# INTRODUCTION TO BAYESIAN STATISTICS

THIRD EDITION

WILLIAM M. BOLSTAD James M. Curran



## INTRODUCTION TO BAYESIAN STATISTICS

# INTRODUCTION TO BAYESIAN STATISTICS

Third Edition

WILLIAM M. BOLSTAD JAMES M. CURRAN



Copyright © 2017 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey. Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 978-1-118-09315-8

Printed in the United States of America

 $10 \hspace{0.2cm} 9 \hspace{0.2cm} 8 \hspace{0.2cm} 7 \hspace{0.2cm} 6 \hspace{0.2cm} 5 \hspace{0.2cm} 4 \hspace{0.2cm} 3 \hspace{0.2cm} 2 \hspace{0.2cm} 1$ 

This book is dedicated to

Sylvie, Ben, Rachel, Emily, Mary, and Elizabeth

## Contents

Preface

1	Intro	oduction to Statistical Science	1
	1.1	The Scientific Method: A Process for Learning	3
	1.2	The Role of Statistics in the Scientific Method	5
	1.3	Main Approaches to Statistics	5
	1.4	Purpose and Organization of This Text	8
2	2 Scientific Data Gathering		13
	2.1	Sampling from a Real Population	14
	2.2	Observational Studies and Designed Experiments	17
		Monte Carlo Exercises	23
3	3 Displaying and Summarizing Data		
	3.1	Graphically Displaying a Single Variable	32
	3.2	Graphically Comparing Two Samples	39
	3.3	Measures of Location	41
	3.4	Measures of Spread	44
			vii

xiii

	3.5	Displaying Relationships Between Two or More Variables	46
	3.6	Measures of Association for Two or More Variables	49
		Exercises	52
4	Logic,	Probability, and Uncertainty	59
	4.1	Deductive Logic and Plausible Reasoning	60
	4.2	Probability	62
	4.3	Axioms of Probability	64
	4.4	Joint Probability and Independent Events	65
	4.5	Conditional Probability	66
	4.6	Bayes' Theorem	68
	4.7	Assigning Probabilities	74
	4.8	Odds and Bayes Factor	75
	4.9	Beat the Dealer	76
		Exercises	80
5	Discre	ete Random Variables	83
	5.1	Discrete Random Variables	84
	5.2	Probability Distribution of a Discrete Random Variable	86
	5.3	Binomial Distribution	90
	5.4	Hypergeometric Distribution	92
	5.5	Poisson Distribution	93
	5.6	Joint Random Variables	96
	5.7	Conditional Probability for Joint Random Variables	100
		Exercises	104
6	Bayes	ian Inference for Discrete Random Variables	109
	6.1	Two Equivalent Ways of Using Bayes' Theorem	114
	6.2	Bayes' Theorem for Binomial with Discrete Prior	116
	6.3	Important Consequences of Bayes' Theorem	119
	6.4	Bayes' Theorem for Poisson with Discrete Prior	120
		Exercises	122
		Computer Exercises	126
7	Conti	nuous Random Variables	129
	7.1	Probability Density Function	131
	7.2	Some Continuous Distributions	135
	7.3	Joint Continuous Random Variables	143

	7.4	Joint Continuous and Discrete Random Variables	144
		Exercises	147
8	Baye	sian Inference for Binomial Proportion	149
	8.1	Using a Uniform Prior	150
	8.2	Using a Beta Prior	151
	8.3	Choosing Your Prior	154
	8.4	Summarizing the Posterior Distribution	158
	8.5	Estimating the Proportion	161
	8.6	Bayesian Credible Interval	162
		Exercises	164
		Computer Exercises	167
9	Com	paring Bayesian and Frequentist Inferences for Proportion	169
	9.1	Frequentist Interpretation of Probability and Parameters	170
	9.2	Point Estimation	171
	9.3	Comparing Estimators for Proportion	174
	9.4	Interval Estimation	175
	9.5	Hypothesis Testing	178
	9.6	Testing a One-Sided Hypothesis	179
	9.7	Testing a Two-Sided Hypothesis	182
		Exercises	187
		Monte Carlo Exercises	190
9 10	Baye	sian Inference for Poisson	193
	10.1	Some Prior Distributions for Poisson	194
	10.2	Inference for Poisson Parameter	200
		Exercises	207
		Computer Exercises	208
11	Baye	sian Inference for Normal Mean	211
	11.1	Bayes' Theorem for Normal Mean with a Discrete Prior	211
	11.2	Bayes' Theorem for Normal Mean with a Continuous Prior	218
	11.3	Choosing Your Normal Prior	222
	11.4	Bayesian Credible Interval for Normal Mean	224
	11.5	Predictive Density for Next Observation	227
		Exercises	230
		Computer Exercises	232

