

ACTEX Academic Series

Risk Models and Their Estimation

Stephen G. Kellison, FSA, EA, MAAA
University of Central Florida (Retired)

Richard L. London, FSA
University of Connecticut (Retired)

ACTEX Publications, Inc.
Winsted, CT

Copyright © 2011 by ACTEX Publications, Inc.

All rights reserved. No portion of this textbook may be reproduced by any means without prior written permission from the copyright owner.

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

Cover design by Christine Phelps and Jeffrey D. Melaragno

ISBN: 978-1-62542-912-4

PREFACE

The basic subject matter of actuarial science is the analysis and management of financial risk, often within the framework of insurance and other risk-sharing arrangements. To accomplish this, actuaries develop *risk models* to describe and measure the various risks that arise in their work. These models are probabilistic in nature. Increasingly, actuaries are also finding applications for these risk models in a variety of other contexts.

The basic idea is that each risk model is governed by an underlying, operative (but unknown) probability distribution. We attempt to *estimate* this underlying distribution by gathering sample data, and processing the data according to an estimation procedure (possibly involving some simplifying assumptions). We also analyze the quality of the resulting estimate.

There are two major classes of risk models that are of interest to us:

(1) *Survival Models*, which address the number and location in time of failure events. These models are included with the topic commonly referred to as *life contingencies*, and are examined on Exam MLC of the Society of Actuaries (SOA) and Exam 3L of the Casualty Actuarial Society (CAS).

(2) *Loss Models*, subdivided as (a) *frequency models*, which address the number of losses or insurance claims occurring in some interval of time, (b) *severity models*, which address the magnitude of such losses or claims, and (c) *aggregate models*, which combine frequency and severity models to address the total, or aggregate, amount of loss in some interval of time. The study of loss models has traditionally been referred to as *risk theory*. It is examined on SOA Exam C and CAS Exam 4.

In this text we presume a thorough understanding of survival models. For those needing a substantial review of these models, we recommend Chapters 5 and 6 of the textbook *Models for Quantifying Risk* (Fourth Edition), by Cunningham, et al. [7]. A condensed summary of these models is presented in Chapter 3 of this text.

A second area of required background presumed in this text is that of mathematical probability and statistics. We presume that the reader has had (at least) the standard two-semester university course in this topic, covering (at least) discrete and continuous random variables and their distributions, techniques of statistical estimation, and the development of confidence intervals and tests of hypotheses. Again, as a review, we have included here summaries of both probability theory (in Chapter 1), and statistical inference, or estimation (in Chapter 2).

A number of risk models, and some statistical analyses as well, can be conveniently pursued via stochastic simulation. Accordingly, we present, in Chapter 4, the general approach to simulating outcomes from a number of distributions, both discrete and continuous. Applications of simulation techniques to some of the problems discussed in the text are then presented throughout the later chapters.

With an understanding of basic probability theory presumed, the second major class of risk models is then developed in Part II of the text. We follow the logical pattern of developing the frequency models first (in Chapter 5), then the severity models (in Chapter 6), and finally the aggregate models (in Chapter 7). In Chapter 8 we present a common application of aggregate loss models, traditionally known as *ruin theory*. Most of the discrete frequency distributions developed in Chapter 5 and most of the continuous severity distributions developed in Chapter 6 are summarized in Appendices B and A, respectively.

Survival models can be presented in either parametric or tabular (non-parametric) form, with the tabular form predominating in practice. Frequency models are generally parametric, and necessarily discrete. Severity models are mostly parametric and continuous, but can also be tabular. The estimation of tabular models is presented in Part III of the text (Chapters 9 through 12), and estimation of parametric models is presented in Part IV of the text (Chapters 13 through 16). Each part begins with introductory comments describing the contents of that part.

The text includes a number of appendices that summarize some of the text results, provide derivations deemed to be too lengthy for convenient inclusion in the chapters themselves, provide additional background information, or present further topics not considered fundamental to the main subject of the text.

Answers to the textbook exercises are included. A separate solutions manual showing in detail how to solve the exercises is also available from ACTEX Publications. A number of the textbook exercises have been taken from examinations published by the Society of Actuaries and the Casualty Actuarial Society, and we wish to thank the professional societies for making these questions available to us.

A bibliography is included to suggest further reading for those interested in a deeper presentation of certain topics.

This text is appropriate for use in a two-semester university course on the topic of risk theory and the estimation of both classes of risk models.

The aforementioned SOA Exam C and CAS Exam 4 cover three major topics:

(1) Traditional risk theory, as presented in Chapters 5-7 of this text.
(2) Estimation of both survival models and risk models, as presented in Chapters 9-16 of this text.
(3) Basic aspects of *credibility theory*, as presented in Chapters 1-9 of the textbook *Introduction to Credibility Theory* (Fourth Edtion), by T.N. Herzog [15].

