ECONOMETRIC EXERCISES 7

# Bayesian Econometric Methods


Gary Koop

Dale J. Poirier

Justin L. Tobias



CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521855716

This page intentionally left blank

Econometric Exercises, Volume 7

# Bayesian Econometric Methods

This book is a volume in the Econometric Exercises series. It teaches principles of Bayesian econometrics by posing a series of theoretical and applied questions, and providing detailed solutions to those questions. This text is primarily suitable for graduate study in econometrics, though it can be used for advanced undergraduate courses, and should generate interest from students in related fields, including finance, marketing, agricultural economics, business economics, and other disciplines that employ statistical methods. The book provides a detailed treatment of a wide array of models commonly employed by economists and statisticians, including linear regression-based models, hierarchical models, latent variable models, mixture models, and time series models. Basics of random variable generation and simulation via Markov Chain Monte Carlo (MCMC) methods are also provided. Finally, posterior simulators for each type of model are rigorously derived, and Matlab computer programs for fitting these models (using both actual and generated data sets) are provided on the Web site accompanying the text.

Gary Koop is Professor of Economics at the University of Strathclyde. He has published numerous articles in Bayesian econometrics and statistics in journals such as the *Journal of Econometrics*, *Journal of the American Statistical Association*, and the *Journal of Business and Economic Statistics*. He is an associate editor for several journals, including the *Journal of Econometrics* and *Journal of Applied Econometrics*. He is the author of the books *Bayesian Econometrics*, *Analysis of Economic Data*, and *Analysis of Financial Data*.

Dale J. Poirier is Professor of Economics at the University of California, Irvine. He is a Fellow of the Econometric Society, the American Statistical Association, and the *Journal of Econometrics*. He has been on the Editorial Boards of the *Journal of Econometrics* and *Econometric Theory* and was the founding editor of *Econometric Reviews*. His professional activities have been numerous, and he has held elected positions in the American Statistical Association and the International Society for Bayesian Analysis. His previous books include *Intermediate Statistics and Econometrics: A Comparative Approach* and *The Econometrics of Structural Change*.

Justin L. Tobias is Associate Professor of Economics, Iowa State University, and has also served as an Assistant and Associate Professor of Economics at the University of California, Irvine. Professor Tobias has authored numerous articles in leading journals, including the *International Economic Review*, *Journal of Applied Econometrics*, *Journal of Business and Economic Statistics*, *Journal of Econometrics*, and the *Review of Economics and Statistics*.

## Econometric Exercises

**Editors:**

Karim M. Abadir, *Tanaka Business School, Imperial College London, UK*
Jan R. Magnus, *CentER and Department of Econometrics and Operations Research, Tilburg University, The Netherlands*
Peter C. B. Phillips, *Cowles Foundation for Research in Economics, Yale University, USA*

**Titles in the Series (* = planned):**

1 Matrix Algebra (K. M. Abadir and J. R. Magnus)
2 Statistics (K. M. Abadir, R. D. H. Heijmans, and J. R. Magnus)
3 Econometric Theory, I (P. Paruolo)
4 Empirical Applications, I (A. van Soest and M. Verbeek)
5 Econometric Theory, II
6 Empirical Applications, II
7 Bayesian Econometric Methods (G. Koop, D. J. Poirier, and J. L. Tobias)
* Time Series Econometrics, I
* Time Series Econometrics, II
* Microeconometrics
* Panel Data
* Nonlinear Models
* Nonparametrics and Semiparametrics
* Simulation-Based Econometrics
* Computational Methods
* Financial Econometrics
* Robustness
* Econometric Methodology

# Bayesian Econometric Methods

Gary Koop
*University of Strathclyde, Scotland*

Dale J. Poirier
*University of California, Irvine, USA*

Justin L. Tobias
*Iowa State University, USA*

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521855716



First published in print format 2007

ISBN-13 978-0-511-29242-2 eBook (Adobe Reader)
ISBN-10 0-511-29242-2 eBook (Adobe Reader)

ISBN-13 978-0-521-85571-6 hardback
ISBN-10 0-521-85571-3 hardback

ISBN-13 978-0-521-67173-6 paperback
ISBN-10 0-521-67173-6 paperback

*To* Lise
*To* the Reverend but not the Queen
*To* Melissa, Madeline, and Drew

# Contents

# List of exercises

## Chapter 4: Frequentist properties of Bayesian estimators



## Chapter 5: Interval estimation



## Chapter 6: Hypothesis testing



## Chapter 7: Prediction



## Chapter 8: Choice of prior



## Chapter 9: Asymptotic Bayes

# Preface to the series

The past two decades have seen econometrics grow into a vast discipline. Many different branches of the subject now happily coexist with one another. These branches interweave econometric theory and empirical applications and bring econometric method to bear on a myriad of economic issues. Against this background, a guided treatment of the modern subject of econometrics in volumes of worked econometric exercises seemed a natural and rather challenging idea.

The present series, *Econometric Exercises*, was conceived in 1995 with this challenge in mind. Now, almost a decade later it has become an exciting reality with the publication of the first installment of a series of volumes of worked econometric exercises. How can these volumes work as a tool of learning that adds value to the many existing textbooks of econometrics? What readers do we have in mind as benefiting from this series? What format best suits the objective of helping these readers learn, practice, and teach econometrics? These questions we now address, starting with our overall goals for the series.

*Econometric Exercises* is published as an organized set of volumes. Each volume in the series provides a coherent sequence of exercises in a specific field or subfield of econometrics. Solved exercises are assembled together in a structured and logical pedagogical framework that seeks to develop the subject matter of the field from its foundations through to its empirical applications and advanced reaches. As the Schaum series has done so successfully for mathematics, the overall goal of *Econometric Exercises* is to develop the subject matter of econometrics through solved exercises, providing a coverage of the subject that begins at an introductory level and moves through to more advanced undergraduate- and graduate-level material.

Problem solving and worked exercises play a major role in every scientific subject. They are particularly important in a subject like econometrics in which there is a rapidly growing literature of statistical and mathematical technique and an ever-expanding core to the discipline. As students, instructors, and researchers, we all benefit by seeing carefully worked-out solutions to problems that develop the subject and illustrate its methods and workings. Regular exercises and problem sets consolidate learning and reveal applications of textbook material. Clearly laid out solutions, paradigm answers, and alternate routes to solution all develop problem-solving skills. Exercises train students in clear analytical thinking and help them in preparing for tests and exams. Teachers, as well as students, find solved exercises useful in their classroom preparation and in designing problem sets, tests, and examinations. Worked problems and illustrative empirical applications appeal to researchers and professional economists wanting to learn about specific econometric techniques. Our intention for the *Econometric Exercises* series is to appeal to this wide range of potential users.

