Springer Texts in Statistics

For further volumes:
http://www.springer.com/series/417

Mary Kathryn Cowles

Applied Bayesian Statistics

With R and OpenBUGS Examples

Springer

Mary Kathryn Cowles
Department of Statistics
and Actuarial Science
University of Iowa
Iowa City, Iowa, USA

ISSN 1431-875X
ISBN 978-1-4614-5695-7 ISBN 978-1-4614-5696-4 (eBook)
DOI 10.1007/978-1-4614-5696-4
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012951150

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Brendan, Lucy, and Donald.

Preface

I have taught a course called "Bayesian Statistics" at the University of Iowa every academic year since 1998–1999. This book is intended to fit the goals and audience addressed by my course. The "Course Objectives" section of my syllabus reads:

> Through hands-on experience with real data from a variety of applications, students will learn the basics of designing and carrying out Bayesian analyses, and interpreting and communicating the results. Students will learn to use software packages including R and OpenBUGS to fit Bayesian models.

The course is intended to be intensely practical, focussing on building understanding of the concepts and procedures required to perform Bayesian analysis of real data to answer real questions. Emphasis is given to such issues as determining what data is needed to address a particular question; choosing an appropriate probability distribution for sample data; quantifying already-existing knowledge in the form of a prior distribution on model parameters; verifying that the posterior distribution will be proper if improper prior distributions are used; and when and how to specify hierarchical models. Interpretation and communication of results are stressed, including differences from, and similarities to, classical approaches to the same problems.

WinBUGS and OpenBUGS currently are the dominant software in applied use of Bayesian methods. I have chosen to introduce OpenBUGS as the primary data analysis software in this textbook because, unlike WinBUGS, OpenBUGS is undergoing continuing development and has versions that run natively under Linux and Macintosh operating systems as well as Windows. Although some background is provided on the Markov chain Monte Carlo sampling procedures employed by WinBUGS and OpenBUGS, the emphasis is on those tasks that a *user* must carry out correctly for reasonably trustworthy inference. These include using appropriate tools to assess whether and when a sampler has converged to the target distribution, deciding how many iterations are needed for acceptable accuracy in estimation, and how to report results of a Bayesian analysis conducted with OpenBUGS. Caveats about the fallibility of convergence diagnostics are emphasized.

Students of different levels and disciplines take the course, including: undergraduate mathematics and statistics majors; master's students in statistics, biostatistics, statistical genetics, educational testing and measurement, and engineering; and PhD students in economics, marketing, psychology, and geography as well as the previously listed fields. In addition, several practicing statisticians employed by the University of Iowa and American College Testing (ACT) have taken the course.

The goal of the course, and of this book, is to provide an introduction to Bayesian principles and practice that is clear, useful, and unintimidating to motivated students even if they do not have an advanced background in mathematics and probability. I emphasize intuitive insight without sacrificing mathematical correctness. Prerequisites are one or two semesters of calculus-based probability and mathematical statistics (at least at the Hogg and Tannis level) and one or two semesters of classical statistical methods, including linear regression (David Moore's *Basic Practice of Statistics* level). Elementary integral and differential calculus is occasionally used in lectures and homework. Linear algebra is not required.

Coralville, Iowa Mary Kathryn Cowles

Contents

Chapter 1
What Is Bayesian Statistics?

1.1 The Scientific Method (But It Is Not Just for Science…)

In almost every field of human activity, people use data to further their learning and to guide decision-making and action. The following steps, paraphrased from Berry (1996), have been described as "the scientific method." However, they are equally appropriate for use by a biologist seeking to better understand the behavior of monarch butterflies, the marketing director of a grocery store chain determining where to open a new store, or a new university graduate deciding whether to accept a particular job offer.