12	Com	paring Bayesian and Frequentist Inferences for Mean	237
	12.1	Comparing Frequentist and Bayesian Point Estimators	238
	12.2	Comparing Confidence and Credible Intervals for Mean	241
	12.3	Testing a One-Sided Hypothesis about a Normal Mean	243
	12.4	Testing a Two-Sided Hypothesis about a Normal Mean	247
		Exercises	251
13	Baye	sian Inference for Difference Between Means	255
	13.1	Independent Random Samples from Two Normal	
		Distributions	256
	13.2	Case 1: Equal Variances	257
	13.3	Case 2: Unequal Variances	262
	13.4	Bayesian Inference for Difference Between Two Proportions	
		Using Normal Approximation	265
	13.5	Normal Random Samples from Paired Experiments	266
		Exercises	272
14	Baye	sian Inference for Simple Linear Regression	283
	14.1	Least Squares Regression	284
	14.2	Exponential Growth Model	288
	14.3	Simple Linear Regression Assumptions	290
	14.4	Bayes' Theorem for the Regression Model	292
	14.5	Predictive Distribution for Future Observation	298
		Exercises	303
		Computer Exercises	312
15	Baye	sian Inference for Standard Deviation	315
	15.1	Bayes' Theorem for Normal Variance with a Continuous	
		Prior	316
	15.2	Some Specific Prior Distributions and the Resulting	
		Posteriors	318
	15.3	Bayesian Inference for Normal Standard Deviation	326
		Exercises	332
		Computer Exercises	335
16	Robu	st Bayesian Methods	337
	16.1	Effect of Misspecified Prior	338
	16.2	Bayes' Theorem with Mixture Priors	340

		Exercises	349
		Computer Exercises	351
17	Baye	sian Inference for Normal	
	with	Unknown Mean and Variance	355
	17.1	The Joint Likelihood Function	358
	17.2	Finding the Posterior when Independent Jeffreys' Priors	
		for $\mu$ and $\sigma^2$ Are Used	359
	17.3	Finding the Posterior when a Joint Conjugate Prior for $\mu$	
		and $\sigma^2$ Is Used	361
	17.4	Difference Between Normal Means with Equal Unknown	
		Variance	367
	17.5	Difference Between Normal Means with Unequal Unknown	
		Variances	377
		Computer Exercises	383
		Appendix: Proof that the Exact Marginal Posterior	205
		Distribution of $\mu$ is Student's t	385
18	Baye	sian Inference for Multivariate Normal Mean Vector	393
	18.1	Bivariate Normal Density	394
	18.2	Multivariate Normal Distribution	397
	18.3	The Posterior Distribution of the Multivariate Normal	
		Mean Vector when Covariance Matrix Is Known	398
	18.4	Credible Region for Multivariate Normal Mean Vector	
		when Covariance Matrix Is Known	400
	18.5	Multivariate Normal Distribution with Unknown	
		Covariance Matrix	402
		Computer Exercises	406
19	Baye	sian Inference for the Multiple Linear Regression Model	411
	19.1	Least Squares Regression for Multiple Linear Regression	
	10.1	Model	412
	19.2	Assumptions of Normal Multiple Linear Regression Model	414
	19.3	Bayes' Theorem for Normal Multiple Linear Regression	
		Model	415
	19.4	Inference in the Multivariate Normal Linear Regression	
		Model	419
	19.5	The Predictive Distribution for a Future Observation	425

xii	CONTENTS	
хп	CONTENTS	

	Com	puter Exercises	428	
20	Computational Bayesian Statistics Including Markov Chain Monte Carlo			
	20.1 Direct Methods for Sampling from the Posterior			
	20.2 Samp	pling–Importance–Resampling	450	
	20.3 Mark	ov Chain Monte Carlo Methods	454	
	20.4 Slice	Sampling	470	
	20.5 Infer	ence from a Posterior Random Sample	473	
	20.6 When	re to Next?	475	
Α	Introductio	on to Calculus	477	
В	Use of Sta	tistical Tables	497	
С	Using the	Included Minitab Macros	523	
D	Using the Included R Functions		543	
Е	Answers to Selected Exercises		565	
References			591	
Index 5			595	

## PREFACE

Our original goal for this book was to introduce Bayesian statistics at the earliest possible stage to students with a reasonable mathematical background. This entailed coverage of a similar range of topics as an introductory statistics text, but from a Bayesian perspective. The emphasis is on statistical inference. We wanted to show how Bayesian methods can be used for inference and how they compare favorably with the frequentist alternatives. This book is meant to be a good place to start the study of Bayesian statistics. From the many positive comments we have received from many users, we think the book succeeded in its goal. A course based on this goal would include Chapters 1–14.