Each volume of the series follows the same general template. Chapters begin with a short outline that emphasizes the main ideas and overviews the most relevant theorems and results. The introductions are followed by a sequential development of the material by solved examples and applications, and by computer exercises when appropriate. All problems are solved and they are graduated in difficulty with solution techniques evolving in a logical, sequential fashion. Problems are asterisked when they require more creative solutions or reach higher levels of technical difficulty. Each volume is self-contained. There is some commonality in material across volumes to reinforce learning and to make each volume accessible to students and others who are working largely, or even completely, on their own.

Content is structured so that solutions follow immediately after the exercise is posed. This makes the text more readable and avoids repetition of the statement of the exercise when it is being solved. More importantly, posing the right question at the right moment in the development of a subject helps to anticipate and address future learning issues that students face. Furthermore, the methods developed in a solution and the precision and insights of the answers are often more important than the questions being posed. In effect, the inner workings of a good solution frequently provide benefit beyond what is relevant to the specific exercise.

Exercise titles are listed at the start of each volume, following the table of contents, so that readers may see the overall structure of the book and its more detailed contents. This organization reveals the exercise progression, how the exercises relate to one another, and where the material is heading. It should also tantalize readers with the exciting prospect of advanced material and intriguing applications.

The series is intended for a readership that includes undergraduate students of econometrics with an introductory knowledge of statistics, first- and second-year graduate students of econometrics, as well as students and instructors from neighboring disciplines (such as statistics, psychology, or political science) with interests in econometric methods. The volumes generally increase in difficulty as the topics become more specialized.

The early volumes in the series (particularly those covering matrix algebra, statistics, econometric models, and empirical applications) provide a foundation to the study of econometrics. These volumes will be especially useful to students who are following the first-year econometrics course sequence in North American graduate schools and need to prepare for graduate comprehensive examinations in econometrics and to write an applied econometrics paper. The early volumes will equally be of value to advanced undergraduates studying econometrics in Europe, to advanced undergraduates and honors students in the Australasian system, and to masters and doctoral students in general. Subsequent volumes will be of interest to professional economists, applied workers, and econometricians who are working with techniques in those areas, as well as students who are taking an advanced course sequence in econometrics and statisticians with interests in those topics.

The *Econometric Exercises* series is intended to offer an independent learning-by-doing program in econometrics and it provides a useful reference source for anyone wanting to learn more about econometric methods and applications. The individual volumes can be used in classroom teaching and examining in a variety of ways. For instance, instructors can work through some of the problems in class to demonstrate methods as they are introduced; they can illustrate theoretical material with some of the solved examples; and they can show real data applications of the methods by drawing on some of the empirical examples. For examining purposes, instructors may draw freely from the solved exercises in test preparation. The systematic development of the subject in individual volumes will make the material easily accessible both for students in revision and for instructors in test preparation.

In using the volumes, students and instructors may work through the material sequentially as part of a complete learning program, or they may dip directly into material in which they are experiencing difficulty to learn from solved exercises and illustrations. To promote intensive study, an instructor might announce to a class in advance of a test that some questions in the test will be selected from a certain chapter of one of the volumes. This approach encourages students to work through most of the exercises in a particular chapter by way of test preparation, thereby reinforcing classroom instruction.

Further details and updated information about individual volumes can be obtained from the *Econometric Exercises* Web site:

http://us.cambridge.org/economics/ee/econometricexercises.htm

The Web site also contains the basic notation for the series, which can be downloaded along with the LaTeX style files.

As series editors, we welcome comments, criticisms, suggestions, and, of course, corrections from all our readers on each of the volumes in the series as well as on the series itself. We bid you as much happy reading and problem solving as we have had in writing and preparing this series.

*York, Tilburg, New Haven*
*June 2005*

Karim M. Abadir
Jan R. Magnus
Peter C. B. Phillips

# Preface

Bayesian econometrics has enjoyed an increasing popularity in many fields. This popularity has been evidenced through the recent publication of several textbooks at the advanced undergraduate and graduate levels, including those by Poirier (1995), Bauwens, Lubrano, and Richard (1999), Koop (2003), Lancaster (2004), and Geweke (2005). The purpose of the present volume is to provide a wide range of exercises and solutions suitable for students interested in Bayesian econometrics at the level of these textbooks.

The Bayesian researcher should know the basic ideas underlying Bayesian methodology (i.e., Bayesian theory) and the computational tools used in modern Bayesian econometrics (i.e., Bayesian computation). The Bayesian should also be able to put the theory and computational tools together in the context of substantive empirical problems. We have written this book with these three activities – theory, computation, and empirical modeling – in mind. We have tried to construct a wide range of exercises on all of these aspects. Loosely speaking, Chapters 1 through 9 focus on Bayesian theory, whereas Chapter 11 focuses primarily on recent developments in Bayesian computation. The remaining chapters focus on particular models (usually regression based). Inevitably, these chapters combine theory and computation in the context of particular models. Although we have tried to be reasonably complete in terms of covering the basic ideas of Bayesian theory and the computational tools most commonly used by the Bayesian, there is no way we can cover all the classes of models used in econometrics. Accordingly, we have selected a few popular classes of models (e.g., regression models with extensions and panel data models) to illustrate how the Bayesian paradigm works in practice. Particularly in Chapters 12 through 18 we have included substantive empirical exercises – some of them based closely on journal articles. We hope that the student who works through these chapters will have a good feeling for how serious Bayesian empirical work is done and will be well placed to write a Ph.D. dissertation or a journal article using Bayesian methods.

For the student with limited time, we highlight that a division in this book occurs between the largely theoretical material of Chapters 1 through 9 and the largely regression-based material in Chapters 10 through 18. A student taking a course on Bayesian statistical theory could focus on Chapters 1 through 9, whereas a student taking a Bayesian econometrics course (or interested solely in empirical work) could focus more on Chapters 10 through 18 (skimming through the more methodologically oriented material in the early chapters).