1. Define the question or problem to be addressed.
2. Assess the relevant information already available. Decide whether it is sufficient for the purpose at hand.
   a. If yes, draw appropriate conclusions, make appropriate decisions, and take appropriate action.
   b. If no, proceed to step 3.
3. Determine what additional information is needed and design a study or experiment to attempt to obtain it.
4. Carry out the study designed in step 3.
5. Use the data obtained in step 4 to update what was previously known. Return to step 2.

Statistics is central to steps 2, 3, and 5. *Bayesian* statistics is particularly well suited to steps 2 and 5, because it provides a quantitative framework for representing current knowledge and for rationally integrating new information.

M.K. Cowles, *Applied Bayesian Statistics: With R and OpenBUGS Examples*,
Springer Texts in Statistics 98, DOI 10.1007/978-1-4614-5696-4_1,

1.2 A Bit of History

One could argue that the Great Fire of London in 1666 sparked the philosophy and methods that now are called Bayesian statistics. Destroying over 13,000 houses, 89 churches, and dozens of public buildings, the Great Fire led to the rise of insurance protection as we understand it today. The following year, one Nicholas Barbon opened an office to insure buildings, and in 1680, he established the first full-fledged fire insurance company in England. By the early eighteenth century, the idea of life insurance as well as property insurance was taking hold in England. However, lack of adequate vital statistics and of probability theory led to the failure of many early life insurers.

Enter Thomas Bayes. Born in London in 1702, Bayes became an ordained Presbyterian minister by profession and a mathematician and scientist by avocation. He applied his mind to the questions urgently raised by the insurers and laid out his resulting theory of probability in his *Essay towards solving a problem in the doctrine of chances*. After Bayes' death in 1761, his friend Richard Price sent the paper to the Royal Society of London.The paper was published in the Philosophical Transactions of the Royal Society of London in 1764.

Bayes' conclusions were accepted enthusiastically by Pierre-Simon Laplace and other contemporary leading probabilists. However, George Boole questioned them in his 1854 treatise on logic called *Laws of Thought*. Bayes' method became controversial in large part because mathematicians and scientists did not yet know how to treat prior probabilities (a topic that we will deal with throughout this book!). In the first half of the twentieth century, a different approach to statistical inference arose, which has come to be called the *frequentist* school. However, Bayesian thinking continued to progress with the works of Bruno de Finetti in Italy, Harold Jeffreys and Dennis Lindley in England, Jimmy Savage in the USA, and others.

Until about 1990, the application of Bayesian methods to statistical analysis in real-world problems was very limited because the necessary mathematical computations could be done analytically only for very simple models. In the early 1990s, the increasing accessibility of powerful computers, along with the development of new computing algorithms for fitting Bayesian models, opened the door to the use of Bayesian methods in complex, real-world applications. The subsequent explosion of interest in Bayesian statistics has led not only to extensive research in Bayesian methodology but also to the use of Bayesian methods to address pressing questions in diverse application areas such as astrophysics, weather forecasting, health-care policy, and criminal justice.

Today, Bayesian statistics is widely used to guide learning and decision-making in business and industry as well as in science. For example, software using Bayesian analysis guides Google's driverless robotic cars (McGrayne 2011a), and Bayesian methods have attained sufficiently wide acceptance in medical research that, in 2006, the United States Food and Drug Administration (FDA) put into place a set of guidelines for designing clinical trials of medical devices using Bayesian methods ("Guidance for the Use of Bayesian Statistics in Medical Device

Clinical Trials", http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm). McGrayne (2011b) offers a lively introduction to the history and current status of Bayesian statistics.

1.3 Example of the Bayesian Method: Does My Friend Have Breast Cancer?

The National Cancer Institute recommends that women aged 40 and above should have mammograms every one to two years. A mammogram produces x-ray images of tissues and structure inside the breast and may help detect and identify cancerous tumors, benign cysts, and other breast conditions. A mammogram administered to a woman who has no signs or symptoms associated with breast cancer is called a "screening mammogram."

