickets on A from your opponent) and $S < 0$ means that you are betting against A (i.e., you have sold S tickets on A to your opponent). If you bet on A, and A occurs, you

Table 2.1. *Coherency: Restrictions on p*

Event	Your gain
A	$S - pS = (1 - p)S$
A^c	$-pS$

receive the 1 that you bet plus k for each ticket you bought, or $S(1 + k)$, where k is the odds *against* A,

$$k = \frac{1 - P(A)}{P(A)};$$

(Berry, 1996). If A occurs and you bet against it, you would "receive" $S(1 + k)$, a negative number because $S < 0$ if you bet against A.

In the de Finetti betting setup, the price of a ticket, denoted by p, is chosen by you, the payout is fixed at 1, and your opponent chooses S. Although you set p, the fact that your opponent determines whether you bet for or against A forces you to set a fair value. This shows the connection between p and $P(A)$. If the price of a ticket is p rather than one, as in the standard betting situation, a winning ticket on A would pay $p + pk = p(1 + k)$. But in the de Finetti setup, the payout is one; that is, $p(1 + k) = 1$, or $k = (1 - p)/p$, which implies $p = P(A)$. Accordingly, in the following discussion you can interpret p as your subjective belief about the value of $P(A)$.

Consider a simple bet on or against A, where you have set the price of a ticket at p and you are holding S tickets for which you have paid pS; your opponent has chosen S. (Remember that $S > 0$ means that you are betting on A, and $S < 0$ means you are betting against A.) If A occurs, you pay pS and collect S. If A does not occur, you collect pS. Verify that these results are valid for both positive and negative values of S. Your gains are summarized in Table 2.1, where the rows denote disjoint events and cover all possible outcomes.

This table shows that the principle of coherency restricts the value of p you set. If $p < 0$, your opponent, by choosing $S < 0$, will inflict a loss (a negative gain) on you whether or not A occurs. By coherency, therefore, $p \geq 0$. Similarly, if you set $p > 1$, your opponent can set $S > 0$, and you are again sure to lose. Axiom 1 is therefore implied by the principle of coherency.

If you are certain that A will occur, coherency dictates that you set $p = 1$: you will have zero loss if A is true, and the second row of the loss table is not relevant, because A^c is impossible in your opinion. This verifies Axiom 2.

To examine the subjective assignment of $P(A_1 \cup A_2)$ if A_1 and A_2 are disjoint, consider the table of gains when you set prices p_1, p_2, p_3, and your opponent