Although there have been some attempts to create specifically Bayesian software (e.g., BUGS, which is available at http://www.mrc-bsu.cam.ac.uk/bugs, or BACC, which is available at http://www2.cirano.qc.ca/∼bacc), in our estimation, most Bayesians still prefer to create their own programs using software such as Matlab, OX, or GAUSS. We have used Matlab to create answers to the empirical problems in this book. Our Matlab code is provided on the Web site associated with this book:

http://www.econ.iastate.edu/faculty/tobias/Bayesian_exercises.html

A few notational conventions are applied throughout the book, and it is worthwhile to review some of these prior to diving into the exercises. In regression-based problems, which constitute a majority of the exercises in the later chapters, lowercase letters such as $y$ and $x_i$ are reserved to denote scalar or vector quantities whereas capitals such as $X$ or $X_j$ are used to denote matrices. In cases in which the distinction between vectors and scalars is critical, this will be made clear within the exercise. In the regression-based problems, $y$ is assumed to denote the $n \times 1$ vector of stacked responses for the dependent variable, $y_i$ the $i$th element of that vector, $x_i$ a $k$ vector of covariate data, and $X$ the $n \times k$ matrix obtained from stacking the $x_i$ over $i$. Latent variables, which are often utilized in the computational chapters of the book, are typically designated with a "*" superscript, such as $y_i^*$. In Chapters 1 through 9, many exercises are presented that are not directly related to linear regression models or models that can be viewed as linear on suitably defined latent data. In these exercises, the distinction between random variables and realizations of those variables is sometimes important. In such cases, we strive to use capital letters to denote random variables, which are unknown *ex ante*, and lowercase letters to denote their realizations, which are known *ex post*. So, in the context of discussing a posterior distribution (which conditions on the data), we will use $\overline{y}$, but if we are interested in discussing the sampling properties of the sample mean, $\overline{Y}$ would be the appropriate notation. Finally, "$\times$" is used to denote multiplication in multiline derivations, and specific parameterizations of various densities are provided in the Appendix associated with this book.

On the issue of parameterization, the reader who is somewhat familiar with the Bayesian literature may realize that researchers often employ different parameterizations for the same model, with no particular choice being "correct" or "ideal." A leading example is the linear regression model, in which the researcher can choose to parameterize this model in terms of the error variance or the error precision (the reciprocal of the variance). In this book, we try and remain consistent in terms of parameterization within individual chapters, though some departures from this trend do exist, particularly in Chapters 11 and 16. These differences arise from our own individual tastes and styles toward approaching these models, and they are superficial rather than substantive. In our view it is quite valuable to expose the student to the use of different parameterizations, since this is the reality that he or she will face when exploring the Bayesian literature in more detail. In all cases, the parameterization employed is clearly delineated within each exercise.

We would like to thank the editors of the *Econometrics Exercises* series – Karim Abadir, Jan Magnus, and Peter Phillips – for their helpful comments and support during the planning and writing of this book. Hoa Jia, Babatunde Abidoye, and Jingtao Wu deserve special recognition for reviewing numerous exercises and helping to reduce the number of typographical errors. The list of other colleagues and students who have helped us – through designing, solving, and pointing out errors in our problems or solutions – is too long to enumerate here. We would, however, like to thank our students at the University of California, Irvine; Leicester University; University of Toronto; and the Institute for Advanced Studies and CIDE (Italy) for their participation, wise insights, and enthusiasm.

# 1

# The subjective interpretation of probability

Reverend Thomas Bayes (born circa 1702; died 1761) was the oldest son of Reverend Joshua Bayes, who was one of the first ordained nonconformist ministers in England. Relatively little is known about the personal life of Thomas Bayes. Although he was elected a Fellow of the Royal Society in 1742, his only known mathematical works are two articles published posthumously by his friend Richard Price in 1763. The first dealt with the divergence of the Stirling series, and the second, "An Essay Toward Solving a Problem in the Doctrine of Chances," is the basis of the paradigm of statistics named for him. His ideas appear to have been independently developed by James Bernoulli in 1713, also published posthumously, and later popularized independently by Pierre Laplace in 1774. In their comprehensive treatise, Bernardo and Smith (1994, p. 4) offer the following summarization of *Bayesian statistics*:

> Bayesian Statistics offers a rationalist theory of personalistic beliefs in contexts of uncertainty, with the central aim of characterizing how an individual should act in order to avoid certain kinds of undesirable behavioral inconsistencies. The theory establishes that expected utility maximization provides the basis for rational decision making and that Bayes' Theorem provides the key to the ways in which beliefs should fit together in the light of changing evidence. The goal, in effect, is to establish rules and procedures for individuals concerned with disciplined uncertainty accounting. The theory is not descriptive, in the sense of claiming to model actual behavior. Rather, it is prescriptive, in the sense of saying "if you wish to avoid the possibility of these undesirable consequences you must act in the following way."

*Bayesian econometrics* consists of the tools of Bayesian statistics applicable to the models and phenomena of interest to economists. There have been numerous axiomatic formulations leading to the central unifying Bayesian prescription of maximizing subjective utility as the guiding principle of Bayesian statistical analysis. Bernardo and Smith (1994, Chapter 2) is a valuable segue into this vast literature. Deep issues are involved regarding meaningful separation of probability and utility assessments, and we do not address these here.

Non-Bayesians, who we hereafter refer to as *frequentists*, argue that situations not admitting repetition under essentially identical conditions are not within the realm of statistical enquiry, and hence "probability" should not be used in such situations. Frequentists define the probability of an event as its long-run relative frequency. This frequentist interpretation cannot be applied to (i) unique, once-and-for-all type of phenomenon, (ii) hypotheses, or (iii) uncertain past events. Furthermore, this definition is nonoperational since only a finite number of trials can ever be conducted. In contrast, the desire to expand the set of relevant events over which the concept of probability can be applied, and the willingness to entertain formal introduction of "nonobjective" information into the analysis, led to the subjective interpretation of probability.

**Definition 1.1 (Subjective interpretation of probability)** Let $\kappa$ denote the body of knowledge, experience, or information that an individual has accumulated about the situation of concern, and let $A$ denote an uncertain event (not necessarily repetitive). The *probability of $A$ afforded by $\kappa$* is the "degree of belief" in $A$ held by an individual in the face of $\kappa$.

Since at least the time of Ramsey (1926), such degrees of belief have been operationalized in terms of agreed upon reference lotteries. Suppose you seek your degree of belief, denoted $p = P(A)$, that an event $A$ occurs. Consider the following two options.

1. Receiving a small reward \$$r$ if $A$ occurs, and receiving \$0 if $A$ does not occur.
2. Engaging in a lottery in which you win \$$r$ with probability $p$, and receiving \$0 with probability $1 - p$.

If you are indifferent between these two choices, then your degree of belief in $A$ occurring is $p$. Requiring the reward to be "small" is to avoid the problem of introducing utility into the analysis; that is, implicitly assuming utility is linear in money for small gambles.

Bruno de Finetti considered the interesting situation in which an individual is asked to quote betting odds (ratios of probabilities) on a set of uncertain events and accept any wagers others may decide to make about these events. According to de Finetti's *coherence principle* the individual should never assign "probabilities" so that someone else can select stakes that guarantee a sure loss *(Dutch book)* for the individual whatever the eventual outcome. A sure loss amounts to the "undesirable consequences" contained in the earlier quote of Bernardo and Smith. This simple principle implies the axioms of probability discussed in Abadir, Heijmans, and Magnus (2006, Chapter 1) except that the additivity of probability of intersections for disjoint events is required to hold only for *finite* intersections. Nonetheless, for purposes of convenience, we consider only countably additive probability in this volume.