Our feedback also showed that many users were taking up the book at a more intermediate level instead of the introductory level original envisaged. The topics covered in Chapters 2 and 3 would be old hat for these users, so we would have to include some more advanced material to cater for the needs of that group. The second edition aimed to meet this new goal as well as the original goal. We included more models, mainly with a single parameter. Nuisance parameters were dealt with using approximations. A course based on this goal would include Chapters 4–16.

#### **Changes in the Third Edition**

Later feedback showed that some readers with stronger mathematical and statistical background wanted the text to include more details on how to deal with multi-parameter models. The third edition contains four new chapters to satisfy this additional goal, along with some minor rewriting of the existing chapters. Chapter 17 covers Bayesian inference for Normal observations where we do not know either the mean or the variance. This chapter extends the ideas in Chapter 11, and also discusses the two sample case, which in turn allows the reader to consider inference on the difference between two means. Chapter 18 introduces the *Multivariate Normal* distribution, which we need in order to discuss multiple linear regression in Chapter 19. Finally, Chapter 20 takes the user beyond the kind of conjugate analysis is considered in most of the book, and into the realm of computational Bayesian inference. The covered topics in Chapter 20 have an intentional light touch, but still give the user valuable information and skills that will allow them to deal with different problems. We have included some new exercises and new computer exercises which use new Minitab macros and R-functions. The Minitab macros can be downloaded from the book website: http://introbayes.ac.nz. The new R functions have been incorporated in a new and improved version of the R package Bolstad, which can either be downloaded from a CRAN mirror or installed directly in R using the internet. Instructions on the use and installation of the Minitab macros and the Bolstad package in R are given in Appendices C and D respectively. Both of these appendices have been rewritten to accommodate changes in R and Minitab that have occurred since the second edition.

#### **Our Perspective on Bayesian Statistics**

A book can be characterized as much by what is left out as by what is included. This book is our attempt to show a coherent view of Bayesian statistics as a good way to do statistical inference. Details that are outside the scope of the text are included in footnotes. Here are some of our reasons behind our choice of the topics we either included or excluded.

In particular, we did not mention decision theory or loss functions when discussing Bayesian statistics. In many books, Bayesian statistics gets compartmentalized into decision theory while inference is presented in the frequentist manner. While decision theory is a very interesting topic in its own right, we want to present the case for Bayesian statistical inference, and did not want to get side-tracked.

We think that in order to get full benefit of Bayesian statistics, one really has to consider all priors subjective. They are either (1) a summary of what you believe or (2) a summary of all you allow yourself to believe initially. We consider the subjective prior as the relative weights given to each possible parameter value, before looking at the data. Even if we use a flat prior to give all possible values equal prior weight, it is subjective since we chose it. In any case, it gives all values equal weight only in that parameterization, so it can be considered "objective" only in that parameterization. In this book we do not wish to dwell on the problems associated with trying to be objective in Bayesian statistics. We explain why universal objectivity is not possible (in a footnote since we do not want to distract the reader). We want to leave him/her with the "relative weight" idea of the prior in the parameterization in which they have the problem in.

In the first edition we did not mention Jeffreys' prior explicitly, although the  $beta(\frac{1}{2},\frac{1}{2})$  prior for binomial and flat prior for normal mean are in fact the Jeffreys' prior for those respective observation distributions. In the second edition we do mention Jeffreys' prior for binomial, Poisson, normal mean, and normal standard deviation. In third edition we mention the independent Jeffreys priors for *normal* mean and standard deviation. In particular, we don't want to get the reader involved with the problems about Jeffreys' prior, such as for mean and variance together, as opposed to independent Jeffreys' priors, or the Jeffreys' prior violating the likelihood principal. These are beyond the level we wish to go. We just want the reader to note the Jeffreys' prior in these cases as possible priors, the relative weights they give, when they may be appropriate, and how to use them. Mathematically, all parameterizations are equally valid; however, usually only the main one is very meaningful. We want the reader to focus on relative weights for their parameterization as the prior. It should be (a) a summary of their prior belief (conjugate prior matching their prior beliefs about moments or median), (b) flat (hence objective) for their parameterization, or (c) some other form that gives reasonable weight over the whole range of possible values. The posteriors will be similar for all priors that have reasonable weight over the whole range of possible values.