The authors would like to thank a number of people for their contributions to the development of this text.

Early drafts of the manuscript were reviewed by Thomas N. Herzog, ASA (National Association of Insurance Commissioners), Samuel A. Broverman, ASA (University of Toronto), Thomas P. Edwalds, FSA (Munich American Reassurance Company), Emiliano Valdez, FSA (University of Connecticut).

Thanks also to the professional staff at ACTEX Publications, particularly Gail A. Hall, FSA, the project editor, Marilyn J. Baleshiski, who did the typesetting and graphic arts, and Christine Phelps and Jeffrey D. Melaragno, who designed the text's cover.

We wish to acknowledge the work of the many authors who have written on these topics before we undertook this project, and whose writings we have researched over the course of our academic careers. We especially wish to thank our mutual instructor, the incomparable Geoffrey Crofts, FSA, to whom we dedicate the publication of this work.

We wish you good luck with your studies and preparation for your professional exams and actuarial careers.

Stephen G. Kellison, FSA
Orlando, Florida

Winsted, Connecticut
August 2011

Richard L. London, FSA
Storrs, Connecticut

TABLE OF CONTENTS

Preface iii

PART I
Review and Background Material

Chapter One
Review of Probability

1.1 Random Variables and Their Distributions 3
1.1.1 Discrete Random Variables 3
1.1.2 Continuous Random Variables 6
1.1.3 Mixed Random Variables 7
1.1.4 More on Moments of Random Variables 8
1.2 Survey of Particular Discrete Distributions 9
1.2.1 The Discrete Uniform Distribution 9
1.2.2 The Binomial and Multinomial Distributions 10
1.2.3 The Negative Binomial Distribution 11
1.2.4 The Geometric Distribution 12
1.2.5 The Poisson Distribution 12
1.3 Survey of Particular Continuous Distributions 13
1.3.1 The Continuous Uniform Distribution 13
1.3.2 The Normal Distribution 14
1.3.3 The Exponential Distribution 15
1.3.4 The Gamma Distribution 16
1.4 Multivariate Probability 17
1.4.1 The Discrete Case 17
1.4.2 The Continuous Case 20
1.5 Sums of Independent Random Variables 21
1.5.1 The Moments of S 21
1.5.2 Distributions Closed Under Convolution 22
1.5.3 The Method of Convolutions 23
1.5.4 Approximating the Distribution of S 24
1.6 Compound Distributions 24
1.6.1 The Moments of S 25
1.6.2 The Compound Poisson Distribution 26

Chapter Two
Review of Statistical Inference

2.1 Introduction 27
2.2 Point Estimation 27
2.2.1 Mean 28
2.2.2 Median 31
2.2.3 Mode 33
2.3 Properties of Estimators 35
2.3.1 Unbiasedness 35
2.3.2 Consistency 38
2.3.3 Efficiency 39
2.4 Interval Estimation 40
2.4.1 Confidence Intervals 40
2.4.2 Central Limit Theorem 43
2.5 Hypothesis Testing 44
2.5.1 Definitions and Concepts 44
2.5.2 Application and Interpretation 45

Chapter Three
Review of Survival Models

3.1 Introduction 49
3.2 Review of Parametric Survival Models 49
3.2.1 The Continuous Uniform Distribution 50
3.2.2 The Exponential Distribution 51
3.2.3 The Gompertz Distribution 51
3.2.4 The Makeham Distribution 52
3.3 Review of the Single-Decrement Tabular Survival Model 52
3.3.1 Discrete Functions Derived from ℓ_x 53
3.3.2 Continuous Functions Derived from ℓ_x 54
3.3.3 Approximation of Continuous Functions 55
3.3.4 Estimation of the Life Table Model 56
3.4 Review of the Double-Decrement Tabular Survival Model 57

Chapter Four
Review of Stochastic Simulation

4.1 Introduction 61
4.1.1 The Simulation Procedure 61
4.1.2 Linear Congruential Random Number Generators 62
4.2 The Inversion Method for Continuous Distributions 63
4.2.1 The Exponential Distribution 65
4.2.2 The Pareto Distribution 65
4.2.3 Additional Continuous Distributions 66

4.3 The Table Look-Up Method for Discrete Distributions 67
4.3.1 The Discrete Uniform Distribution 67
4.3.2 The Bernoulli Distribution 68
4.3.3 The Binomial Distribution 68
4.3.4 The Negative Binomial Distribution 69
4.3.5 The Poisson Distribution 70
4.3.6 Applications of the Material of Section 4.3 71
4.4 Generating Simulated Normal Distribution Values 72
4.4.1 The Polar Method 72
4.4.2 The Box-Muller Method 73
4.5 Sample Size 74
4.6 The Bootstrap Method 75
4.7 Applications in Loss Models and Statistical Estimation 78
4.8 Exercises 79