De Finetti's Dutch book arguments also lead to the standard rule for conditional probability. Consider two events A and B. By using the *factorization rule for conditional probability* [Abadir et al. (2006, p. 5)],

$$P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B),$$

the simplest form of Bayes' theorem follows immediately:

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}.$$

In words, we are interested in the event $B$ to which we assign the *prior* probability $P(B)$ for its occurrence. We observe the occurrence of the event $A$. The probability of $B$ occurring given that $A$ has occurred is the *posterior* probability $P(B|A)$. More generally, we have the following result.

**Theorem 1.1 (Bayes' theorem for events)** Consider a probability space $[S, \tilde{A}, P(\cdot)]$ and a collection $B_n \in \tilde{A}$ $(n = 1, 2, \ldots N)$ of mutually disjoint events such that $P(B_n) > 0$ $(n = 1, 2, \ldots, N)$ and $B_1 \cup B_2 \cup \cdots \cup B_N = S$. Then

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{N} P(A|B_j)P(B_j)} \quad (n = 1, 2, \ldots, N) \tag{1.1}$$

for every $A \in \tilde{A}$ such that $P(A) > 0$.

**Proof**: The proof follows directly upon noting that the denominator in (1.1) is $P(A)$.

An important philosophical topic is whether the conditionalization in Bayes theorem warrants an unquestioned position as the model of learning in the face of knowledge of the event $A$. Conditional probability $P(B|A)$ refers to *ex ante* beliefs on events not yet decided. *Ex post* experience of an event can sometimes have a striking influence on the probability assessor (e.g., experiencing unemployment, stock market crashes, etc.), and the experience can bring with it more information than originally anticipated in the event. Nonetheless, we adopt such conditionalization as a basic principle.

The subjective interpretation reflects an individual's personal assessment of the situation. According to the subjective interpretation, probability is a property of an individual's perception of reality, whereas according to classical and frequency interpretations, probability is a property of reality itself. For the subjectivist there are no "true unknown probabilities" in the world out there to be discovered. Instead, "probability" is in the eye of the beholder.

Bruno de Finetti assigned a fundamental role in Bayesian analysis to the concept of *exchangeability*, defined as follows.

**Definition 1.2** A finite sequence $Y_t$ $(t = 1, 2, \ldots, T)$ of events (or random variables) is *exchangeable* iff the joint probability of the sequence, or any subsequence, is invariant under permutations of the subscripts, that is,

$$P(y_1, y_2, \ldots, y_T) = P(y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(T)}), \tag{1.2}$$

where $\pi(t)(t = 1, 2, \ldots, T)$ is a permutation of the elements in $\{1, 2, \ldots, T\}$. An infinite sequence is exchangeable iff any finite subsequence is exchangeable.

Exchangeability provides an operational meaning to the weakest possible notion of a sequence of "similar" random quantities. It is "operational" because it only requires probability assignments of *observable* quantities, although admittedly this becomes problematic in the case of infinite exchangeability. For example, a sequence of Bernoulli trials is *exchangeable* iff the probability assigned to particular sequences does not depend on the order of "successes" $(S)$ and "failures" $(F)$. If the trials are exchangeable, then the sequences FSS, SFS, and SSF are assigned the same probability.

Exchangeability involves recognizing symmetry in beliefs concerning only observables, and presumably this is something about which a researcher may have intuition. Ironically, subjectivists emphasize observables (data) and objectivists focus on unobservables (parameters). Fortunately, Bruno de Finetti provided a subjectivist solution to this perplexing state of affairs. De Finetti's representation theorem and its generalizations are interesting because they provide conditions under which exchangeability gives rise to an isomorphic world in which we have iid observations conditional on a mathematical construct, namely, a parameter. These theorems provide an interpretation of parameters that differs substantively from the interpretation of an objectivist.

As in the case of iid sequences, the individual elements in an exchangeable sequence are identically distributed, but they are *not* necessarily independent, and this has important predictive implications for learning from experience. The importance of the concept of exchangeability is illustrated in the following theorem.

**Theorem 1.2 (de Finetti's representation theorem)** Let $Y_t$ $(t = 1, 2, \ldots)$ be an infinite sequence of Bernoulli random variables indicating the occurrence (1) or nonoccurrence (0) of some event of interest. For any finite sequence $Y_t$ $(t = 1, 2, \ldots, T)$, define the average number of occurrences

$$\overline{Y}_T = \frac{1}{T}\sum_{t=1}^{T} Y_t. \tag{1.3}$$

Let $h(y_1, y_2, \ldots, y_T) = \Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_T = y_T)$ denote a probability mass function (p.m.f.) reflecting *exchangeable* beliefs for an arbitrarily long finite sequence $Y_t$ $(t = 1, 2, \ldots, T)$, and let $H(y) = \Pr(Y \leq y)$ denote its associated cumulative distribution function (c.d.f.). Then $h(\cdot)$ has the representation

$$h(y_1, y_2, \ldots, y_T) = \int_0^1 L(\theta)dF(\theta), \tag{1.4}$$

where

$$L(\theta) = \prod_{t=1}^{T} \theta^{y_t}(1-\theta)^{(1-y_t)}, \tag{1.5}$$

$$F(\theta) = \lim_{T\to\infty} P_H(\overline{Y}_T \leq \theta), \tag{1.6}$$

and $P_H(\cdot)$ denotes probability with respect to the c.d.f. $H(\cdot)$ corresponding to p.m.f. (1.4).

**Proof**: See de Finetti (1937) or the simpler exposition of Heath and Sudderth (1976).

Theorem 1.1 implies that it is *as if*, given $\theta$, $Y_t$ $(t = 1, 2, \ldots, T)$ are iid Bernoulli trials where the probability of a success is $\theta$, and the "parameter" $\theta$ is assigned a probability distribution with c.d.f. $F(\cdot)$ that can be interpreted as belief about the long-run relative frequency of $\overline{Y}_T \leq \theta$ as $T \to \infty$. From de Finetti's standpoint, both the quantity $\theta$ and the notion of independence are "mathematical fictions" implicit in the researcher's subjective assessment of arbitrarily long *observable* sequences of successes and failures. The parameter $\theta$ is of interest primarily because it constitutes a limiting form of predictive inference about the observable $\overline{Y}_T$ via (1.6). The mathematical construct $\theta$ may nonetheless be useful. However, Theorem 1.2 implies that the subjective probability distribution need not apply to the "fictitious $\theta$" but only to the *observable* exchangeable sequence of successes and failures. When the c.d.f. is absolutely continuous, so that $f(\theta) = \partial F(\theta)/\partial\theta$ exists, then (1.4) becomes

$$h(y_1, y_2, \ldots, y_T) = \int_0^1 \prod_{t=1}^{T} \theta^{(y_t)}(1-\theta)^{(1-y_t)} f(\theta) d\theta. \tag{1.7}$$

