ACTEX ACADEMIC SERIES

Introduction to Credibility Theory

4th Edition

THOMAS N. HERZOG, PhD, ASA

ACTEX Academic Series

Introduction to Credibility Theory

4th Edition

Thomas N. Herzog, PhD, ASA

ACTEX Publications, Inc.
Winsted, CT

To Tracy, Steve, and Evelyn

Requests for permission should be addressed to
ACTEX Learning
PO Box 715
New Hartford CT 06057

Cover design: Christine Phelps

ISBN: 978-1-62542-905-6

PREFACE

This text is intended to introduce the reader to a wide variety of credibility models and, in so doing, trace the historical development of the subject. The Bayesian approach to statistics is emphasized, revealing the author's personal preference. The reader should be able to use this work as a foundation for understanding more sophisticated treatments in other works. For example, by seeing how various formulas are derived in the Bayesian paradigm, the reader should be able to understand other works describing the Bayesian approach to credibility. Another goal is to present the key assumptions underlying the various credibility models and to discuss the advantages and disadvantages of the various approaches.

This work is intended to be largely self-contained. Although numerous references to the technical literature are provided, few are necessary for an understanding of the material discussed here. Rather, they are provided for those who would like to consult original sources and/or obtain some insight into the more advanced topics omitted from this introductory work. A large number of exercises are provided to help reinforce understanding of the material. Most of these have been taken from past examinations of the Casualty Actuarial Society. Complete solutions to all of the text exercises are available in a companion solutions manual.

The emphasis in the first ten chapters of this introductory text is on basic statistical concepts.

In Chapter 1, we (a) discuss two major statistical paradigms, (b) offer a glimpse into the nature of credibility, (c) introduce a simple practical problem later solved in Chapter 6 using credibility procedures, and (d) present a brief review of the key historical developments in credibility theory and its application to practical insurance problems.

In Chapter 2, we review the basic concepts of Bayesian analysis, and in Chapter 3 we discuss statistical loss functions. In Chapter 4, we use an example originally employed by Hewitt [1970] to illustrate the use of Bayesian concepts in the insurance ratemaking process. The key ideas are the use of the predictive distribution of aggregate claim amounts and the use of the (Bayesian) conditional mean to estimate pure premium amounts.

In Chapter 5, we describe the limited fluctuation credibility model. This is primarily of historical interest, because it is not in wide use today.

In Chapter 6, we present the development of an alternative credibility model proposed by Bühlmann [1967] as well as a special case of a more general model proposed by Bühlmann and Straub [1972]. The general Bühlmann-Straub model is presented in Chapter 7.

In Chapter 8, we discuss an important general result of Ericson [1970]. It turns out that some specific results described in Mayerson [1964] are simply special cases of Ericson's more general result. We also present an example which shows that the Bühlmann estimate does not always equal the corresponding Bayesian estimate. In Chapter 9, we use the statistical machinery developed to describe Ericson's result to construct the predictive distribution of a more realistic two-stage model. Here the number of claims is assumed to follow a Poisson distribution and the claim amounts are based on an exponential distribution.

In Chapter 10, we show that Bühlmann's model produces least squares linear approximations to the Bayesian estimate of the pure premium.

In the first ten chapters, we do not discuss important practical issues such as how to apply these procedures in dealing with issues likely to be encountered in real-life situations. Moreover, we do not attempt to discuss more sophisticated theoretical concepts such as multivariate extensions of the results presented here. These are all left for a more advanced treatment in later chapters and elsewhere.

The three prior editions of this text have been in use since 1994. Our goal over all of these years has been to make this text of practical use to the working actuary/actuarial student. To further this goal, we have added three new chapters to this fourth edition. Each of Chapters 11 through 15 now deals in depth with a practical application of the concepts developed earlier in the text. Chapter 11 discusses a Bayesian procedure for comparing two binomial proportions. Many researchers, including us, feel that this is far superior to the frequentist scheme of deciding whether or not to reject the null hypothesis that the two proportions are equal. Chapter 12 describes a procedure suggested by Fuhrer [1988] that has application to health insurance. Chapter 13 summarizes work completed by Rosenberg and Farrell [2008] that describes a scheme for predicting the frequency and severity of hospitalization cost for a group of young children suffering from cystic fibrosis. Chapters 14 and 15 continue from earlier editions. In Chapter 14, we use the concept of conjugate prior distributions to estimate probabilities arising from a data quality problem. In Chapter 15, we present an application of empirical Bayesian procedures to a problem in automobile insurance ratemaking. This is the application and solution proposed by Morris and van Slyke [1979].

The other major change in the fourth edition is in Chapter 1 where we have expanded the historical discussion surrounding the work of Bayes himself as well as Laplace.

We assume here that the reader has a working knowledge of (1) the integration techniques normally taught during the second semester of a university course in calculus, (2) basic probability and statistics as taught during a two-semester university course having a calculus prerequisite, and (3) matrix manipulation and multiplication techniques. In particular regard to integral calculus, we assume that the reader can perform integration by parts and integration by change of variable (as done in Section 9.4). We also note that in a number places we have changed either the order of integration or summation. In general such manipulation requires the verification of one or more conditions to ensure that the sums actually converge. Because we are dealing here with probabilities that sum to one, we are never in danger of diverging to infinity. Hence, we will omit the verification step in this work.

For many years, the topic of credibility theory has been included in the preliminary exams jointly sponsored by the Casualty Actuarial Society and the Society of Actuaries. All of the topics in that collection as well as the general learning objectives and the sample exam questions are thoroughly covered by this text. It has thus been an officially approved reference for this topic for most of those years.

Many people have contributed to the completion of this project. In particular I am grateful to the late Professor James C. Hickman, FSA, ACAS, University of Wisconsin, and Gary G. Venter, FCAS, ASA, CERA, MAAA for their generous assistance with an initial version of this work. Jim taught me a lot about Bayesian statistics and Gary introduced me to the Bühlmann and limited fluctuation approaches. Gary also read this and earlier versions and suggested a number of the figures that appear in Chapters 6 and 10.

I also want to thank Professor Donald B. Rubin of Harvard University for introducing me to the Bayesian paradigm. One of my other early teachers was Professor Dennis V. Lindley, at George Washington University. I would also like to express my appreciation to Professor James C. Owings, Jr., my dissertation advisor in the Department of Mathematics at the University of Maryland.

In addition to Gary Venter, I am also indebted to the other members of the initial editorial review team for their suggestions that improved the exposition and notation of this work. This group included Professor Elias Shiu, ASA, University of Iowa, Professor Samuel H. Cox, FSA, University of Manitoba, Professor Larry Santoro, College of Insurance, Professor Michael R. Powers, Temple University, and David F. Mohrman, FCAS, and Stephen W. Philbrick, FCAS, both of Towers Watson. Professor Shiu supplied me with the clever proofs presented in Sections 3.2 and 5.2.