PART II
Loss Models

Chapter Five
Claim Frequency Models

5.1 Introduction 85
5.2 Section 1.2 (Discrete Distributions) Revisited 85
5.2.1 The Binomial Distribution 85
5.2.2 The Poisson Distribution 86
5.2.3 The Negative Binomial Distribution 90
5.2.4 The Geometric Distribution 92
5.2.5 Summary of the Recursive Relationships 92
5.2.6 Probability Generating Functions 93
5.3 Creating Additional Counting Distributions 95
5.3.1 Compound Frequency Models 95
5.3.2 Mixture Frequency Models 99
5.3.3 Truncation or Modification at Zero 102
5.4 Counting Processes 105
5.4.1 Properties of Counting Processes 105
5.4.2 The Poisson Counting Process 106
5.4.3 Further Properties of the Poisson Counting Process 107
5.4.4 Poisson Mixture Processes 109
5.4.5 The Nonstationary Poisson Counting Process 110
5.5 Summary 112
5.6 Exercises 112

Chapter Six
Claim Severity Models

6.1 Introduction 117
6.1.1 The Normal and Exponential Distributions 117
6.1.2 The Pareto Distribution 118
6.1.3 Analytic Measures of Tail Weight 120
6.2 Generating Additional Distributions 121
6.2.1 Summation 121
6.2.2 Scalar Multiplication 122
6.2.3 Power Operations 123
6.2.4 Exponentiation and the Lognormal Distribution 126
6.2.5 Summary of Severity Distributions 127
6.2.6 Mixtures of Distributions 129
6.2.7 Frailty Models 133
6.2.8 Spliced Distributions 134
6.2.9 Limiting Distributions 136
6.3 Modifications of the Loss Random Variable 137
6.3.1 Deductibles 137
6.3.2 Policy Limits 139
6.3.3 Relationships between Deductibles and Policy Limits 141
6.3.4 Coinsurance Factors 144
6.3.5 The Effect of Inflation 145
6.3.6 Effect of Coverage Deductibles on Frequency Models 147
6.4 Tail Weight Revisited; Risk Measures 152
6.4.1 The Mean Excess Loss Function 152
6.4.2 Conditional Tail Expectation 154
6.4.3 Value at Risk 155
6.4.4 Distortion Risk Measures 157
6.4.5 Risk Measures Using Discrete Distributions 159
6.5 Empirical Loss Distributions 160
6.6 Summary 163
6.7 Exercises 163

Chapter Seven
Models for Aggregate Payments

7.1 Introduction 171
7.2 Individual Risk versus Collective Risk 171
7.3 Selection of Frequency and Severity Distributions 174
7.3.1 Frequency 174
7.3.2 Severity 175
7.3.3 Frequency-Severity Interaction 175
7.4 More on the Collective Risk Model 176
7.4.1 Convolutions of the Probability Function of X 176
7.4.2 Convolutions of the CDF of X 181
7.4.3 Continuous Severity Distributions 184
7.4.4 A Final Thought Regarding Convolutions 189

7.5 Effect of Coverage Modifications 191
7.5.1 Modifications Applied to Individual Losses 191
7.5.2 Modifications Applied to the Aggregate Loss (Stop-Loss Reinsurance) 192
7.6 Infinitely Divisible Distributions 197
7.6.1 Definition of Infinite Divisibility 198
7.6.2 The Poisson Distribution 198
7.6.3 The Negative Binomial Distribution 198
7.7 Summary 198
7.8 Exercises 199

Chapter Eight
Surplus Process Models (Risk Theory)

8.1 Introduction 205
8.2 The Compound Poisson Process 205
8.2.1 Moments of the Compound Poisson Process 206
8.2.2 Other Properties of the Compound Poisson Process 206
8.3 The Surplus Process Model 207
8.4 The Probability of Ruin 210
8.4.1 The Adjustment Coefficient 210
8.4.2 The Probability of Ruin 213
8.5 The Distribution of Surplus Deficit 215
8.5.1 The Event of $U(t) < u$ 215
8.5.2 The Cumulative Loss of Surplus 217
8.6 Probability of Ruin in Finite Time 220
8.7 Summary 222
8.8 Exercises 222