It is clear from (1.4) and (1.7) that exchangeable beliefs assign probabilities acting as if the $Y_t$'s are iid Bernoulli random variables given $\theta$, and then average over values of $\theta$ using the weight $f(\theta)$ to obtain a marginal density for the $Y_t$'s. Let $S_T = T\overline{Y}_T$ be the number of successes in $T$ trials. Since there are $\binom{T}{r}$ ways in which to obtain $S_T = r$ successes in $T$ trials, it follows immediately from (1.4) and (1.5) that

$$\Pr(S_T = r) = \binom{T}{r} \int_0^1 \theta^r (1-\theta)^{T-r} dF(\theta) \quad (r = 0, 1, \ldots, T), \tag{1.8}$$

where

$$F(\theta) = \lim_{T \to \infty} \Pr(T^{-1} S_T \leq \theta). \tag{1.9}$$

Thus, given $\theta$, it follows from (1.8) that exchangeable beliefs assign probabilities acting as if $S_T$ has a binomial distribution given $\theta$, and then average over values of $\theta$ using the weight $f(\theta) = \partial F(\theta)/\partial\theta$. Bayes and Laplace suggest choosing the "mixing" distribution $F(\theta)$ for $\theta$ to be uniform over $[0, 1]$, in which case (1.8) reduces to

$$\Pr(S_T = r) = (T+1)^{-1}, \quad r = 0, 1, \ldots, T. \tag{1.10}$$

In words, (1.10) describes beliefs that in $T$ trials, any number $r$ of successes are equally likely. In the degenerate case in which the distribution of $\theta$ assigns probability one to some value $\theta_0$, then de Finetti's theorem implies that $S_T$ follows the standard binomial distribution

$$\Pr(S_T = r) = \binom{T}{r} \theta_0^r (1-\theta_0)^{T-r}, \tag{1.11}$$

and (1.9) implies

$$\lim_{T \to \infty} \overline{Y}_T = \theta_0 \tag{1.12}$$

with "probability one." This last result, as a special case of de Finetti's Theorem, is equivalent to the *strong law of large numbers*.

De Finetti's representation theorem has been generalized by seeking more stringent forms of "symmetry" than simple exchangeability, in the process rationalizing sampling models other than the binomial [see Bernardo and Smith (1994, Chapter 4)]. Although these theorems do not hold exactly for infinite sequences, they hold approximately for sufficiently large finite sequences.

The pragmatic value of de Finetti's theorem depends on whether it is easier to assess the left-hand side of (1.8), which involves only observable quantities, or instead, the integrand on the right-hand side of (1.8), which involves two distributions and the mathematical fiction $\theta$. Most statisticians think in terms of the right-hand side. Frequentists implicitly do so with a degenerate distribution for $\theta$ that in effect treats $\theta$ as a constant, and Bayesians do so with a nondegenerate "prior" distribution for $\theta$. What is important to note here, however, is the isomorphism de Finetti's theorem suggests between two worlds, one involving only observables and the other involving the parameter $\theta$. De Finetti put parameters in their proper perspective: (i) They are mathematical constructs that provide a convenient index for a probability distribution, and (ii) they induce conditional independence for a sequence of observables.

**Exercise 1.1 (Let's make a deal)** Consider the television game show "Let's Make a Deal" in which host Monty Hall asks contestants to choose the prize behind one of three curtains. Behind one curtain lies the grand prize; the other two curtains conceal only relatively small gifts. Assume Monty knows what is behind every curtain. Once the contestant has made a choice, Monty Hall reveals what is behind one of the two curtains that were not chosen. Having been shown one of the lesser prizes, the contestant is offered a chance to switch curtains. Should the contestant switch?

***Solution***

Let $C$ denote which curtain hides the grand prize. Let $\hat{C}$ denote the curtain the contestant chooses first, and let $M$ denote the curtain Monty shows the contestant. Assume $\Pr(C = i) = 1/3,\ i = 1,2,3$, $\Pr(\hat{C} = k|C) = 1/3,\ k = 1,2,3$, and that $C$ and $\hat{C}$ are independent. Without loss of generality, suppose $C = 1$ and $M = 2$. Then use Bayes' theorem for events to compute the numerator and denominator of the following ratio:

$$\frac{\Pr(C=3|M=2,\hat{C}=1)}{\Pr(C=1|M=2,\hat{C}=1)} = \frac{\frac{\Pr(M=2,\hat{C}=1|C=3)\Pr(C=3)}{\Pr(M=2,\hat{C}=1)}}{\frac{\Pr(M=2,\hat{C}=1|C=1)\Pr(C=1)}{\Pr(M=2,\hat{C}=1)}} \tag{1.13}$$

$$= \frac{\Pr(M=2,\hat{C}=1|C=3)}{\Pr(M=2,\hat{C}=1|C=1)}$$

$$= \frac{\Pr(M=2|\hat{C}=1,C=3)\Pr(\hat{C}=1|C=3)}{\Pr(M=2|\hat{C}=1,C=1)\Pr(\hat{C}=1|C=1)}$$

$$= \frac{\Pr(M=2|\hat{C}=1,C=3)}{\Pr(M=2|\hat{C}=1,C=1)}.$$

The numerator of the last line of (1.13) is one because Monty has no choice but to choose $M = 2$ when $\hat{C} = 1$ and $C = 3$. The denominator of (1.13), however, is ambiguous because when $\hat{C} = 1$ and $C = 1$, Monty can choose either $M = 2$ or $M = 3$. The problem formulation does not contain information on Monty's choice procedure in this case. But since this probability must be less than or equal to one, ratio (1.13) can never be less than one. Unless $\Pr(M = 2|\hat{C} = 1, C = 1) = 1$, the contestant is better off switching curtains. If $\Pr(M = 2|\hat{C} = 1, C = 1) = \Pr(M = 3|\hat{C} = 1, C = 1) = 1/2$, then the contestant doubles the probability of winning the grand prize by switching.

**Exercise 1.2 (Making Dutch book)** Consider a horse race involving $N$ horses. Suppose a bettor's beliefs are such that he believes the probability of horse $n$ winning is $p_n$, where $p_1 + p_2 + \cdots + p_N < 1$. Show how to make Dutch book with such an individual.

***Solution***

Consider a bet with this person of $p_n$ dollars that pays one dollar if horse $n$ wins, and place such a bet on each of the $N$ horses. Then you are guaranteed winning one dollar (since one of the horses has to win) and earning a profit of $1 - (p_1 + p_2 + \cdots + p_N) > 0$.