The fourth edition was also reviewed by Steven Craighead, Margie Rosenberg, Dennis Tolley, Chuck Fuhrer, Bob Cumming, Joan Barrett, and Vincent Kane. The author is grateful for their generous and thoughtful assistance with this text.

Finally, I would like to thank the staff of ACTEX Publications, Inc, especially Marilyn Baleshiski and Gail Hall for their assistance with the publication of this text.

Reston, Virginia — Thomas N. Herzog, Ph.D., ASA
June 2010

CONTENTS

CHAPTER 3 LOSS FUNCTIONS 31

CHAPTER 4 DISCRETE FREQUENCY-SEVERITY INSURANCE MODEL UNDER INDEPENDENCE 43

CHAPTER 5 LIMITED FLUCTUATION CREDIBILITY APPROACH 53

CHAPTER 6 BÜHLMANN'S APPROACH 71

CHAPTER 7 BÜHLMANN-STRAUB MODEL 113

CHAPTER 8 CREDIBILITY AND BAYESIAN INFERENCE 137

CHAPTER 9 FREQUENCY-SEVERITY INSURANCE MODEL WITH CONTINUOUS SEVERITY COMPONENT 163

CHAPTER 10 CREDIBILITY AND LEAST SQUARES 183

CHAPTER 11 CASE STUDY: CLINICAL TRIALS – AN APPLICATION OF BAYESIAN INFERENCE ABOUT BINOMIAL PROPORTIONS 193

CHAPTER 12 CASE STUDY: APPLICATION OF CREDIBILITY TO GROUP HEALTH INSURANCE 205

CHAPTER 13 A CASE STUDY IN BAYESIAN ANALYSIS: THE IMPACT OF NEWBORN SCREENING ON CYSTIC FIBROSIS HOSPITALIZATION COSTS 221

CHAPTER 14 A CASE STUDY: THE MISCODED MORTGAGE RECORDS 235

APPENDIX B: COMPARING CREDIBILITY ESTIMATES OF HEALTH INSURANCE CLAIMS COSTS 293

CHAPTER 1

INTRODUCTION AND HISTORY

1.1 INTRODUCTION

According to Rodermund [1989, page 3], "the concept of credibility has been the casualty actuaries' most important and enduring contribution to casualty actuarial science."

In order to present a brief history of credibility, it will be helpful to begin by describing two major statistical paradigms and three major approaches to credibility. This will facilitate our description of the historical development.

1.2 STATISTICAL PARADIGMS

Credibility is an example of a statistical estimate. Statistical estimates are obtained through the use of statistical formulas or models which, in turn, are based on statistical approaches or paradigms. There are two major statistical paradigms of current interest, which are (a) the **frequentist** or **classical** paradigm, and (b) the **Bayesian** paradigm.

In the frequentist paradigm, the probability of an event is based on its relative frequency. All prior and/or collateral information is ignored. Proponents of the frequentist paradigm view it as being objective, because all attention is devoted to the observations (data). Some of the key constructs of the frequentist paradigm are the Neyman-Pearson Lemma, tests of statistical hypotheses, confidence intervals, and unbiased estimates.

In the Bayesian paradigm, probability is treated as a rational measure of belief. Thus, the Bayesian paradigm is based on personal or subjective probabilities and involves the use of Bayes' theorem. Prior and/or collateral information is incorporated explicitly into the model via the prior distribution and the likelihood. Some of the key constructs of the Bayesian paradigm, in addition to Bayes' theorem itself, are conditional probabilities, prior distributions, predictive distributions, and (posterior) odds ratios.

1.3 WHAT IS CREDIBILITY?

Suppose we have two collections of data, as illustrated in the following figure.

Prior Observations			
#	#	#	#
#	#	#	#
#	#	#	#
#	#	#	#

Current Observations			
#	#	#	#
#	#	#	#
#	#	#	#
#	#	#	#

FIGURE 1.1

One collection consists of current observations, taken from the most recent period of observation. The second collection has observations for one or more prior periods. The various approaches to credibility give us different "recipes" for combining the two collections of observations to obtain an overall estimate.

Under some approaches to credibility, a compromise estimator, C, is calculated from the relationship

$$C = ZR + (1-Z)H, \tag{1.1}$$

where R is the mean of the current observations (for example, the data), H is the prior mean (for example, an estimate based on the actuary's prior data and/or opinion), and Z is the credibility factor, satisfying the condition $0 \le Z \le 1$. Under these approaches, the credibility estimator of the quantity of interest is derived as a linear compromise between the current observations and the actuary's prior opinion. Graphically we see that the compromise estimator, C, is somewhere on the line segment between R and H, as shown in Figure 1.2.

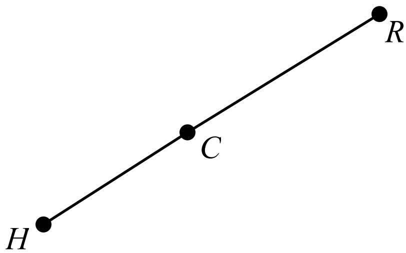

FIGURE 1.2

The symbol Z denotes the weight assigned to the (current) data and $(1-Z)$ the weight assigned to the prior data. This formulation of Equation (1.1), which includes the concept of prior data, is in the spirit of the Bayesian paradigm. As an insurance example, a new insurance rate, C, is derived as a weighted average of an old insurance rate, H, and an insurance rate, R, whose calculation is based solely on observations from a recent period. An alternative interpretation of Equation (1.1) is to let C be the insurance rate for a particular class of business, to let R be the insurance rate whose calculation is based solely on the recent experience of that class, and to let H be the insurance rate whose computation takes into account the experience of all classes combined.

To illustrate the types of practical problems that are addressed by credibility theory, we present here the statement of a problem typical of those solved in this text. Because we have not yet developed the technical machinery required to solve such a problem, we defer its solution until Section 6.6.1 (see Example 6.5).

EXAMPLE 1.1

An insurance company has two policies of group workers' compensation. The aggregate claim amounts in millions of dollars for the first three policy years are summarized in the table below. Estimate the aggregate claim amount during the fourth policy year for each of the two group policies.

Aggregate Claim Amounts			
Group Policy	**Policy Year**		
	1	2	3
1	5	8	11
2	11	13	12

Over the years there have been three major approaches to credibility: limited fluctuation, greatest accuracy, and Bayesian. The first two approaches fall under the frequentist paradigm, as neither entails the use of Bayes' theorem. Moreover, neither approach explicitly requires prior information (i.e., a formal prior probability distribution) in order to compute either the credibility factor, Z, or the estimate, C. The most well-developed approach to greatest accuracy credibility is least squares credibility. Because this approach was popularized by Hans Bühlmann, it is referred to in this text as Bühlmann's approach.