The Bayesian inference on the standard deviation of the normal was done where the mean is considered a known parameter. The conjugate prior for the variance is the *inverse chi-squared* distribution. Our intuition is about the standard deviation, yet we are doing Bayes' theorem on the variance. This required introducing the change of variable formula for the prior density.

In the second edition we considered the mean as known. This avoided the mathematically more advanced case where both mean and standard deviation are unknown. In the third edition we now cover this topic in Chapter 17. In earlier editions the *Student's* t is presented as the required adjustment to credible intervals for the mean when the variance is estimated from the data. In the third edition we show in Chapter 17 that in fact this would be the result when the joint posterior found, and the variance marginalized out. Chapter 17 also covers inference on the difference in two means. This problem is made substantially harder when one relaxes the assumption that both populations have the same variance. Chapter 17 derives the Bayesian solution to the well-known Behrens-Fisher problem for the difference in two population means with unequal population variances. The function bayes.t.test in the R

package for this book actually gives the user a numerical solution using Gibbs sampling. Gibbs sampling is covered in Chapter 20 of this new edition.

#### Acknowledgments

WMB would like to thank all the readers who have sent him comments and pointed out misprints in the first and second editions. These have been corrected. WMB would like to thank Cathy Akritas and Gonzalo Ovalles at Minitab for help in improving his Minitab macros. WMB and JMC would like to thank Jon Gurstelle, Steve Quigley, Sari Friedman, Allison McGinniss, and the team at John Wiley & Sons for their support.

Finally, last but not least, WMB wishes to thank his wife Sylvie for her constant love and support.

WILLIAM M. "BILL' BOLSTAD

Hamilton, New Zealand

JAMES M. CURRAN

Auckland, New Zealand

## INTRODUCTION TO STATISTICAL SCIENCE

Statistics is the science that relates data to specific questions of interest. This includes devising methods to gather data relevant to the question, methods to summarize and display the data to shed light on the question, and methods that enable us to draw answers to the question that are supported by the data. Data almost always contain uncertainty. This uncertainty may arise from selection of the items to be measured, or it may arise from variability of the measurement process. Drawing general conclusions from data is the basis for increasing knowledge about the world, and is the basis for all rational scientific inquiry. *Statistical inference* gives us methods and tools for doing this despite the uncertainty in the data. The methods used for analysis depend on the way the data were gathered. It is vitally important that there is a probability model explaining how the uncertainty gets into the data.

#### Showing a Causal Relationship from Data

Suppose we have observed two variables X and Y. Variable X appears to have an association with variable Y. If high values of X occur with high values of variable Y and low values of X occur with low values of Y, then we say the



Figure 1.1 Association between two variables.



Figure 1.2 Association due to causal relationship.

association is positive. On the other hand, the association could be negative in which high values of variable X occur in with low values of variable Y. Figure 1.1 shows a schematic diagram where the association is indicated by the dashed curve connecting X and Y. The unshaded area indicates that Xand Y are observed variables. The shaded area indicates that there may be additional variables that have not been observed.

We would like to determine why the two variables are associated. There are several possible explanations. The association might be a causal one. For example, X might be the cause of Y. This is shown in Figure 1.2, where the causal relationship is indicated by the arrow from X to Y.

On the other hand, there could be an unidentified third variable Z that has a causal effect on both X and Y. They are not related in a direct causal relationship. The association between them is due to the effect of Z. Z is called a *lurking* variable, since it is hiding in the background and it affects the data. This is shown in Figure 1.3.



**Figure 1.3** Association due to lurking variable.



Figure 1.4 Confounded causal and lurking variable effects.

It is possible that both a causal effect and a lurking variable may both be contributing to the association. This is shown in Figure 1.4. We say that the causal effect and the effect of the lurking variable are *confounded*. This means that both effects are included in the association.

Our first goal is to determine which of the possible reasons for the association holds. If we conclude that it is due to a causal effect, then our next goal is to determine the size of the effect. If we conclude that the association is due to causal effect confounded with the effect of a lurking variable, then our next goal becomes determining the sizes of both the effects.

#### 1.1 The Scientific Method: A Process for Learning

In the Middle Ages, science was deduced from principles set down many centuries earlier by authorities such as Aristotle. The idea that scientific theories should be tested against real world data revolutionized thinking. This way of thinking known as the scientific method sparked the Renaissance.