PART III
Estimation of Tabular Models

Chapter Nine
Estimation from Complete Data

9.1 Introduction 231
9.2 Complete Exact Data 231
9.2.1 Frequency Models 231
9.2.2 Severity Models 233
9.2.3 Survival Models 234
9.2.4 Analysis of the Observed (Empirical) Survival Function 236
9.2.5 Interval Estimates 239
9.2.6 Estimation of $S(t)$ Using Cumulative Hazard 241
9.3 Complete Grouped Data 245
9.3.1 The General Case 245

9.3.2 Survival Model Estimation Using Equal Intervals 249
9.3.3 Estimation of $S(i)$, ${}_i|q_0$, and q_i 249
9.3.4 Estimation of $S(i)$ from $\{\hat{p}_j\}$ 252
9.4 Summary 253
9.5 Exercises 255

Chapter Ten
Estimation from Incomplete Data

10.1 Introduction 259
10.2 Causes of Incomplete Data 259
10.2.1 Truncation of Data 259
10.2.2 Censoring of Data 260
10.3 Estimation Procedures 262
10.3.1 Product-Limit Estimation with Incomplete Data 262
10.3.2 Nelson-Aalen Estimation with Incomplete Data 264
10.3.3 Determination of the Risk Set 265
10.3.4 Application to Severity Model Estimation 265
10.4 Properties of the Product-Limit Estimator 266
10.4.1 Expected Value of $\hat{S}(t_i)$ 267
10.4.2 Variance of $\hat{S}(t_i)$ 267
10.5 Application to Actuarial Survival Studies 270
10.6 Summary 273
10.7 Exercises 274

Chapter Eleven
Actuarial Survival Model Estimation

11.1 Introduction 279
11.2 Estimation of q_x from Exact Data 280
11.2.1 Decimal Years 280
11.2.2 Exact Ages 282
11.2.3 Estimation of q_x 284
11.3 Modified Product-Limit Estimation 285
11.4 Insuring Age Studies 286
11.4.1 Valuation Year of Birth 287
11.4.2 Anniversary-to-Anniversary Studies 287
11.4.3 Select Studies 290
11.5 Fiscal Age Studies 291
11.5.1 Fiscal Year of Birth 292
11.5.2 Observation Periods for Fiscal Age Studies 292
11.5.3 New Members and Withdrawals 292
11.6 Withdrawal or Double-Decrement Studies 294
11.7 Summary 296
11.8 Exercises 297

Chapter Twelve
Revision of Tabular Models

12.1 Introduction 303
12.2 Kernel Density Estimation 303
12.2.1 The Uniform Kernel Density Estimator 304
12.2.2 The Triangular Kernel Density Estimator 308
12.2.3 Other Kernel Density Estimators 310
12.2.4 Summary of the Kernel Density Estimator 312
12.3 Piece-Wise Parametric Revision 313
12.3.1 Basic Properties of Graduation by Splines 313
12.3.2 The Two-Arc Cubic Spline 314
12.3.3 The Three-Arc Cubic Spline 316
12.3.4 The General Case 317
12.3.5 The Knots 318
12.3.6 Summary of Graduation by Splines 318
12.4 Whittaker Graduation 318
12.4.1 The Basic Graduation Formula 319
12.4.2 The Minimization of M 319
12.4.3 Variations on the Standard Whittaker Form 320
12.5 Summary 321
12.6 Exercises 321

PART IV
Estimation of Parametric Models

Chapter Thirteen
Parameter Estimation Using Moments or Percentiles

13.1 Introduction 329
13.2 Method of Moments – Complete Exact Data 329
13.2.1 Continuous Distributions 330
13.2.2 Discrete Distributions 332
13.2.3 Special Handling for the Binomial 334
13.3 Method of Moments – Extensions 336
13.3.1 Complete Grouped Data 336
13.3.2 Incomplete Data – Censoring 339
13.3.3 Incomplete Data – Truncation 342
13.4 Percentile Matching 344
13.4.1 Smoothed Empirical Percentiles 345
13.4.2 Complete Exact Data 348
13.4.3 Extensions 351
13.5 Some Advanced Illustrations 351
13.6 Summary 354
13.7 Exercises 355

Chapter Fourteen
Parameter Estimation Using Maximum Likelihood

14.1 Introduction 361
14.2 Complete Exact Data – One Parameter 361
14.2.1 Continuous Distributions 362
14.2.2 Discrete Distributions 365
14.3 Extensions 367
14.3.1 Complete Grouped Data 367
14.3.2 Incomplete Data – Censoring 369
14.3.3 Incomplete Data – Truncation 370
14.3.4 Multiple Parameters 373
14.4 Variances of Parameter Estimators 375
14.4.1 One Parameter 375
14.4.2 Multiple Parameters 380
14.4.3 Confidence Intervals 382
14.5 Variances of Functions of Parameter Estimators 384
14.5.1 One Parameter 384
14.5.2 Multiple Parameters 385
14.6 Some Advanced Illustrations 387
14.7 Summary 391
14.8 Exercises 392