1.4 THREE APPROACHES TO CREDIBILITY

The limited fluctuation and Bühlmann approaches both involve the explicit calculation of the credibility factor, Z, and the use of Equation (1.1) to obtain the compromise estimator, C. On the other hand, the Bayesian approach requires neither the direct calculation of Z nor the use of Equation (1.1).

1.4.1 LIMITED FLUCTUATION APPROACH

Mowbray [1914] described a limited fluctuation approach for deriving the number of exposures required for full credibility, the case where $Z=1$. Perryman [1932] proposed a limited fluctuation approach to partial credibility problems, those for which $Z<1$. More modern treatments of the limited fluctuation approach to both full credibility and partial credibility are found in Longley-Cook [1962] and in Chapter 8 of Hossack, Pollard, and Zehnwirth [1983]. Outside of North America this approach is sometimes called "American credibility."

1.4.2 BÜHLMANN'S APPROACH

Bühlmann's approach, as described in this text, is based on Bühlmann [1967], which had its origins in a paper by Bailey [1942 and 1943]. Bühlmann and Straub [1972] describe an important generalization of the 1967 Bühlmann work.

1.5 BAYESIAN APPROACH TO CREDIBILITY

1.5.1 BAYESIAN STATISTICAL INFERENCE – THE EARLY YEARS

The Bayesian approach goes all the way back to the Reverend Thomas Bayes who was born in London, England around 1702. According to Stigler [1986], "Bayes was an ordained Nonconformist minister in Turnbridge Wells (about 35 miles southeast of London)." Although Bayes was elected a fellow of the Royal Society in 1742, his major work was not published until 1764, almost three years after his death. For a long time, his membership in the Royal Society was something of a mystery. Recently-discovered letters, however, now indicate that he did indeed have private correspondence with the other leading intellectuals of his era in London. When Bayes died in 1761, he left £100 and his scientific papers to his friend, Richard Price. After adding an introduction and an appendix, Price presented Bayes' essay "Toward Solving a Problem in the Doctrine of Chance" to the Royal Society.

The famous French astronomer, probabilist and mathematician Pierre Simon Laplace, who lived from 1749-1827, both championed and extended Bayes' work. In his text entitled *Essai philosophie sur les probabilities* (*Philosophical Essay on Probabilities*), Laplace described a mathematical framework for conducting statistical inference. This extended the work of Bayes and constituted the essence of Bayesian statistical inference. Laplace took this work seriously as the following passage from the beginning of his *Essay* indicates:

> *"Here I will present ... the principles and general results of the theory, applying them to the most important questions of life, which are indeed, for the most part, only questions of probability."*

Inverse Probabilities and Statistical Inference

Bayes' theorem has practical application in many fields. Kanellos [2003] presents one in a recent article about the application of Bayes' theorem to data searches entitled "18th Century Theory is New Force in Computing." Bayes' Theorem is important to actuaries because it enables them to perform statistical inference by computing *inverse probabilities*.

What exactly do we mean by "inverse probabilities"? We use the term "inverse" because we are inferring backwards from results (or effects) to causes. Let's look at some simple examples to examine this further.

A typical probability problem might be stated as follows: I have a standard die with six sides numbered from "one" through "six" and throw the die three times. What is the probability that the result of each of these three tosses of the die will be a "six"?

Now, I might have a second (non-standard) die with three sides numbered "1" and three sides numbered "six." Again I can ask the same question: What is the probability that the result of each of these three tosses of the die will be a "six"?

The idea behind inverse probabilities is to turn the question around. Here, we might observe that the results of three throws of a die were all "sixes." We then ask the question: What is the probability that we threw the standard die (as opposed to the non-standard die), given these results?

1.5.2 WHITNEY'S VIEW OF CREDIBILITY

Whitney [1918] stated that the credibility factor, Z, needed to be of the form

$$Z = \frac{n}{n+k}$$

where n represents "earned premiums" and k is a constant to be determined. The problem was how to determine k. Whitney noted that, "In practice k must be determined by judgment."[1] Whitney also noted that, "The detailed solution to this problem depends upon the use of inverse probabilities" via Bayes' Theorem.[2]

Predictive Distributions

In insurance work, we typically experience a number of claims or an aggregate amount of losses in one or more prior observation periods. The questions we want to answer are:

---

[1] See Whitney [1918, page 289].
[2] See Whitney [1918, page 277].

(1) Given such results, how many claims will we experience during the next observation period?

(2) Given such results, what will be the aggregate loss amount during the next observation period?

Using Bayes' Theorem, we can construct an entire probability distribution for such future claim frequencies or loss amounts. Probability distributions of this type are usually called **predictive distributions**. Predictive distributions give the actuary much more information than would an average or other summary statistic. A predictive distribution provides the actuary with much more information than just the expected aggregate amount of losses in the next period. It provides the actuary with a complete profile of the tail of the probability distribution of aggregate losses for use in a "value-at-risk" analysis. Thus, predictive distributions can provide the actuary and her client an important tool with which to make business decisions under uncertainty.

1.5.3 BAYESIAN STATISTICAL INFERENCE AND MODERN TIMES

Perhaps, in part, because the frequentist paradigm of statistics dominated the statistical community during the first half of the twentieth century, it remained for Bailey [1950] to rediscover and advance Whitney's ideas. During the second half of the twentieth century, Bayesian methods gained increased adherents. Two of the earliest influential books on Bayesian statistics were Savage [1954] and Raiffa and Schlaifer [1961]. Mayerson [1965] brought together the statistical developments in Bayesian statistical inference and the actuary's credibility problem, reexamining Bailey's results using the concept of a "conjugate prior distribution" and other more modern notation and terminology. Ericson [1970] and Jewell [1974] generalized Mayerson's results. Whereas Whitney and Bailey had considered only the distribution of the *number* of claims, Mayerson, Jones, and Bowers [1968] and Hewitt [1971] considered both the distribution of the *number* of claims and the distribution of the *amount* of those claims. Hewitt used some clever, artificial examples to illustrate the use of a full Bayesian approach to insurance ratemaking. It remained for Klugman [1987 and 1992], who had the advantage of modern computing equipment, to extend Hewitt's ideas and actually apply them to a major practical insurance-ratemaking problem.

1.5.4 BAYESIAN STATISTICAL INFERENCE AND MODERN COMPUTING

With the increased power of 21st-century computing equipment, advances in statistical algorithms (e.g., the EM algorithm and Markov chain Monte Carlo methods) that implement the Bayesian approach, and widely-available software that performs Bayesian inference (i.e., WinBUGS[3]), a wider class of problems is becoming susceptible to solution via the Bayesian approach.