#### 4 INTRODUCTION TO STATISTICAL SCIENCE

The scientific method rests on the following premises:

- A scientific hypothesis can never be shown to be absolutely true.
- However, it must potentially be disprovable.
- It is a useful model until it is established that it is not true.
- Always go for the simplest hypothesis, unless it can be shown to be false.

This last principle, elaborated by William of Ockham in the 13<sup>th</sup> century, is now known as *Ockham's razor* and is firmly embedded in science. It keeps science from developing fanciful overly elaborate theories. Thus the scientific method directs us through an improving sequence of models, as previous ones get falsified. The scientific method generally follows the following procedure:

- 1. Ask a question or pose a problem in terms of the current scientific hypothesis.
- 2. Gather all the relevant information that is currently available. This includes the current knowledge about parameters of the model.
- 3. Design an investigation or experiment that addresses the question from step 1. The predicted outcome of the experiment should be one thing if the current hypothesis is true, and something else if the hypothesis is false.
- 4. Gather data from the experiment.
- 5. Draw conclusions given the experimental results. Revise the knowledge about the parameters to take the current results into account.

The scientific method searches for cause-and-effect relationships between an experimental variable and an outcome variable. In other words, how changing the experimental variable results in a change to the outcome variable. Scientific modeling develops mathematical models of these relationships. Both of them need to isolate the experiment from outside factors that could affect the experimental results. All outside factors that can be identified as possibly affecting the results must be controlled. It is no coincidence that the earliest successes for the method were in physics and chemistry where the few outside factors could be identified and controlled. Thus there were no *lurking* variables. All other relevant variables could be identified and could then be physically controlled by being held constant. That way they would not affect results of the experiment, and the effect of the experimental variable on the outcome variable could be determined. In biology, medicine, engineering, technology, and the social sciences it is not that easy to identify the relevant factors that must be controlled. In those fields a different way to control outside factors is needed, because they cannot be identified beforehand and physically controlled.

### 1.2 The Role of Statistics in the Scientific Method

Statistical methods of inference can be used when there is *random* variability in the data. The probability model for the data is justified by the design of the investigation or experiment. This can extend the scientific method into situations where the relevant outside factors cannot even be identified. Since we cannot identify these outside factors, we cannot control them directly. The lack of direct control means the outside factors will be affecting the data. There is a danger that the wrong conclusions could be drawn from the experiment due to these uncontrolled outside factors.

The important statistical idea of *randomization* has been developed to deal with this possibility. The unidentified outside factors can be "averaged out" by randomly assigning each unit to either treatment or control group. This contributes variability to the data. Statistical conclusions always have some uncertainty or error due to variability in the data. We can develop a probability model of the data variability based on the randomization used. Randomization not only reduces this uncertainty due to outside factors, it also allows us to measure the amount of uncertainty that remains using the probability model. Randomization lets us control the outside factors statistically, by averaging out their effects.

Underlying this is the idea of a statistical *population*, consisting of all possible values of the observations that could be made. The data consists of observations taken from a *sample* of the population. For valid inferences about the population *parameters* from the sample *statistics*, the sample must be "representative" of the population. Amazingly, choosing the sample randomly is the most effective way to get representative samples!

#### **1.3** Main Approaches to Statistics

There are two main philosophical approaches to statistics. The first is often referred to as the *frequentist* approach. Sometimes it is called the *classical* approach. Procedures are developed by looking at how they perform over all possible random samples. The probabilities do not relate to the particular random sample that was obtained. In many ways this indirect method places the "cart before the horse."

The alternative approach that we take in this book is the *Bayesian* approach. It applies the laws of probability directly to the problem. This offers many fundamental advantages over the more commonly used frequentist approach. We will show these advantages over the course of the book.

#### Frequentist Approach to Statistics

Most introductory statistics books take the frequentist approach to statistics, which is based on the following ideas:

- Parameters, the numerical characteristics of the population, are fixed but unknown constants.
- Probabilities are always interpreted as long-run relative frequency.
- Statistical procedures are judged by how well they perform in the long run over an infinite number of hypothetical repetitions of the experiment.

Probability statements are only allowed for random quantities. The unknown parameters are fixed, not random, so probability statements cannot be made about their value. Instead, a sample is drawn from the population, and a sample statistic is calculated. The probability distribution of the statistic over all possible random samples from the population is determined and is known as the *sampling distribution* of the statistic. A parameter of the population will also be a parameter of the sampling distribution. The probability statement that can be made about the statistic based on its sampling distribution is converted to a *confidence* statement about the parameter. The confidence is based on the average behavior of the procedure over all possible samples.