Chapter Fifteen
Bayesian Techniques

15.1 Introduction 399
15.2 Bayes' Theorem 400
15.2.1 Discrete Case 400
15.2.2 Continuous Case 406
15.3 Bayesian Parameter Estimation 408
15.3.1 Continuous Mixture Models 409
15.3.2 Estimating Parameters 414
15.4 Loss Functions 417
15.4.1 Squared Error Loss 417
15.4.2 Absolute Loss 419
15.4.3 Almost Constant Loss 420
15.5 Confidence (Credibility) Intervals 421
15.5.1 Equal Probability 421
15.5.2 Highest Posterior Density 422
15.5.3 Bayesian Central Limit Theorem 424
15.6 Bayesian Graduation 425
15.7 Summary 433
15.8 Exercises 434

Chapter Sixteen
Testing and Selecting Models

16.1 Introduction 443
16.2 Graphical Comparisons 444
16.2.1 Overlays of PDFs and CDFs 444
16.2.2 $D(x)$ Plot 450
16.2.3 *p-p* Plot 452
16.3 Formal Tests of Fit 454
16.3.1 Chi-Square Test 455
16.3.2 Kolmogorov-Smirnov Test 459
16.3.3 Anderson-Darling Test 463
16.4 Selecting a Model 465
16.4.1 Judgment-Based Approaches 466
16.4.2 Score-Based Approaches 467
16.5 Parsimony 469
16.5.1 Likelihood Ratio Test 470
16.5.2 Schwarz Bayesian Criterion 471
16.6 Summary 472
16.7 Exercises 473

Appendix A: Summary of Continuous Distributions 483
Appendix B: Summary of Discrete Distributions 497
Appendix C: Statistical Tables 499
Appendix D: The Forward Finite Difference Operator 501
Appendix E: Derivation of the Whittaker Equation (12.21) Minimizing Vector 503
Appendix F: Iteration Methods 505
Appendix G: Asymptotic Properties of Maximum Likelihood Estimators 507
Appendix H: Inverting a 2 x 2 Matrix 511
Appendix I: Other Parameter Estimation Techniques 513
Appendix J: Derivation of Equation (16.8c) for the Anderson-Darling Statistic 517
Appendix K: Review of the Incomplete Gamma Function 521

Answers to the Exercises 527

Bibliography 551

Index 555

PART I

REVIEW AND BACKGROUND MATERIAL

The primary purpose of this text is to present a detailed description of models for insurance losses, including the estimation of those models from sample data. In addition, we also discuss the estimation of survival models from sample data, with the description of such models contained in Cunningham, et al. [7], which could be viewed as a companion text to this one.

The topic of loss models, traditionally referred to as risk theory, relies heavily on general probability theory, and the estimation of models relies on the general theory of mathematical statistics. Accordingly, as background to the major work of this text, we first present a broad review of both probability (in Chapter 1) and statistics (in Chapter 2).

With survival models covered in detail in the Cunningham text, and examined on SOA Exam MLC and CAS Exam 3L, we present a summary review of that topic in Chapter 3 in advance of our discussion of estimating survival models from sample data.

The topic of simulation is treated differently on Exam C/4 than are the other topics reviewed in this part of the text, in that exam questions directly addressing the theory and application of simulation can be expected. In light of this, we have included both examples and end-of-chapter exercises for this topic in Chapter 4.

CHAPTER ONE

REVIEW OF PROBABILITY

In this text we make major use of many results from probability theory. We assume that the reader has completed at least one full semester in calculus-based probability at the university level, so that most of the material contained in this chapter will be somewhat familiar. For several selected topics (see, in particular, Sections 1.5 and 1.6), we are less sure that the reader will have this prior familiarity and we will present those topics in greater detail.

Specialized applications of basic probability concepts to claim frequency models, claim severity models, and sums of random variables models are considered in Chapters 5, 6, and 7, respectively. Extensions of probability theory that are needed for these specialized applications, which would not normally be covered in a basic probability course, will be introduced as needed in the later chapters.

The most basic concepts of probability are not included in this review; the reader should refer to any standard probability textbook if a review of these concepts is needed.[1] Among the basic concepts not reviewed here are the notion of the probability of an event, negation, union, intersection, mutual exclusion, the general addition rule, conditional probability, independence, the general multiplication rule, the law of total probability, and Bayes' Theorem.