A friend of mine recently was referred by her physician for her first screening mammogram. My friend does not have a family history of breast cancer, and before being referred for the screening mammogram, she had given no thought whatsoever to breast cancer as something that could conceivably happen to her. However, as the date of the mammogram approached, she began to wonder about her chances of being diagnosed with breast cancer. She was at step 1 of the scientific method—she had defined a question that she wanted to address. In other words, she was *uncertain* about her status with respect to breast cancer and wanted to learn more about it.

In the next sections, we will couch my friend's learning process in the framework of the simplest possible application of Bayes' rule within the scientific method. We will introduce the notion of using probabilities to quantify knowledge or uncertainty and of using data to update such probabilities in a rational way.

1.3.1 Quantifying Uncertainty Using Probabilities

In science, business, and daily life, people quantify uncertainty in the form of probabilities. The weather forecaster says there is a 30% probability of precipitation today; the seismologist says that there is a 21% chance of a major earthquake along the San Andreas fault by the year 2032; a doctor may tell a cancer patient that she has a 50% chance of surviving for 5 years or longer. Two different interpretations of *probability* are in common use.

1.3.1.1 The Long-Run Frequency Interpretation of Probability

In previous statistics or math classes, you undoubtedly have encountered the long-run frequency interpretation of the probability of an event. For example, Moore (2007, page 248) says:

> The probability of any outcome of a random phenomenon is the proportion of the times the outcome would occur in a very long series of repetitions.

Coin flipping is an obvious example of this interpretation of probability. Saying that the probability of a fair coin coming up heads is 0.5 means that we expect to get a head about half the time if we flip the coin a huge number of times under exactly the same conditions.

Although this interpretation is useful (and the *frequentist* or *classical* approach to statistics is based on it), it has serious shortcomings. Trying to use the long-run frequency interpretation as a *definition* of probability results in a circular argument; see Woodworth (2004, page 25) for a summary of the mathematical issue. From a more intuitive standpoint, the frequency interpretation is limited to situations in which a sequence of repeatable experiments is possible (or at least imaginable). No frequency interpretation is possible for probabilities of many kinds of events about which we would like to quantify uncertainty. For example, in the winter of 2007–2008, editorial writers were assessing the probability that the United States economy was headed for a major recession. Thousands of homes purchased in the previous several years were in foreclosure, and some mortgage companies had gone bankrupt due to bad loans. The price of oil was over $100 a barrel. Although everyone certainly wanted to know the probability that a recession was coming, obviously, the question could not be couched as the proportion of the time that countries facing exactly the economic, social, and political conditions that then existed in the USA would go into recession.

1.3.1.2 Subjective Probability

The subjective interpretation of probability is:

> A probability of an event or of the truth of a statement is a number between 0 and 1 that quantifies a particular person's subjective opinion as to how likely that event is to occur (or to have already occurred) or how likely the statement is to be true.

This interpretation of probability clearly is not limited to repeatable events. Note that the subjective interpretation was of "a" probability, not "the" probability, of an event or statement. Not only may different people have different subjective probabilities regarding the same event, but the same person's subjective probability is likely to change as more information becomes available. (As we will see shortly, these updates to a person's subjective probability are where the mathematical identity called Bayes' rule comes in.)

Some people object to admitting that there is any place for subjectivity in science. However, that does not make it any less true that two different scientists can look at the same data (experimental results, observational results, or whatever) and come to different conclusions because of their previously acquired knowledge of the subject.

Here is a hypothetical example in which your own life experience and knowledge of the world might lead you to different conclusions from identical experimental results obtained from different applications. Suppose that I tell you that I have carried out a study consisting of 6 trials. Each trial has only two possible outcomes, which I choose to call "success" and "failure." I believe that the probability of success was the same for each trial and that the six trials were independent. I tell you that the outcome was six successes in six trials, and I ask you to predict the outcome if I carry out a seventh trial of my experiment.