### **Bayesian Approach to Statistics**

The Reverend Thomas Bayes first discovered the theorem that now bears his name. It was written up in a paper An Essay Towards Solving a Problem in the Doctrine of Chances. This paper was found after his death by his friend Richard Price, who had it published posthumously in the Philosophical Transactions of the Royal Society in 1763 (1763). Bayes showed how inverse probability could be used to calculate probability of antecedent events from the occurrence of the consequent event. His methods were adopted by Laplace and other scientists in the 19<sup>th</sup> century, but had largely fallen from favor by the early 20<sup>th</sup> century. By the middle of the 20<sup>th</sup> century, interest in Bayesian methods had been renewed by de Finetti, Jeffreys, Savage, and Lindley, among others. They developed a complete method of statistical inference based on Bayes' theorem.

This book introduces the Bayesian approach to statistics. The ideas that form the basis of the this approach are:

- Since we are uncertain about the true value of the parameters, we will consider them to be random variables.
- The rules of probability are used directly to make inferences about the parameters.
- Probability statements about parameters must be interpreted as "degree of belief." The *prior distribution* must be subjective. Each person can have his/her own prior, which contains the relative weights that person

gives to every possible parameter value. It measures how "plausible" the person considers each parameter value to be before observing the data.

• We revise our beliefs about parameters after getting the data by using Bayes' theorem. This gives our *posterior distribution* which gives the relative weights we give to each parameter value after analyzing the data. The posterior distribution comes from two sources: the prior distribution and the observed data.

This has a number of advantages over the conventional frequentist approach. Bayes' theorem is the only consistent way to modify our beliefs about the parameters given the data that actually occurred. This means that the inference is based on the actual occurring data, not all possible data sets that might have occurred but did not! Allowing the parameter to be a random variable lets us make probability statements about it, posterior to the data. This contrasts with the conventional approach where inference probabilities are based on all possible data sets that could have occurred for the fixed parameter value. Given the actual data, there is nothing random left with a fixed parameter value, so one can only make *confidence* statements, based on what could have occurred. Bayesian statistics also has a general way of dealing with a *nuisance parameter*. A nuisance parameter is one which we do not want to make inference about, but we do not want them to interfere with the inferences we are making about the main parameters. Frequentist statistics does not have a general procedure for dealing with them. Bayesian statistics is predictive, unlike conventional frequentist statistics. This means that we can easily find the conditional probability distribution of the next observation given the sample data.

#### Monte Carlo Studies

In frequentist statistics, the parameter is considered a fixed, but unknown, constant. A statistical procedure such as a particular estimator for the parameter cannot be judged from the value it gives. The parameter is unknown, so we can not know the value the estimator should be giving. If we knew the value of the parameter, we would not be using an estimator.

Instead, statistical procedures are evaluated by looking how they perform in the long run over all possible samples of data, for fixed parameter values over some range. For instance, we fix the parameter at some value. The estimator depends on the random sample, so it is considered a random variable having a probability distribution. This distribution is called the *sampling distribution* of the estimator, since its probability distribution comes from taking all possible random samples. Then we look at how the estimator is distributed around the parameter value. This is called sample space averaging. Essentially it compares the performance of procedures before we take any data.

Bayesian procedures consider the parameter to be a random variable, and its posterior distribution is conditional on the sample data that actually occurred, not all those samples that were possible but did not occur. However, *before* the experiment, we might want to know how well the Bayesian procedure works at some specific parameter values in the range.

To evaluate the Bayesian procedure using sample space averaging, we have to consider the parameter to be both a random variable and a fixed but unknown value at the same time. We can get past the apparent contradiction in the nature of the parameter because the probability distribution we put on the parameter measures our uncertainty about the true value. It shows the relative belief weights we give to the possible values of the unknown parameter! After looking at the data, our belief distribution over the parameter values has changed. This way we can think of the parameter as a fixed, but unknown, value at the same time as we think of it being a random variable. This allows us to evaluate the Bayesian procedure using sample space averaging. This is called *pre-posterior* analysis because it can be done before we obtain the data.

In Chapter 4, we will find out that the laws of probability are the best way to model uncertainty. Because of this, Bayesian procedures will be optimal in the post-data setting, given the data that actually occurred. In Chapters 9 and 11, we will see that Bayesian procedures perform very well in the pre-data setting when evaluated using *pre-posterior* analysis. In fact, it is often the case that Bayesian procedures outperform the usual frequentist procedures even in the pre-data setting.