**Exercise 1.3 (Independence and exchangeability)** Suppose $Y = [Y_1 \ Y_2 \ \cdots \ Y_T]' \sim N(0_T, \Sigma)$, where $\Sigma = (1 - \alpha)I_T + \alpha\iota_T\iota_T'$ is positive definite for some scalar $\alpha$ and $\iota$ is a $T \times 1$ vector with each element equal to unity. Let $\pi(t)$ $(t = 1, 2, \ldots, T)$ be a permutation of $\{1, 2, \ldots, T\}$ and suppose $[Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(T)}] = AY$, where $A$ is a $T \times T$ selection matrix such that, for $t = 1, 2, \ldots, T$, row $t$ in $A$ consists of all zeros except column $\pi(t)$, which is unity. Show that these beliefs are exchangeable.

***Solution***

Note that $AA' = I_T$ and $A\iota_T = \iota_T$. Then, $AY \sim N(0_T, \Omega)$, where

$$
\begin{aligned}
\Omega &= A\Sigma A' \\
&= A[(1-\alpha)I_t + \alpha\iota_T\iota_T']A' \\
&= (1-\alpha)AA' + \alpha A\iota_T\iota_T'A' \\
&= (1-\alpha)I_T + \alpha\iota_T\iota_T' \\
&= \Sigma.
\end{aligned}
$$

Hence, beliefs regarding $Y_t (t = 1, 2, \ldots, T)$ are exchangeable. Despite this exchangeability, it is interesting to note that if $\alpha \neq 0$, $Y_t$ $(t = 1, 2, \ldots, T)$ are *not independent*.

**Exercise 1.4 (Predicted probability of success of a Bernoulli random variable)** Suppose a researcher makes a coherent probability assignment to an infinite sequence $Y_t (t = 1, 2, 3, \ldots)$ of exchangeable Bernoulli random variables. Given an observed sequence of $T$ trials with $r$ successes, find the probability that the next outcome, $Y_{T+1}$, is $y_{T+1}$.

***Solution***
Applying the definition of conditional probability and then Theorem 1.2 to both the numerator and denominator yields

$$\begin{aligned}\Pr(Y_{T+1} = y_{T+1}|T\overline{Y}_T = r) &= \frac{\Pr(T\overline{Y}_T = r, Y_{T+1} = y_{T+1})}{\Pr(T\overline{Y}_T = r)} \\ &= \frac{\int_0^1 \theta^{(r+y_{T+1})}(1-\theta)^{(T+1-r-y_{T+1})}p(\theta)d\theta}{\int_0^1 \theta^r(1-\theta)^{(T-r)}p(\theta)d\theta} \\ &= \frac{\int_0^1 \theta^{(y_{T+1})}(1-\theta)^{(1-Y_{T+1})}p(\theta)L(\theta)d\theta}{\int_0^1 L(\theta)p(\theta)d\theta} \\ &= \int_0^1 \theta^{(y_{T+1})}(1-\theta)^{(1-y_{T+1})}p(\theta|y)d\theta,\end{aligned} \tag{1.14}$$

where

$$p(\theta|y) = \frac{p(\theta)L(\theta)}{p(y)}. \tag{1.15}$$

Therefore $\Pr(Y_{T+1} = y_{T+1}|T\overline{Y}_T = r)$ is simply

$$E(\theta|y) \text{ if } y_{T+1} = 1,$$

or

$$1 - E(\theta|y) \text{ if } y_{T+1} = 0.$$

The simplicity of this exercise hides its importance because it demonstrates most of the essential operations that characterize the Bayesian approach to statistics. First, the existence of the density $p(\theta)$ is a result of Theorem 1.2, *not* an assumption. Second, the updating of prior beliefs captured in (1.15) amounts to nothing more than Bayes' theorem. Third, although $Y_t$ $(t = 1, 2, \ldots, T)$ are independent conditional on $\theta$, unconditional on $\theta$ they are dependent. Finally, the parameter $\theta$ is merely a mathematical entity indexing the integration in (1.14). Its "real-world existence" is a question only of metaphysical importance.

**Exercise 1.5 (Independence and conditional independence)** Consider three events $A_i$ $(i = 1, 2, 3)$, where $\Pr(A_i) = p_i,\ i = 1, 2, 3$. Show that the following statements are totally unrelated: (a) $A_1$ and $A_2$ are independent and (b) $A_1$ and $A_2$ are conditionally independent given $A_3$.

***Solution***
There are $2^3 = 8$ possible three-element strings that can occur when considering $A_i$ $(i = 1, 2, 3)$ and their complements $A_i^c$ $(i = 1, 2, 3)$. This leaves assessment of $7 = 8 - 1$ probabilities since the eighth is determined by the adding-up condition. These can be assessed in terms of the following probabilities: $\Pr(A_1 \cap A_2) = q_{12}$, $\Pr(A_1 \cap A_3) = q_{13}$,

$\Pr(A_2 \cap A_3) = q_{23}$, and $\Pr(A_1 \cap A_2 \cap A_3) = s$. Independence of $A_1$ and $A_2$ places a restriction on $\Pr(A_1 \cap A_2)$, namely $q_{12} = p_1 p_2$. Conditional independence places a restriction on the remaining probabilities $q_{13},\ q_{23}, p_3$, and $s$. To see this note $\Pr(A_1 \cap A_2 | A_3) = s/p_3$ by simply expressing the conditional as the joint divided by the marginal, and conditional independence implies $\Pr(A_1 \cap A_2 | A_3) = \Pr(A_1 | A_3)\Pr(A_2 | A_3) = (q_{13}/p_3)(q_{23}/p_3)$. Putting these equalities together implies $s = q_{13} q_{23}/p_3$. Note that the restrictions implied by independence and conditional independence share no common probabilities.

# 2

# Bayesian inference

In this chapter we extend Chapter 1 to cover the case of random variables. By *Bayesian inference* we mean the updating of prior beliefs into posterior beliefs conditional on observed data. This chapter covers a variety of standard sampling situations in which prior beliefs are sufficiently regular that the updating can proceed in a fairly mechanical fashion. Details of point estimation, interval estimation, hypothesis testing, and prediction are covered in subsequent chapters. We remind the reader that the definitions of many common distributions are provided in the Appendix to this book. Further details on the underlying probability theory are available in Chapters 1 and 2 of Poirier (1995).