1.6 EXPLORATORY DATA ANALYSIS

Some of the followers of John Tukey [1977] consider "exploratory data analysis" to be another distinct approach to data analysis[4]. While it is not the intention here to enter this philosophical discussion, it is often important to do substantial exploratory data analysis prior to constructing formal models, doing statistical inference, or carrying out other types of more involved statistical procedures. There are several reasons for doing this. First, substantial insight can often be gained by using simple approaches. In some situations, especially when the actuary thoroughly understands the subject matter, exploratory data analysis may yield a complete solution. As an example, we consider the following table that summarizes the experience of some mortgages insured by the Federal Housing Administration (FHA) – a component of the U. S. Department of Housing and Urban Development.

---

[3] The BUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling) Project (begun by the MRC Biostatistics unit at Imperial College, London) is concerned with the development of flexible software for Bayesian analysis of complex statistical models using Markov chain Monte Carlo methods. The "**Win**" prefix refers to Microsoft's Windows operating system. For more details about BUGS, actuaries should read David Scollnik [2001]: "Actuarial Modeling with MCMC and BUGS."

[4] In addition to Tukey's seminal reference work, cited above, other (perhaps more refined) references on exploratory data analysis include Mosteller and Tukey [1977] and Velleman and Hoaglin [1981].

EXAMPLE 1.2

Claim Rates[5] through July 1, 1989 on FHA-insured Single-family Mortgages Originated during 1981 Owner-Occupied Only

Loan-to-value ratio	Mortgage Amount (In Dollars)					
	< 25,000	25,001-35,000	35,001-50,000	50,001-60,000	Over 60,000	Overall
≤ 80.0%	8.38%	6.88%	6.74%	10.01%	6.94%	7.63%
80.1 – 85.0	20.43	12.47	11.92	11.68	8.20	11.39
85.1 - 90.0	24.33	17.43	12.59	11.76	11.43	13.69
90.1 – 95.0	27.70	23.53	18.53	19.10	17.46	19.94
95.1 – 97.0	33.48	32.42	26.76	25.88	23.51	27.77
97.1 – 100.0	42.86	52.05	40.99	31.09	18.42	42.13

Because we know from a companion table that there are only a small number of mortgages whose loan-to-value ratio is in the 97.1 – 100.0% category, we ignore that line of the table. We find a strong pattern indicating that the claim rate goes down as (1) the mortgage amount goes up and (2) as the loan-to-value ratio goes down. In particular, we note that in the roughly eight years covered by the table, more than 27% of the loans having a loan-to-value ratio in excess of 95% resulted in an insurance claim. The message of this table is clear. If you originate mortgages with little or no down-payment, the proportion of mortgages ending up in foreclosure may be substantial. It does not come as a surprise then that, after lenders originated a large number of mortgages with little or no down-payment during the period 2003-2007, a substantial number of these mortgages ended up in foreclosure. Should it come as a surprise that the housing "bubble" burst?

Second, exploratory data analysis often gives useful insight into the process generating the data. Such insight could be critical to the selection of a good model.

---

[5] Claim rate is defined as the proportion of claims received for a given origination year, on single-family mortgages insured by FHA, i.e.,

$$claim\ rate\ = \frac{number\ of\ claims}{number\ of\ mortgages\ originated}.$$

Too often large databases/data warehouses have material deficiencies involving erroneous or missing data elements, missing records, and/or duplicate records. Health insurance companies are concerned with avoiding duplicate claim payments to policyholders. Life insurance companies are concerned with (1) making payments to deceased annuitants and (2) failing to pay beneficiaries of life insurance policyholders because they are not aware that the policyholder has died. Hansen and Wang [1991] describe major deficiencies in a wide range of databases. Thus, the existence of material errors is not an unusual occurrence. Exploratory data analyses can often reveal such errors in the database under study. For a more complete discussion of how to prevent, identify, and correct faulty data, the interested reader should see Herzog, Scheuren, and Winkler [2007].

1.7 RECENT APPLICATIONS

We conclude this chapter by citing some recent applications of credibility theory to actuarial problems. Jewell [1989 and 1990] shows how to use Bayesian procedures to calculate incurred but not yet reported reserve requirements. Russo [1995] extends the work of Jewell. In order to estimate insurance reserves, Russo develops continuous time models of claim reporting and payment processes. In so doing, he employs both the Bayesian paradigm and a multistate model of the incurred claims process.

Klugman [1987] uses a full Bayesian approach to analyze actual data on worker's compensation insurance. Klugman investigates two problems. First, he calculates the joint posterior distribution of the relative frequency of claims in each of 133 rating groups. He employs three distinct prior distributions and shows that the results are virtually identical in all three instances. Second, Klugman analyzes the loss ratio for three years of experience in 319 rating classes in Michigan. He uses these data to construct prediction intervals for future observations (i.e., the fourth year). He then compares his predictions to the actual results.

The Bayesian paradigm has been used to graduate (or smooth) various types of mortality data. London [1985], building on the pioneering work of Kimeldorf and Jones [1967], provides a general description of this method. London also provides a Bayesian rationale for the historically popular Whittaker graduation method. A specific application of Bayesian graduation is found in Herzog [1983].

Young [1997, 1998] has done some research on credibility and spline functions. Her work enables the actuary to estimate future claims as a function of a statistic other than the sample mean. For example, Young [1998] argues that the use of a regression model with the predictor variable being a function of the sample geometric mean may lead to a more accurate estimator, i.e., one whose squared error loss is reduced.

As discussed in Chapter 8 of this text, Ericson [1970] and Jewell [1974] have shown that the Bühlmann estimate is equal to the Bayesian estimate of the pure premium when the claim distribution belongs to the exponential family of probability distributions and the conjugate prior is employed. Landsman and Makov [1998a] have extended this result to claim distributions belonging to the "exponential dispersion family" of distributions. Landsman and Makov [1998b] suggest a totally new approach to deal with the situation in which the claim distribution is not a member of either of the two previously-mentioned families of distributions.

Frees, et al., [1999] and Frees, et al., [2001] delineate the relationship between (1) credibility models and (2) parametric statistical models used for panel (longitudinal) data analysis.

Prior to the advent of Markov Chain Monte Carlo (MCMC) numerical methods, it was only feasible to implement a full Bayesian approach for a limited class of models. Scollnik [2001] shows how to implement Bayesian methods in actuarial models using the BUGS software package. Fellingham, Tolley, and Herzog [2005] also use BUGS to construct a Bayesian hierarchical model in order to estimate health insurance claim costs. Finally, Rosenberg and Farrell [2008] use version 1.4 of WinBUGS to construct a Bayesian statistical model in order to predict the incident and cost of hospitalization for a group of children with cystic fibrosis.

1.8 EXERCISES

1.1 Introduction

1-1 According to Rodermund, what has been the casualty actuaries' most important and enduring contribution to casualty actuarial science?

1.2 Statistical Paradigms

1-2 Name the two major statistical paradigms of current interest.