1.1 RANDOM VARIABLES AND THEIR DISTRIBUTIONS

The concept of the *random variable* is the foundation for much of the material presented in this text. Levels of risk can be quantitatively represented by random variables, and understanding the properties of these random variables allows us to analyze and manage the risk so represented. In this section of this introductory chapter we review basic aspects of random variables and their properties.

1.1.1 DISCRETE RANDOM VARIABLES

A random variable, denoted X, is said to be *discrete* if it can take on only a finite (or countably infinite) number of different values. Each value it can take on is called an *outcome* of the random variable. The set of all possible outcomes is called the *domain*, or *support*, of the random variable.[2] We let x denote a particular value in the domain.

[1] For those needing a good probability text, we recommend Hassett and Stewart's *Probability for Risk Management* [12].

[2] Technically, we should say the domain (or support) of the random variable's probability function, but the shorter phrase "domain of the random variable" is often used.

Associated with each value of x is a probability value for the random variable taking on that particular outcome. The probability value is a function of the value of the outcome, denoted $p(x)$, and is called, appropriately, the *probability function* (PF). That is, $p(x)$ gives the probability of the event $X = x$. (In some textbooks $p(x)$ is called the *probability mass function*.) The set of all probability values constitutes the *distribution* of the random variable. It is necessarily true that

$$\sum_x p(x) = 1, \tag{1.1}$$

where the summation is taken over all values of x in the domain with non-zero probability.

The *expected value* of the random variable, denoted $E[X]$, is a weighted average of all values in the domain, using the associated probability values as weights. Thus we have

$$E[X] = \sum_x x \cdot p(x), \tag{1.2}$$

where the summation is again taken over all values of x with non-zero probability. The expected value is also called the *mean of the random variable* or the *mean of the distribution*.

The expected value is a special case of the more general idea of finding the weighted average of a function of the random variable, again using the associated probability values as the weights. If $g(X)$ is any real function of the random variable X, then it can be shown that

$$E[g(X)] = \sum_x g(x) \cdot p(x) \tag{1.3}$$

gives the expected value of the function of the random variable. Note that the mean of the random variable is simply the special case that results when $g(X) = X$.

An important special case is $g(X) = X^k$, and $E[g(X)] = E[X^k]$ is called the k^{th} *moment* of the random variable. (Note that the mean is therefore the *first moment* of the random variable.) Another special case is $g(X) = (X - E[X])^2$, where the expected value of $g(X)$ is called the *variance* of the random variable and is denoted by $Var(X)$. That is,

$$Var(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 \cdot p(x). \tag{1.4a}$$

The reader will recall that an equivalent expression for $Var(X)$ is

$$Var(X) = E[X^2] - (E[X])^2, \tag{1.4b}$$

a form often more convenient for calculating $Var(X)$ than is Equation (1.4a). The positive square root of the variance is called the *standard deviation* of X, and is denoted by $SD(X)$ or sometimes by σ.

The moments of a random variable can be generated from a function called, appropriately, the *moment generating function* (MGF), and denoted by $M_X(t)$. It is defined as

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_x e^{tx} \cdot p(x). \end{aligned} \tag{1.5}$$

We recognize that this is just another example of finding the expected value of a particular function of the random variable; in this case the function is $g(X) = e^{tX}$. Note that $M_X(t)$ is a function of *t*, with the subscript *X* merely reminding us of what the random variable is for which $M_X(t)$ is the MGF.

The reader will recall that the moments are then obtained from the MGF by differentiating $M_X(t)$ with respect to *t* and evaluating at $t = 0$. The first derivative evaluated at $t = 0$ produces the first moment, the second derivative so evaluated gives the second moment, and so on. In general,

$$E[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0} = M_X^{(k)}(0) \tag{1.6}$$

gives the k^{th} moment of the random variable *X*.

Several other characteristics of the random variable are also important.

The *mode* of the distribution is the value of *x* at which the greatest amount of probability is located (i.e., the value of *x* that maximizes $p(x)$). Note that several values of *x* could be tied for the greatest amount, in which case the distribution would have several modes.

The *cumulative distribution function* (CDF) of the random variable, denoted $F(x)$, gives the accumulated amount of probability at all values of the random variable less than or equal to *x*. That is,

$$F(x) = Pr(X \le x) = \sum_{y \le x} p(y), \tag{1.7}$$

where the summation is taken over all values of *y* less than or equal to *x*.