One of the appealing things about Bayesian analysis is that it requires only a few general principles that are applied over and over again in different settings. Bayesians begin by writing down a joint distribution of all quantities under consideration (except known constants). Quantities to become known under sampling are denoted by the $T$-dimensional vector $y$, and remaining unknown quantities by the $K$-dimensional vector $\theta \in \Theta \subseteq \mathcal{R}^K$. Unless noted otherwise, we treat $\theta$ as a continuous random variable. Working in terms of densities, consider

$$p(y, \theta) = p(\theta)p(y|\theta) = p(y)p(\theta|y), \tag{2.1}$$

where $p(\theta)$ is the *prior density* and $p(\theta|y)$ is the *posterior density*. Viewing $p(y|\theta)$ as a function of $\theta$ for known $y$, any function proportional to it is referred to as a *likelihood function*. We will denote the likelihood function as $L(\theta)$. Unless noted otherwise, we will work with $L(\theta) = p(y|\theta)$ and thus include the integrating constant for $y|\theta$ in our description of the likelihood. We also note that

$$p(y) = \int_{\Theta} p(\theta)L(\theta)d\theta \tag{2.2}$$

is the *marginal density of the observed data* (also known as the *marginal likelihood*).

From (2.1) Bayes' *theorem for densities* follows immediately:

$$p(\theta|y) = \frac{p(\theta)L(\theta)}{p(y)} \propto p(\theta)L(\theta). \tag{2.3}$$

The shape of the posterior can be learned by plotting the right-hand side of (2.3) when $k = 1$ or 2. Obtaining moments or quantiles, however, requires the integrating constant (2.2). Fortunately, in some situations the integration in (2.2) can be performed analytically, in which case the updating of prior beliefs $p(\theta)$ in light of the data $y$ to obtain the posterior beliefs $p(\theta|y)$ is straightforward. These situations correspond to cases where $p(\theta)$ and $L(\theta)$ belong to the *exponential family* of densities (see Exercise 2.13). In this case the prior density can be chosen so that the posterior density falls within the same elementary family of distributions as the prior. These families are called *conjugate families*.

The denominator in (2.3) serves as a factor of proportionality (not involving $\theta$) that ensures that the posterior density integrates to unity. To simplify much of the analysis that follows, we calculate a posterior density by dropping all factors of proportionality from the prior density and the likelihood function, concentrating attention on the resulting posterior kernel, and then compute the required posterior integrating constant at the end. This works particularly well when using easily recognized conjugate families. Note also that this implies that when considering experiments employing the same prior, and that yield proportional likelihoods for the observed data, identical posteriors will emerge. This reflects the important fact that Bayesian inference is consistent with the *likelihood principle* [see Berger and Wolpert (1988)].

In most practical situations not all elements of $\theta$ are of direct interest. Let $\theta_1 \in \Theta_1,\ \theta_2 \in \Theta_2$, and $\theta = [\theta_1, \theta_2] \in\ \Theta_1 \times \Theta_2$ be partitioned into *parameters of interest* $\theta_1$ and *nuisance parameters* $\theta_2$ not of direct interest. For example, $\theta_1$ may be the mean and $\theta_2$ the variance of some sampling distribution. Nuisance parameters are well named for frequentists, because dealing with them in a general setting is one of the major problems frequentist researchers face. In contrast, Bayesians have a universal approach to eliminating nuisance parameters from the problem: They are integrated out of the posterior density, yielding the marginal posterior density for the parameters of interest, that is,

$$p(\theta_1|y) = \int_{\Theta_2} p(\theta_1, \theta_2|y)\, d\theta_2, \quad \theta_1 \in \Theta_1. \tag{2.4}$$

Many of the following exercises involve particular distributions. The Appendix of this book contains definitions and properties of many common distributions. It is worth noting that there are two common parameterizations of the gamma distribution and we use both in this chapter (see Appendix Definition 2).

**Exercise 2.1 (Conjugate Bernoulli analysis)** Given the parameter $\theta$, where $0 < \theta < 1$, consider $T$ iid Bernoulli random variables $Y_t$ $(t = 1, 2, \ldots, T)$, each with p.m.f.

$$p(y_t|\theta) = \begin{cases} \theta & \text{if } y_t = 1, \\ 1-\theta & \text{if } y_t = 0. \end{cases} \tag{2.5}$$

The likelihood function is

$$L(\theta) = \theta^m (1-\theta)^{T-m}, \tag{2.6}$$

where $m = T\overline{y}$ is the number of successes (i.e., $y_t = 1$) in $T$ trials. Suppose prior beliefs concerning $\theta$ are represented by a beta distribution with p.d.f.

$$p_B(\theta|\underline{\alpha}, \underline{\delta}) = [\mathcal{B}(\underline{\alpha}, \underline{\delta})]^{-1} \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1}, \quad 0 < \theta < 1, \tag{2.7}$$

where $\underline{\alpha} > 0$ and $\underline{\delta} > 0$ are known, and $\mathcal{B}(\underline{\alpha}, \underline{\delta}) = \Gamma(\underline{\alpha})\Gamma(\underline{\delta})/\Gamma(\underline{\alpha} + \underline{\delta})$ is the beta function defined in terms of the gamma function $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt$. This class of priors can represent a wide range of prior opinions. Find the posterior density of $\theta$.

***Solution***

The denominator (2.2) of posterior (2.3) density is easy to compute. Define

$$\overline{\alpha} = \underline{\alpha} + m, \tag{2.8}$$

$$\overline{\delta} = \underline{\delta} + T - m, \tag{2.9}$$

and consider

$$\begin{aligned} p(y) &= \int_0^1 [\mathcal{B}(\underline{\alpha}, \underline{\delta})]^{-1} \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1} \theta^m (1-\theta)^{T-m} d\theta \qquad (2.10) \\ &= \left[\frac{\mathcal{B}(\overline{\alpha}, \overline{\delta})}{\mathcal{B}(\underline{\alpha}, \underline{\delta})}\right] \int_0^1 [\mathcal{B}(\overline{\alpha}, \overline{\delta})]^{-1} \theta^{\overline{\alpha}-1}(1-\theta)^{\overline{\delta}-1} d\theta \\ &= \left[\frac{\mathcal{B}(\overline{\alpha}, \overline{\delta})}{\mathcal{B}(\underline{\alpha}, \underline{\delta})}\right], \end{aligned}$$

where the integral in (2.10) equals unity because the integrand is a beta p.d.f. for $\theta$. From (2.3) and (2.8)–(2.10) it follows that the posterior density of $\theta$ is

$$\begin{aligned} p(\theta|y) &= \frac{[\mathcal{B}(\underline{\alpha}, \underline{\delta})]^{-1} \theta^{\underline{\alpha}-1}(1-\theta)^{\underline{\delta}-1} \theta^m (1-\theta)^{T-m}}{\mathcal{B}(\overline{\alpha}, \overline{\delta})/\mathcal{B}(\underline{\alpha}, \underline{\delta})} \qquad (2.11) \\ &= [\mathcal{B}(\overline{\alpha}, \overline{\delta})]^{-1} \theta^{\overline{\alpha}-1}(1-\theta)^{\overline{\delta}-1}, \qquad 0 < \theta < 1. \end{aligned}$$

Therefore, because posterior density (2.11) is itself a beta p.d.f. with parameters $\overline{\alpha}$ and $\overline{\delta}$ given by (2.8) and (2.9), it follows that the conjugate family of prior distributions for a Bernoulli likelihood is the beta family of p.d.f.s

**Exercise 2.2 (Application of Bayes' theorem)** A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for one percent of the healthy people tested. If 0.1 percent of the population actually has the disease, what is the probability that a person has the disease given that her test result is positive?