1.3 What Is Credibility?

1-3 Using Equation (1.1), determine the realization of the compromise estimator C, given that (i) the mean of the current observations is 10, (ii) the prior mean is 6, and (iii) the credibility factor is .25.

1-4 Using Equation (1.1), determine the insurance rate, C, for a particular class of business given that (i) the insurance rate calculated strictly from the experience data of that class of business is \$100, (ii) the insurance rate for all classes combined is \$200, and (iii) the credibility factor for the class is .40.

1.4 Three Approaches to Credibility

1-5 List the three major approaches to credibility.

CHAPTER 2

MATHEMATICAL PRELIMINARIES

In this chapter we review some basic probability and statistics concepts. In Section 2.1, we first define the term "conditional probability" which forms the basis for Bayes' theorem. Bayes' theorem, in turn, is the foundation of the Bayesian paradigm, a useful tool for solving a wide range of practical problems. After the statement and proof of Bayes' theorem, we present the Theorem of Total Probability, which is often useful in applying Bayes' theorem. In Section 2.2 we consider some examples of the use of Bayes' theorem. The first example is based on the target-shooting example of Philbrick [1981]; the second is taken from Hewitt [1970]. Prior and posterior probabilities are defined in Section 2.3, and the concepts of conditional expectation and unconditional expectation are reviewed in Sections 2.4 and 2.5, respectively. Hewitt's example is used to illustrate the results of Sections 2.4 and 2.5.

Exercises 2-8, 2-9, 2-14, and 2-15 deal with the estimation of the number of future insurance claims, a key component of an insurer's future liability of loss. (The provision for such liability is called the **loss reserve**, and the process of estimating the liability is called loss reserving or **loss development**.) These four exercises are based on material discussed by Brosius [1993].

2.1 BAYES' THEOREM

Definition 2.1

Let A and B represent events such that $P[B] > 0$. Then the **conditional probability of *A* given *B*** is defined to be

$$P[A \mid B] = \frac{P[A \text{ and } B]}{P[B]}. \tag{2.1}$$

The following result is named after the Reverend Thomas Bayes, who lived during the eighteenth century.

Theorem 2.1
(Bayes' Theorem)

Let A and B be events such that $P[B] > 0$. Then

$$P[A \mid B] = \frac{P[B \mid A] \cdot P[A]}{P[B]}. \tag{2.2}$$

Proof

By repeated application of the definition of conditional probability, we have

$$P[B \mid A] = \frac{P[A \text{ and } B]}{P[A]},$$

so that $P[A \text{ and } B] = P[B \mid A] \cdot P[A]$. Then

$$P[A \mid B] = \frac{P[A \text{ and } B]}{P[B]} = \frac{P[B \mid A] \cdot P[A]}{P[B]}.$$

Since the value of $P[B]$ does not depend on A, we can consider $P[A \mid B]$ to be the product of a constant, c, and the two functions of A, writing

$$P[A \mid B] = c \cdot P[B \mid A] \cdot P[A]. \tag{2.2a}$$

Alternatively we can consider $P[A \mid B]$ to be proportional to the product of the two functions of A, writing

$$P[A \mid B] \propto P[B \mid A] \cdot P[A], \tag{2.2b}$$

a construct frequently employed in applications appearing later in the text.

The next theorem is often useful in the application of Bayes' theorem.

Theorem 2.2
(Theorem of Total Probability)

Let $A_1, A_2, \ldots$ represent a countable collection of mutually exclusive and exhaustive events, so that

$$A_i \bigcap A_j = \varnothing \quad \text{for} \quad i \neq j$$

and

$$\bigcup_{i=1}^{\infty} A_i = \Omega$$

where Ω denotes the entire sample space. Then

$$P[B] = \sum_{i=1}^{\infty} P[B \mid A_i] \cdot P[A_i]. \tag{2.3}$$

Proof

We have

$$\begin{aligned} P[B] &= P[B \text{ and } \Omega] \\ &= P\left[B \text{ and } \bigcup_{i=1}^{\infty} A_i\right] = P\left[\bigcup_{i=1}^{\infty} (B \text{ and } A_i)\right] \\ &= \sum_{i=1}^{\infty} P[B \text{ and } A_i] = \sum_{i=1}^{\infty} P[B \mid A_i] \cdot P[A_i]. \end{aligned}$$

The Theorem of Total Probability is widely used in this text. Its first application is found in Example 2.2.

2.2 EXAMPLES OF THE USE OF BAYES' THEOREM

Under the notation of Chapter 1, the Bayesian approach does *not* necessarily produce a linear estimate of the true value. In fact, the Bayesian estimate, B, does not even have to be on the line segment joining R and H, as shown in the following figure.

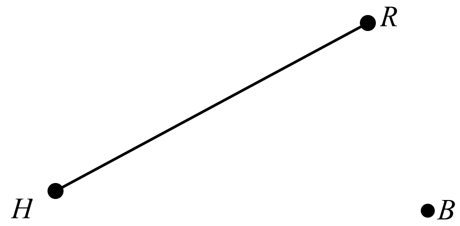

FIGURE 2.1

This is illustrated in the following example.

EXAMPLE 2.1

Consider an abbreviated form of the "target-shooting" example of Philbrick [1981], where one of two shooters, X or Y, is chosen at random (i.e., with probability ½). The shots of each shooter are uniformly distributed over two non-overlapping circular targets, illustrated in Figure 2.2a.

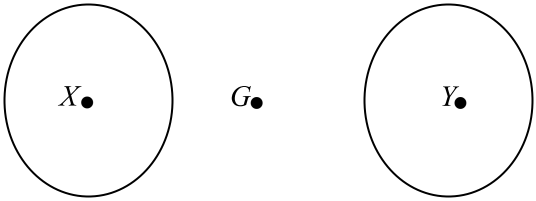

FIGURE 2.2a

The (overall) mean of the two targets is G, the point half-way between the centers of the two circles. (In the terminology of physics, G is known as the **center of gravity**.) A single shot is fired and observed to be at point S on target X, as shown in Figure 2.2b. What is the Bayesian estimate of the next shot?

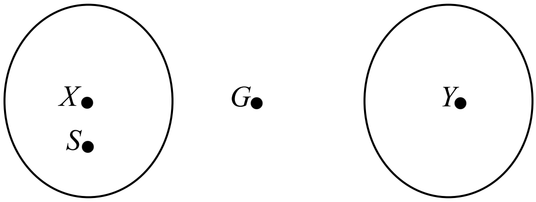

FIGURE 2.2b

SOLUTION

The answer is the center of target X. The reason[1] is that the selected shooter must be shooter X. Since the prior estimate is the point G, half-way between the centers of the two targets, the Bayesian point estimate of the location of the second shot is not on the line segment from G to the location, S, of the first shot. ❑

The following example of the use of Bayes' theorem is taken from the important paper of Hewitt [1970].