After you think about this for a while, I offer to tell you what the experiment was. I explain that I randomly selected six dates from the 2011 calendar year, and on each of those dates, I looked out the window at 11:00 a.m. If I saw daylight, I recorded a success; if it was completely dark outside, I recorded a failure. (By the way, I live in Iowa, not in the extreme northerly or southerly latitudes.) Does this information affect your prediction regarding the outcome of a seventh trial?

Suppose that, instead, I explained my study as follows. Over the last 10 years, a family in my neighborhood has had six children. Each time a new baby was born, I asked the mother whether it was a boy or a girl. If the baby was a girl, I recorded a success; if a boy, I recorded a failure. (Obviously the choice of which gender to designate a "success" and which a "failure" is completely arbitrary!) Now the mother is pregnant again. Would your assessment of whether the seventh trial is likely to be another success be different in this case compared to the previous case of observing daylight?

1.3.1.3 Properties of Probabilities

To help my friend assess her chances of being diagnosed with breast cancer, we must recall two of the elementary properties of probability. Regardless of which interpretation of probability is being used, these must hold:

- Probabilities must not be negative. If A is any event and $P(A)$ denotes "the probability that A occurs," then

$$P(A) \geq 0$$

- All possible outcomes of an experiment or random phenomenon, taken together, must have probability 1.

1.3.2 Models and Prior Probabilities

For my friend, there are two possible true states of the world:

1. She has breast cancer.
2. She does not have breast cancer.

We may refer to these statements as *models* or *hypotheses*—statements about a certain aspect of the real world, which could predict observable data.

Before obtaining any data specifically about her own breast cancer status, it would be rational for my friend to figure that her chance of being discovered to

Table 1.1 Models and prior probabilities

Model	Prior probability
Breast cancer	0.0045
No breast cancer	0.9955

have breast cancer is similar to that of a randomly selected person in the population of all women who undergo screening mammograms. Her physician tells her that published studies (Poplack et al. 2000) have shown that, among women who have a screening mammogram, the proportion who are diagnosed with breast cancer within 1 year is about 0.0045 (4.5 per 1,000).

Therefore, my friend assigns the following *prior probabilities* to the two models (Table 1.1).

Note that at this point, my friend has carried out step 2 in the scientific method from Sect. 1.1. Prior probabilities refer to probabilities assessed *before* new data are gathered in step 3.

1.3.3 Data

By actually having the mammogram, my friend will go on to step 3: She will collect *data*. Although there are several specific possible results of a mammogram, they may be grouped into just two possible outcomes: A "positive" mammogram indicates possible or likely cancer and results in a recommendation of further diagnostic procedures, and a "negative" mammogram does not give any evidence of cancer or any need for further procedures. We will use the notation $D+$ ($D-$) to represent the event that a person has (does not have) breast cancer and $M+$ ($M-$) to indicate the event that the person has a positive (negative) mammogram result.

The probabilities of the two possible mammogram outcomes are different depending on whether a person has breast cancer or not—that is, depending on which model is correct. In this case, these probabilities are properties of the particular test being used. A perfect screening test would always have a positive result if the person had the disease and a negative result if not. That is, for a perfect test, $P(M+|D+)$ would equal 1, and $P(M+|D-)$ would equal 0. (In this standard notation for conditional probability, the vertical bar is read "given" or "conditional on.") However, perfect tests generally do not exist. A study (Poplack et al. 2000) of tens of thousands of people who received screening mammograms found that, if a person does have breast cancer (i.e., has a confirmed diagnosis of breast cancer within 1 year after the mammogram), then the probability that the screening mammogram will be positive is about 0.724. That is, $P(M+|D+) = 0.724$. This is called the *sensitivity* of a screening mammogram. Furthermore, for screening mammograms, $P(M+|D-) = 0.027$. The probabilities of a negative test under the two models are $P(M-|D-) = 0.973$ and $P(M-|D+) = 0.276$.

We can summarize all of this in Table 1.2. [These and subsequent tables are similar in structure to tables in Albert (1997).]