Monte Carlo studies are a useful way to perform sample space averaging. We draw a large number of samples randomly using the computer and calculate the statistic (frequentist or Bayesian) for each sample. The empirical distribution of the statistic (over the large number of random samples) approximates its sampling distribution (over all possible random samples). We can calculate statistics such as mean and standard deviation on this Monte Carlo sample to approximate the mean and standard deviation of the sampling distribution. Some small-scale Monte Carlo studies are included as exercises.

#### 1.4 Purpose and Organization of This Text

A very large proportion of undergraduates are required to take a service course in statistics. Almost all of these courses are based on frequentist ideas. Most of them do not even mention Bayesian ideas. As a statistician, I know that Bayesian methods have great theoretical advantages. I think we should be introducing our best students to Bayesian ideas, from the beginning. There are not many introductory statistics text books based on the Bayesian ideas. Some other texts include Berry (1996), Press (1989), and Lee (1989).

This book aims to introduce students with a good mathematics background to Bayesian statistics. It covers the same topics as a standard introductory statistics text, only from a Bayesian perspective. Students need reasonable algebra skills to follow this book. Bayesian statistics uses the rules of probability, so competence in manipulating mathematical formulas is required. Students will find that general knowledge of calculus is helpful in reading this book. Specifically they need to know that area under a curve is found by integrating, and that a maximum or minimum of a continuous differentiable function is found where the derivative of the function equals zero. However, the actual calculus used is minimal. The book is self-contained with a calculus appendix that students can refer to.

Chapter 2 introduces some fundamental principles of scientific data gathering to control the effects of unidentified factors. These include the need for drawing samples randomly, along with some random sampling techniques. The reason why there is a difference between the conclusions we can draw from data arising from an observational study and from data arising from a randomized experiment is shown. Completely randomized designs and randomized block designs are discussed.

Chapter 3 covers elementary methods for graphically displaying and summarizing data. Often a good data display is all that is necessary. The principles of designing displays that are true to the data are emphasized.

Chapter 4 shows the difference between deduction and induction. Plausible reasoning is shown to be an extension of logic where there is uncertainty. It turns out that plausible reasoning must follow the same rules as probability. The axioms of probability are introduced and the rules of probability, including conditional probability and Bayes' theorem are developed.

Chapter 5 covers discrete random variables, including joint and marginal discrete random variables. The *binomial*, *hypergeometric*, and *Poisson* distributions are introduced, and the situations where they arise are characterized.

Chapter 6 covers Bayes' theorem for discrete random variables using a table. We see that two important consequences of the method are that multiplying the prior by a constant, or that multiplying the likelihood by a constant do not affect the resulting posterior distribution. This gives us the "proportional form" of Bayes' theorem. We show that we get the same results when we analyze the observations sequentially using the posterior after the previous observation as the prior for the next observation, as when we analyze the observations all at once using the joint likelihood and the original prior. We demonstrate Bayes' theorem for binomial observations with a discrete prior and for Poisson observations with a discrete prior.

Chapter 7 covers continuous random variables, including joint, marginal, and conditional random variables. The *beta*, *gamma*, and *normal* distributions are introduced in this chapter.

Chapter 8 covers Bayes' theorem for the population proportion (*binomial*) with a continuous prior. We show how to find the posterior distribution of the population proportion using either a *uniform* prior or a *beta* prior. We explain how to choose a suitable prior. We look at ways of summarizing the posterior distribution.

Chapter 9 compares the Bayesian inferences with the frequentist inferences. We show that the Bayesian estimator (posterior mean using a uniform prior) has better performance than the frequentist estimator (sample proportion) in terms of mean squared error over most of the range of possible values. This kind of frequentist analysis is useful before we perform our Bayesian analysis. We see the Bayesian credible interval has a much more useful interpretation than the frequentist confidence interval for the population proportion. Onesided and two-sided hypothesis tests using Bayesian methods are introduced.

Chapter 10 covers Bayes' theorem for the *Poisson* observations with a continuous prior. The prior distributions used include the *positive uniform*, the *Jeffreys' prior*, and the *gamma* prior. Bayesian inference for the *Poisson* parameter using the resulting posterior include Bayesian credible intervals and two-sided tests of hypothesis, as well as one-sided tests of hypothesis.

Chapter 11 covers Bayes' theorem for the mean of a *normal* distribution with known variance. We show how to choose a *normal* prior. We discuss dealing with nuisance parameters by marginalization. The predictive density of the next observation is found by considering the population mean a nuisance parameter and marginalizing it out.