EXAMPLE 2.2

A die is selected at random (i.e., with probability ½) from a pair of "honest" dice. It is known that one die, a_1, has one marked face and five unmarked faces and the other die, a_2, has three marked faces and three unmarked faces. Let A denote the random variable representing the selection of the die. Let u denote the outcome if a toss of the die produces an unmarked face, and let m denote the outcome if the result is a marked face. Then

A_1 denotes the event $A = a_1$, the selection of the die with one marked face and five unmarked faces, and

A_2 denotes the event $A = a_2$, the selection of the die with three marked faces and three unmarked faces.

Let T_i denote the random variable representing the result of the i^{th} toss of the selected die, for $i = 1, 2, \ldots$. Then

U_i denotes the event $T_i = u$, the result of having an unmarked face showing on the i^{th} toss for $i = 1, 2, \ldots$, and

M_i denotes the event $T_i = m$, the result of having a marked face showing on the i^{th} toss for $i = 1, 2, \ldots$.

[1] While the "reason" is probably intuitive and therefore comfortable, it is not complete. To make the reasoning complete, we must employ a "loss function" (see Chapter 3). In particular, if a squared error loss function is chosen, then the center is the "best" estimate because it minimizes the sum of the squared deviations. The reader may wish to return to this example after reading Chapter 3.

Note that A and T_i denote random variables, whereas A_1, A_2, U_i, and M_i denote events. In calling each die "honest" we simply mean that

$$P[U_i \mid A_1] = \frac{5}{6} \tag{2.4a}$$

and

$$P[U_i \mid A_2] = \frac{3}{6} \tag{2.4b}$$

Calculate the value of $P[A_1 \mid U_1]$, the probability that the die with only one marked face has been drawn, given that a die was selected at random, tossed once, and resulted in an unmarked face.

SOLUTION

By Bayes' theorem we have

$$P[A_1 \mid U_1] = \frac{P[A_1 \text{ and } U_1]}{P[U_1]} = \frac{P[U_1 \mid A_1] \cdot P[A_1]}{P[U_1]}.$$

From Equation (2.4a), we have $P[U_1 \mid A_1] = \frac{5}{6}$. Because each die is chosen with probability $\frac{1}{2}$, we have $P[A_1] = \frac{1}{2}$. The value of $P[U_1]$ is computed using the Theorem of Total Probability as

$$\begin{aligned} P[U_1] &= P[A_1 \text{ and } U_1] + P[A_2 \text{ and } U_1] \\ &= P[U_1 \mid A_1] \cdot P[A_1] + P[U_1 \mid A_2] \cdot P[A_2] \\ &= \left(\frac{5}{6}\right)\left(\frac{1}{2}\right) + \left(\frac{3}{6}\right)\left(\frac{1}{2}\right) = \frac{2}{3}. \end{aligned}$$

Substituting into the Bayes' theorem equation, we obtain the result

$$P[A_1 \mid U_1] = \frac{\left(\frac{5}{6}\right)\left(\frac{1}{2}\right)}{\frac{2}{3}} = \frac{5}{8}.$$

❑

2.3 PRIOR AND POSTERIOR PROBABILITIES

In Example 2.2, we assumed that $\frac{1}{2}$ was our initial or **prior** (estimate of the) **probability** of event A_1. The word "prior" relates to the fact that this probability was assessed before the experiment of tossing the die was performed. After observing the result of the first toss to be an unmarked face, we revised our estimate of the probability of A_1 to be $\frac{5}{8}$. In symbols, we now have $P[A_1 \mid U_1] = \frac{5}{8}$. Thus our final or **posterior** (estimate of the) **probability** of A_1 given U_1 is $\frac{5}{8}$. This modification of our prior probability estimate based on recently observed data is the essence of Bayesian statistics.[2] Such modifications are frequently required in order to solve practical actuarial problems such as the calculation of insurance premium rates.

In terms of the probability distribution of the parameter A, our initial assessment was $P[A=a_1] = \frac{1}{2}$ and $P[A=a_2] = \frac{1}{2}$. Under the Bayesian paradigm, parameters are typically considered to be random variables. After observing the result of the first toss to be an unmarked face, we revised our assessment of the probability distribution of the parameter A to be $P[A=a_1 \mid T_1=u] = \frac{5}{8}$ and $P[A=a_2 \mid T_1=u] = \frac{3}{8}$. In general, the entire distribution of the prior probabilities of a parameter is called its **prior probability distribution**, and the entire distribution of **posterior probabilities** is called its posterior probability distribution. **Prior** and **posterior density functions** are similarly defined.

2.4 CONDITIONAL EXPECTATION

We now move on to a concept that is most useful in calculating insurance premium rates.

---

[2] Edwards, Lindman, and Savage (1963) summarize the Bayesian view of statistics as follows:

> "Probability is orderly opinion, and inference from data is nothing other than the revision of such opinion in the light of relevant new information."

Definition 2.2

Let X be a discrete random variable such that $x_1, x_2, \ldots$ are the only values that X takes on with positive probability. Then the **expectation of X**, denoted $E[X]$, is given by

$$E[X] = \sum_{i=1}^{\infty} x_i \cdot P[X = x_i]. \tag{2.5}$$

Definition 2.3

Using the notation of Definition 2.2, we define the **conditional expectation of X given that event A_1 has occurred**, denoted by $E[X \mid A_1]$, as

$$E[X \mid A_1] = \sum_{i=1}^{\infty} x_i \cdot P[X = x_i \mid A_1]. \tag{2.6}$$

To illustrate the concept of conditional expectation, we consider the following example, also based on Hewitt [1970].

EXAMPLE 2.3

A spinner is selected at random (i.e., with probability $\frac{1}{2}$) from a pair of spinners. It is known that (a) one spinner, b_1, has six equally likely sectors, five of which are marked "two" and one of which is marked "fourteen," and (b) the other spinner, b_2, has six equally likely sectors, three of which are marked "two" and three of which are marked "fourteen." Let B denote the random variable representing the selection of the spinner. Also let

B_1 denote the event $B = b_1$, the selection of the spinner with five "two's" and one "fourteen," and

B_2 denote the event $B = b_2$, the selection of the spinner with three "two's" and three "fourteen's."

Let S_i denote the random variable representing the result of the i^{th} spin, for $i = 1, 2, \ldots$. Calculate (a) the value of $E[S_1 \mid B_1]$, the conditional expectation of the value of a single spin, given that the spinner with one "fourteen" has been selected, and (b) the value of $E[S_1 \mid B_2]$.