The value of *x* for which $F(x) = r$ is called the $100r^{th}$ *percentile* of the distribution. It is the value in the domain of *X* for which the probability of being less than or equal to that value is *r*, and the probability of being greater than that value is therefore $1-r$. In particular, when $r = .50$ we are speaking of the value of *x* for which half the probability lies below (or at) that

value and half lies above that value. The value of x in this case is called the *median* of the distribution.[3]

1.1.2 CONTINUOUS RANDOM VARIABLES

A random variable is said to be *continuous* if it can take on any value within a defined interval (or the union of several disjoint intervals) on the real number axis. If this set of possible values, again called the domain or support of the random variable,[4] includes all values between, say, a and b, then we would define the domain as $a < x < b$. If the domain were all non-negative real values of x we would write $x \geq 0$, and if it were all real values of x we would write $-\infty < x < \infty$. Note that the defined values of x could be in several disjoint intervals, so the domain would then be the union of these disjoint intervals. For example, the domain could be all x satisfying $a < x < b$ or $c < x < d$.

Associated with each possible value of x is an amount of *probability density*, given as a function of x by the *probability density function* (PDF), denoted by $f(x)$. Together the PDF and the domain define the distribution of the random variable. It is necessarily true that

$$\int_x f(x)\,dx = 1, \tag{1.8}$$

where the integral is taken over all values of x in the domain.

Analogous with the discrete case, we again consider the weighted average of a function of the random variable, which is the expected value of that function, this time using the density as the weight associated with each value of x. Thus is can be shown that

$$E[g(x)] = \int_x g(x) \cdot f(x)\,dx. \tag{1.9}$$

The same special cases apply here as in the discrete case. For $g(X) = X$ we have

$$E[X] = \int_x x \cdot f(x)\,dx \tag{1.10}$$

as the expected value (or first moment) of the random variable. For $g(X) = X^k$ in general we have

$$E[X^k] = \int_x x^k \cdot f(x)\,dx \tag{1.11}$$

as the k^{th} moment. As before, the variance is given by

---

[3] The median (or any other percentile) in a discrete distribution is not always clear. For example, if $p(0) = \frac{1}{3}$ and $p(1) = \frac{2}{3}$, then what is the median? Clearly there is no unique value of x for which $F(x) = .50$. Either we would say the median does not exist, or we would adopt a definition to resolve the question in each case.

[4] As in the discrete case of Section 1.1.1, the phrase "domain (or support) of the density function of the random variable" is more technically correct, but the briefer phrase "domain of the random variable" is often used.

$$Var(X) = E[(X - E[X])^2] = \int_x (x - E[X])^2 \cdot f(x)\, dx \tag{1.12}$$

and the moment generating function is given by

$$M_X(t) = E[e^{tX}] = \int_x e^{tx} \cdot f(x)\, dx. \tag{1.13}$$

The mode of the distribution is the value of x associated with the greatest amount of probability density, so it can be described as the value of x that maximizes the density function. If several values of x have the same maximum density, then the distribution has more than one mode.

As in the discrete case, the cumulative distribution function (CDF) of the random variable X is defined by $F(x) = Pr(X \le x)$. It follows that

$$F(x) = \int_{-\infty}^{x} f(y)\, dy, \tag{1.14a}$$

and, conversely,

$$f(x) = \frac{d}{dx} F(x). \tag{1.14b}$$

Just as in the discrete case, the $100r^{th}$ percentile of the distribution is the value of x for which $F(x) = r$, and, in particular, the median of the distribution is the value of x for which $F(x) = .50$.

1.1.3 MIXED RANDOM VARIABLES

On occasion we encounter a random variable that is discrete in one part of its domain and continuous in the rest of the domain. Such random variables are said to have *mixed distributions*. For example, suppose there is a finite probability associated with each of the outcomes $X = a$ and $X = b$, denoted $p(a)$ and $p(b)$, respectively, and a probability density associated with all values of x on the open interval between a and b. Then it would follow that

$$p(a) + \int_a^b f(x)\, dx + p(b) = 1. \tag{1.15}$$

The k^{th} moment of the mixed random variable X would be found as

$$E[X^k] = a^k \cdot p(a) + \int_a^b x^k \cdot f(x)\, dx + b^k \cdot p(b). \tag{1.16}$$

Mixed random variables appear quite often in actuarial models, particularly in connection with insurance coverages involving a deductible, or a policy maximum, or both. We will explore these situations in Chapters 6 and 7.

1.1.4 MORE ON MOMENTS OF RANDOM VARIABLES

Earlier in this section we reviewed the basic idea of the k^{th} moment of a random variable, denoted by $E[X^k]$. This type of moment is called the k^{th} *raw moment of X*, or the k^{th} *moment about the origin.*

By contrast, the quantity $E[(X-\mu)^k]$ is called the k^{th} *central moment of X*, or the k^{th} *moment about the mean*, where $\mu = E[X]$. In particular, the second central moment, denoted by $E[(X-\mu)^2]$, gives the variance of the distribution of X, which is denoted by $Var(X)$ or sometimes by σ^2. Recall that the positive square root of the variance is called the standard deviation, and is denoted by $SD(X)$ or sometimes by σ.