Chapter 12 compares Bayesian inferences with the frequentist inferences for the mean of a normal distribution. These comparisons include point and interval estimation and involve hypothesis tests including both the one-sided and the two-sided cases.

Chapter 13 shows how to perform Bayesian inferences for the difference between normal means and how to perform Bayesian inferences for the difference between proportions using the normal approximation.

Chapter 14 introduces the simple linear regression model and shows how to perform Bayesian inferences on the slope of the model. The predictive distribution of the next observation is found by considering both the slope and intercept to be nuisance parameters and marginalizing them out.

Chapter 15 introduces Bayesian inference for the standard deviation  $\sigma$ , when we have a random sample of *normal* observations with known mean  $\mu$ . This chapter is at a somewhat higher level than the previous chapters and requires the use of the change-of-variable formula for densities. Priors used include *positive uniform for standard deviation*, *positive uniform for variance*, *Jeffreys' prior*, and the *inverse chi-squared* prior. We discuss how to choose an *inverse chi-squared* prior that matches our prior belief about the median. Bayesian inferences from the resulting posterior include point estimates, credible intervals, and hypothesis tests including both the one-sided and two-sided cases.

Chapter 16 shows how we can make Bayesian inference robust against a misspecified prior by using a mixture prior and marginalizing out the mixture parameter. This chapter is also at a somewhat higher level than the others, but it shows how one of the main dangers of Bayesian analysis can be avoided.

Chapter 17 returns to the problem we discussed in Chapter 11 — that is, of making inferences about the mean of a normal distribution. In this chapter, however, we explicitly model the unknown population standard deviation and show how the approximations we suggested in Chapter 11 are exactly true.

We also deal with the two sample cases so that inference can be performed on the difference between two means.

Chapter 18 introduces the multivariate normal distribution and extends the theory from Chapters 11 and 17 to the multivariate case. The multivariate normal distribution is essential for the discussion of linear models and, in particular, multiple regression.

Chapter 19 extends the material from 14 on simple linear regression to the more familiar multiple regression setting. The methodology for making inference about the usefulness of explanatory variables in predicting the response is given, and the posterior predictive distribution for a new observation is derived.

Chapter 20 provides a brief introduction to modern computational Bayesian statistics. Computational Bayesian statistics relies heavily on being able to efficiently sample from potentially complex distributions. This chapter gives an introduction to a number of techniques that are used. Readers might be slightly disappointed that we did not cover popular computer programs such as BUGS and JAGS, which have very efficient general implementations of many computational Bayesian methods and tie in well to R. We felt that these topics require almost an entire book in their own right, and as such we could not do justice to them in such a short space.

## Main Points

- An association between two variables does not mean that one causes the other. It may be due to a causal relationship, it may be due to the effect of a third (lurking) variable on both the other variables, or it may be due to a combination of a causal relationship and the effect of a lurking variable.
- Scientific method is a method for searching for cause-and-effect relationships and measuring their strength. It uses controlled experiments, where outside factors that may affect the measurements are controlled. This isolates the relationship between the two variables from the outside factors, so the relationship can be determined.
- Statistical methods extend the scientific method to cases where the outside factors are not identified and hence cannot be controlled. The principle of *randomization* is used to statistically control these unidentified outside factors by averaging out their effects. This contributes to *variability* in the data.
- We can use the probability model (based on the randomization method) to measure the uncertainty.
- The frequentist approach to statistics considers the parameter to be a fixed but unknown constant. The only kind of probability allowed is long-

run relative frequency. These probabilities are only for observations and sample statistics, given the unknown parameters. Statistical procedures are judged by how they perform in an infinite number of hypothetical repetitions of the experiment.

- The Bayesian approach to statistics allows the parameter to be considered a random variable. Probabilities can be calculated for parameters as well as observations and sample statistics. Probabilities calculated for parameters are interpreted as "degree of belief" and must be subjective. The rules of probability are used to revise our beliefs about the parameters, given the data.
- A frequentist estimator is evaluated by looking at its sampling distribution for a fixed parameter value and seeing how it is distributed over all possible repetitions of the experiment.
- If we look at the sampling distribution of a Bayesian estimator for a fixed parameter value, it is called pre-posterior analysis since it can be done prior to taking the data.
- A Monte Carlo study is where we perform the experiment a large number of times and calculate the statistic for each experiment. We use the empirical distribution of the statistic over all the samples we took in our study instead of its sampling distribution over all possible repetitions.