SOLUTION

(a) By the definition of conditional expectation (Definition 2.3) we have

$$E[S_1 \mid B_1] = 2 \cdot P[S_1 = 2 \mid B_1] + 14 \cdot P[S_1 = 14 \mid B_1]$$
$$= 2\left(\frac{5}{6}\right) + 14\left(\frac{1}{6}\right) = 4.$$

(b) In a similar manner we can find

$$E[S_1 \mid B_2] = 2\left(\frac{3}{6}\right) + 14\left(\frac{3}{6}\right) = 8.$$ ❑

The following example further illustrates the use of Bayes' theorem.

EXAMPLE 2.4

Calculate the value of $P[B_1 \mid S_1 = 2]$.

SOLUTION

As in Example 2.2, from Bayes' theorem we have

$$P[B_1 \mid S_1 = 2] = \frac{P[S_1 = 2 \mid B_1] \cdot P[B_1]}{P[S_1 = 2]},$$

where $P[S_1 = 2 \mid B_1] = \frac{5}{6}$ and $P[B_1] = \frac{1}{2}$, and, from Theorem 2.2, we find

$$P[S_1 = 2] = P[S_1 = 2 \mid B_1] \cdot P[B_1] + P[S_1 = 2 \mid B_2] \cdot P[B_2]$$
$$= \left(\frac{5}{6}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{2}{3}.$$

Then we can calculate

$$P[B_1 \mid S_1 = 2] = \frac{\left(\frac{5}{6}\right)\left(\frac{1}{2}\right)}{\frac{2}{3}} = \frac{5}{8}.$$ ❑

(The reader is encouraged to calculate other such probabilities.)

2.5 UNCONDITIONAL EXPECTATION

The following theorem is useful in calculating pure premium estimates, as will be demonstrated in Chapter 4.

Theorem 2.3

Let $A_1, A_2, \ldots$ represent a countable collection of mutually exclusive and exhaustive events, and let X be a discrete random variable for which $E[X]$ exists. Then

$$E[X] = \sum_{i=1}^{\infty} E\left[X \mid A_i\right] \cdot P\left[A_i\right]. \qquad (2.7)$$

Proof

Because X is a discrete random variable, we have from Definition 2.2

$$E[X] = \sum_{j=1}^{\infty} x_j \cdot P[X = x_j]. \qquad (2.5)$$

By the Theorem of Total Probability we have

$$P[X = x_j] = \sum_{i=1}^{\infty} P\left[X = x_j \mid A_i\right] \cdot P[A_i].$$

Then we can rewrite Equation (2.5) as

$$E[X] = \sum_{j=1}^{\infty} x_j \sum_{i=1}^{\infty} P\left[X = x_j \mid A_i\right] \cdot P[A_i].$$

Interchanging the order of summation we obtain

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_j \cdot P[X = x_i \mid A_i] \cdot P[A_i] \\ &= \sum_{i=1}^{\infty} P[A_i] \cdot \sum_{j=1}^{\infty} x_j \cdot P\left[X = x_j \mid A_i\right]. \end{aligned}$$

By the definition of conditional expectation (Definition 2.3), the second summation is $E[X \mid A_i]$, so the last expression may be written as

$$E[X] = \sum_{i=1}^{\infty} P[A_i] \cdot E[X \mid A_i].$$

EXAMPLE 2.5

Calculate the expected value of the random variable S_1 defined in Example 2.3.

SOLUTION

Using Theorem 2.3 and the results of Example 2.3, we obtain

$$E[S_1] = \sum_{i=1}^{2} P[B_i] \cdot E[S_1 \mid B_i] = \left(\frac{1}{2}\right)(4) + \left(\frac{1}{2}\right)(8) = 6. \qquad \square$$

Some more sophisticated and useful applications of Theorem 2.3 will be discussed in Chapter 4.

Theorem 2.4

Let X and Y be discrete random variables and let $g(Y)$ be a function of Y for which $E[g(Y)]$ exists. Then we may write

$$E\left[g(Y)\right] = E_Y\left[g(Y)\right] = E_X\left[E_Y\left[g(Y) \mid X\right]\right], \tag{2.8}$$

where E_X denotes expectation over the sample space of X.

Proof

By definition of E_Y, we have $E\left[g(Y)\right] = E_Y\left[g(Y)\right]$.

Since Y is a discrete random variable, then

$$\begin{aligned}
E\left[g(Y)\right] &= \sum_{i=1}^{\infty} g(y_i) \cdot P[Y = y_i] \\
&= \sum_{i=1}^{\infty} g(y_i) \cdot \sum_{j=1}^{\infty} P[Y = y_i \mid X = x_j] \cdot P[X = x_j] \\
&= \sum_{j=1}^{\infty} P[X = x_j] \cdot \sum_{i=1}^{\infty} g(y_i) \cdot P[Y = y_i \mid X = x_j] \\
&= \sum_{j=1}^{\infty} P[X = x_j] \cdot E_Y\left[g(Y) \mid X = x_j\right].
\end{aligned}$$

The final expression above is the expectation, with respect to X, of the conditional expectation, with respect to Y, of the random variable $g(Y)$, given that $X = x_j$. Thus it can be rewritten as

$$E\left[g(Y)\right] = E_X\left[E_Y\left[g(Y) \mid X\right]\right].$$

We note that in the last equation the term $E_Y\left[g(Y)\,|\,X\right]$ is a function only of X.

EXAMPLE 2.6

Use the result of Theorem 2.4 to compute $E\left[S_1^2\right]$.

SOLUTION

Recall that B is the random variable representing the result of selecting either spinner b_1 or b_2 with equal probability. Then we have

$$\begin{aligned}
E\left[S_1^2\right] &= E_B\left[E_{S_1}\left[S_1^2\,|\,B\right]\right] \\
&= \frac{1}{2}\cdot E_{S_1}\left[S_1^2\,|\,B=b_1\right] + \frac{1}{2}\cdot E_{S_1}\left[S_1^2\,|\,B=b_2\right] \\
&= \frac{1}{2}\left[(2)^2\left(\frac{5}{6}\right) + (14)^2\left(\frac{1}{6}\right)\right] + \frac{1}{2}\left[(2)^2\left(\frac{1}{2}\right) + (14)^2\left(\frac{1}{2}\right)\right] \\
&= \frac{1}{2}(36) + \frac{1}{2}(100) \;=\; 68.
\end{aligned}$$

❑

2.6 EXERCISES

2.1 Bayes' Theorem
2.2 Examples of the Use of Bayes' Theorem

2-1 What is $P\left[A_2\,|\,U_1\right]$?

2-2 Let the conditions be as in Example 2.2. As before we select a die at random, but now toss it twice rather than just once. What is the probability that the die with only one marked face has been drawn, if both tosses result in unmarked faces? (Symbolically this is given by $P\left[A_1\,|\,U_1 \text{ and } U_2\right]$.)

2-3 Let the conditions be as in Example 2.2, except that $P[A_1]=\frac{5}{8}$ and $P[A_2]=\frac{3}{8}$. What is $P[A_1 \mid U_2]$ in this case?