***Solution***

Let $D$ denote the presence of the disease, $D^c$ denote its absence, and $+$ denote a positive test result. Then $\Pr(+|D) = .95$, $\Pr(+|D^c) = .01$, and $P(D) = .001$. Then according to Bayes' theorem

$$P(D|+) = \frac{P(D)P(+|D)}{P(+)} = \frac{.001(.95)}{.001(.95) + .999(.01)} = .0868.$$

**Exercise 2.3 (Conjugate normal analysis with unknown mean and known variance)** Given $\theta = [\theta_1\ \theta_2]' \in \mathcal{R} \times \mathcal{R}^+$, consider a random sample $Y_t$ $(t = 1, 2, \ldots, T)$ from a $N(\theta_1, \theta_2^{-1})$ population. For reasons that will become clear as we proceed, it is convenient to work in terms of $\theta_2$, the reciprocal of the variance (called the *precision*). (In later exercises, however, particularly those in the computational chapters, we will work directly with the error variance. Both approaches are commonly employed in the literature).

Assume $\theta_2$ is known. Suppose prior beliefs for $\theta_1$ are represented by the normal distribution

$$\theta_1 \sim N(\underline{\mu}, \underline{h}^{-1}), \tag{2.12}$$

where $\underline{\mu}$ and $\underline{h} > 0$ are given. Find the posterior density of $\theta_1$ and marginal likelihood $p(y)$.

***Solution***

For later notational convenience, let

$$h = [\theta_2^{-1}/T]^{-1} = T\theta_2, \tag{2.13}$$

$$\overline{h} = \underline{h} + h, \tag{2.14}$$

and

$$\overline{\mu} = \overline{h}^{-1}(\underline{h}\underline{\mu} + h\overline{y}). \tag{2.15}$$

It is useful to employ two identities. The first identity is

$$\sum_{t=1}^{T}(y_t - \theta_1)^2 = \sum_{t=1}^{T}(y_t - \overline{y})^2 + T(\overline{y} - \theta_1)^2 = \nu s^2 + T(\overline{y} - \theta_1)^2, \tag{2.16}$$

for all $\theta_1$, where

$$\nu = T - 1 \tag{2.17}$$

and

$$s^2 = \nu^{-1}\sum_{t=1}^{T}(y_t - \overline{y})^2. \tag{2.18}$$

The second identity is

$$\underline{h}(\theta_1 - \underline{\mu})^2 + h(\overline{y} - \theta_1)^2 = \overline{h}(\theta_1 - \overline{\mu})^2 + (\underline{h}^{-1} + h^{-1})^{-1}(\overline{y} - \underline{\mu})^2, \tag{2.19}$$

for all $\theta_1$, $\underline{h}$, and $h$.

Now we apply Bayes' theorem to find the posterior density of $\theta_1$. Using identities (2.13) and (2.16), we write the likelihood function as

$$\begin{aligned} L(\theta_1) &= \prod_{t=1}^{T} \phi(y_t|\theta_1, \theta_2^{-1}) \\ &= (2\pi\theta_2^{-1})^{-T/2} \exp\left(-\frac{\theta_2}{2}\sum_{t=1}^{T}(y_t - \theta_1)^2\right) \\ &= (2\pi\theta_2^{-1})^{-T/2} \exp\left(-\frac{h}{2T}[\nu s^2 + T(\overline{y} - \theta_1)^2]\right) \\ &= c_1\phi(\overline{y}|\theta_1, h^{-1}), \end{aligned} \tag{2.20}$$

where

$$c_1 = (2\pi)^{-\nu/2} T^{(-1/2)} \theta_2^{\nu/2} \exp\left(-\frac{1}{2}\theta_2 \nu s^2\right) \tag{2.21}$$

does *not* depend on $\theta_1$. Note that the factorization in (2.20) demonstrates that $\overline{y}$ is a sufficient statistic for $\theta_1$. Also note that density $\phi(\overline{y}|\theta_1, h^{-1})$ corresponds to the sampling density of the sample mean, given $\theta_1$.

Using identity (2.19) and factorization (2.20), the numerator of (2.3) is

$$\begin{aligned} p(\theta_1)L(\theta_1) &= \phi(\theta_1|\underline{\mu}, \underline{h}^{-1})c_1\phi(\overline{y}|\theta_1, h^{-1}) \\ &= c_1(2\pi\underline{h}^{-1})^{-1/2} \exp\left(-\frac{1}{2}\left[\underline{h}(\theta_1 - \underline{\mu})^2 + h(\overline{y} - \theta_1)^2\right]\right) \\ &= c_1(2\pi\underline{h}^{-1})^{-1/2} \exp\left(-\frac{1}{2}\left[\overline{h}(\theta_1 - \overline{\mu})^2 + (\underline{h}^{-1} + h^{-1})^{-1}(\overline{y} - \underline{\mu})^2\right]\right) \\ &= c_1(2\pi)^{(1/2)}[\underline{h}\overline{h}^{-1}(\underline{h}^{-1} + h^{-1})]^{1/2}\phi\left(\overline{y}|\underline{\mu}, \underline{h}^{-1} + h^{-1}\right)\phi(\theta_1|\overline{\mu}, \overline{h}^{-1}). \end{aligned} \tag{2.22}$$

Bayes' theorem tells us that the posterior distribution of $\theta_1$ is proportional to (2.22). Thus, all terms not involving $\theta_1$ that enter (2.22) multiplicatively are absorbed in the normalizing constant, allowing us to focus on the posterior kernel. When looking at the final line of (2.22), we immediately see that

$$p(\theta_1|y) = \phi\left(\theta_1|\overline{\mu}, \overline{h}^{-1}\right). \tag{2.23}$$

The interpretations of quantities (2.14) and (2.15) are now clear from (2.23): They are the posterior precision and posterior mean, respectively. Note that it is the additivity of precisions in (2.14) that motivates working with precisions rather than variances. Because posterior density (2.23) and prior density (2.12) are both members of the normal family, it follows that the conjugate prior for the case of random sampling from a normal population with *known* variance is itself a normal density.