2-4 A box contains 4 red balls and 6 white balls. A sample of size 3 is drawn without replacement from the box. What is the probability of obtaining 1 red ball and 2 white balls, given that at least 2 of the balls in the sample are white?

2-5 Defective items on an assembly line occur independently with probability .05. A random sample of 100 items is taken. What is the probability that the first sampled item is *not* defective, given that at least 99 of the sampled items are *not* defective?

2-6 Box I contains 3 red marbles and 2 blue marbles. Box II contains 3 red marbles and 7 blue marbles. Box I is selected with probability $\frac{2}{3}$ and Box II is selected with probability $\frac{1}{3}$. A box is selected and a red marble is drawn from the selected box. What is the probability that Box I was selected?

2-7 An insured population of individual drivers consists of 1500 youthful drivers and 8500 adult drivers. The probability distribution of claims for individual insureds during a single policy year is as follows:

Number of Claims	**Probability for Youth**	**Adult**
0	.50	.80
1	.30	.15
2	.15	.05
3	.05	.00

A particular policy has exactly 1 claim. What is the probability that the insured is a youthful driver?

2-8 A property-casualty insurance company issues automobile policies on a calendar year basis only. Let X be a random variable representing the number of accident claims reported during calendar year 2005 on policies issued during calendar year 2005. Let Y be a random variable representing the total number of accident claims that will eventually be reported on policies issued during calendar year 2005. The probability that an individual accident claim on a 2005 policy is reported during calendar year 2005 is d. Assume that the reporting times of individual claims are mutually independent. Assume also that Y has the negative binomial distribution, with fixed parameters r and p, given by

$$P[Y=y] = \binom{r+y-1}{y} p^r (1-p)^y, \tag{2.9}$$

for $y=0,1,\ldots$. Calculate $P[Y=y\,|\,X=x]$, the probability that the total number of claims reported on 2005 policies is y, given that x claims have been reported by the end of the calendar year. [Hint: The solution requires the use of Theorems 2.1 and 2.2, and the identity $\binom{y}{x}\binom{r+y-1}{y} = \binom{r+x-1}{x}\binom{(r+x)+(y-x)-1}{y-x}$. In all, substantial algebraic manipulation is involved.]

2-9 An insurer believes that the number of claims, Y, that will occur during calendar year 2005 is uniformly distributed over the set $\{2,3,4,5,6\}$, so that

$$P[Y=y] = \begin{cases} .2 & y=2,3,4,5,6 \\ 0 & \text{elsewhere} \end{cases}.$$

The insurer further believes that any claim occurring during the calendar year has a 50% chance of being reported before the end of the calendar year, and that the reporting of one claim does not influence the reporting of any other claims. Let X be a random variable representing the number of claims occurring during 2005 and reported before the end of 2005. Calculate the values of $P[Y=y\,|\,X=x]$, for $x=0,1,\ldots,6$ and $y=x,x+1,\ldots,6$.

2.3 Prior and Posterior Probabilities
2.4 Conditional Expectation

2-10 Let R_1 and R_2 be stochastically independent random variables, each with probability density function $f(x) = e^{-x}$, for $x \geq 0$. Calculate $E\left[R_1^2 + R_2^2 \mid R_1 = r_1\right]$, for $r_1 > 0$.

2-11 Let X and Y be stochastically independent random variables each with density function $f(x)$ defined by

$$f(x) = \begin{cases} 0 & x < 0 \\ .25 & x = 0 \\ .75e^{-x} & x > 0 \end{cases}.$$

Calculate the expected value of $X^2 + Y^3$, given that $X = 3$ and $Y > 0$.

2-12 Let X and Y be discrete random variables with joint density function $f(x,y)$ concentrated on the four corners of a unit square $\big((0,0),(0,1),(1,0)$ and $(1,1)\big)$. Let $f(x,y)$ be defined as follows:

x	y	$f(x,y)$
0	0	.1
0	1	.2
1	0	.3
1	1	.4

Calculate each of the following:

(a) $E_Y\left[Y \mid X = 1\right]$

(b) $E_X\left[X \mid Y = 0\right]$

2-13 Let X and Y be continuous random variables with joint density function

$$f(x,y) = \begin{cases} 6xy+3x^2, & 0<x<y<1 \\ 0 & \text{otherwise} \end{cases}.$$

Find $E\left[X \mid Y=y\right]$, for $0<y<1$, via the following three steps:

(a) First determine $f(y)$, for $0<x<y<1$.

(b) Then determine $f(x \mid y)$, for $0<x<y<1$.

(c) Finally calculate $E[X \mid Y=y]$.

2-14 Using the results and notation of Exercise 2-9, calculate each of the following:

(a) The conditional expectation $E[Y \mid X=3]$, the total expected number of claims on 2005 policies that will be reported in 2006 and beyond, given that 3 claims on such policies were reported during 2005.

(b) $E[Y \mid X=3]-3$, the expected number of claims reported after 2005 on policies issued during 2005.

2-15 Show that if X and Y are as defined in Exercise 2-8, then

$$E[Y \mid X=x] = x+(r+x)\cdot\frac{(1-d)(1-p)}{1-(1-d)(1-p)},$$

for $x=0,1,\ldots$. This result can be restated as follows:

Given that x claims were reported on the 2005 book of business during calendar year 2005, then the expected number of claims that will eventually be reported on the 2005 book of business is $x+(r+x)\cdot\frac{(1-d)(1-p)}{1-(1-d)(1-p)}$, for $x=0,1,\ldots$. Alternatively, the expected number of claims that will be reported on the 2005 book of business in 2006 and beyond is $\frac{(r+x)(1-d)(1-p)}{1-(1-d)(1-p)}$, for $x=0,1,\ldots$. [Hint: The expected value of a random variable having a negative binomial distribution with parameters n and q is $\frac{n(1-q)}{q}$. The distribution of $Y \mid X=x$ is the same as that of $U+X$, where U has a negative binomial distribution with parameters $n=r+x$ and $q = 1-(1-d)(1-p)$.]

2.5 Unconditional Expectation

2-16 Let the discrete random variables X and Y be defined as in Exercise 2-12. Calculate the unconditional expectation $E[X]$.

2-17 Let the continuous random variables X and Y be defined as in Exercise 2-13.

(a) Show that the marginal probability density function of X is

$$f(x) = \begin{cases} 3x+3x^2-6x^3 & 0<x<1 \\ 0 & \text{otherwise} \end{cases}.$$

(b) Calculate the unconditional expectation $E[X]$.

2-18 Assume that the number of insurance claims, R, filed by an individual in a single policy year has a binomial distribution with parameter for Θ for $r=0,1,2,3$. Assume further that the parameter Θ has density function $g(\theta)=6(\theta-\theta^2)$, for $0<\theta<1$. Determine the unconditional expectation $E[R]$.