The ratio of the standard deviation to the mean of a random variable is called the *coefficient of variation*, and is denoted by $CV(X)$. Thus we have

$$CV(X) = \frac{\sigma}{\mu}, \tag{1.17}$$

for $\mu \neq 0$. It measures the degree of spread of a random variable relative to its mean.

The *skewness* of a distribution measures its symmetry, or lack thereof. It is defined by

$$\gamma_3 = \frac{E[(X-\mu)^3]}{\sigma^3}, \tag{1.18}$$

the ratio of the third central moment to the cube of the standard deviation. A distribution that is symmetric, such as the normal, will have a skewness measure of zero. A positively skewed distribution will have a right hand tail and a negatively skewed distribution will have a left hand tail.

The extent to which a distribution is peaked or flat is measured by its *kurtosis,* which is defined by

$$\gamma_4 = \frac{E[(X-\mu)^4]}{\sigma^4}, \tag{1.19}$$

the ratio of the fourth central moment to the square of the variance (or the fourth power of the standard deviation). The kurtosis of a normal distribution has a value of 3, so the kurtosis of any other distribution will indicate its degree of peakedness or flatness relative to a normal distribution with equal variance. In addition, kurtosis also relates to the thickness of the tails of the distribution.

It is well known (see Equation (1.4b)) that the second central moment (the variance) is equal to the second raw moment minus the first raw moment (the mean) squared. Similar relationships hold for the higher central moments as well. For example, for the third central moment we have

$$\begin{aligned} E[(X-\mu)^3] &= E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] \\ &= E[X^3] - 3 \cdot E[X^2] \cdot E[X] + 3 \cdot E[X] \cdot (E[X])^2 - (E[X])^3 \\ &= E[X^3] - 3 \cdot E[X^2] \cdot E[X] + 2(E[X])^3. \end{aligned} \tag{1.20}$$

1.2 SURVEY OF PARTICULAR DISCRETE DISTRIBUTIONS

In this section we will review five standard discrete distributions with which the reader should be familiar. They are included here simply as a convenient reference.

1.2.1 THE DISCRETE UNIFORM DISTRIBUTION

If there are n discrete values in the domain of a random variable X, denoted $x_1, x_2, \cdots, x_n$, for which an equal amount of probability is associated with each value, then X is said to have a *discrete uniform distribution*. Its probability function is therefore

$$p(x_i) = \frac{1}{n}, \tag{1.21}$$

for all x_i. Its first moment is

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i, \tag{1.22a}$$

and its second moment is

$$E[X^2] = \sum_{i=1}^{n} x_i^2 \cdot p(x_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \tag{1.22b}$$

In the special case where $x_i = i$, for $i = 1, 2, \cdots, n$, then we have

$$E[X] = \frac{n+1}{2} \tag{1.23a}$$

and

$$E[X^2] = \frac{(n+1)(2n+1)}{6}, \tag{1.23b}$$

so that

$$Var(X) = E[X^2] - (E[X])^2 = \frac{n^2 - 1}{12}. \tag{1.24}$$

The moment generating function in the special case is

$$M_X(t) = E[e^{tX}] = \frac{e^t(1-e^{nt})}{n(1-e^t)}. \tag{1.25}$$

1.2.2 THE BINOMIAL AND MULTINOMIAL DISTRIBUTIONS

Recall the binomial (or Bernoulli) model, in which we find the concept of repeated independent trials with each trial ending in either success of failure. The probability of success on a single trial, denoted *p*, is constant over all trials. The random variable *X*, denoting the number of successes out of *n* independent trials, is said to have a *binomial distribution*. (The special case of this distribution with $n=1$ is called a *Bernoulli distribution*.) The probability function is

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \tag{1.26a}$$

for $x = 0,1,2,\cdots,n,$ the expected value is

$$E[X] = np, \tag{1.27a}$$

the variance is

$$Var(X) = np(1-p), \tag{1.28a}$$

and the moment generating function is

$$M_X(t) = (q+pe^t)^n, \tag{1.29}$$

where $q = 1-p.$

As the prefix *bi-* implies, there are only two possible outcomes for each of the *n* trials in the binomial model, which we have called "success" and "failure." Instead of the traditional terminology, we could say that each outcome is either a Type 1 outcome, with probability p_1, or a Type 2 outcome, with probability p_2. If X_1 and X_2 denote the random variables for numbers of Type 1 and Type 2 outcomes, respectively, then X_1 and X_2 have a *joint probability function* given by

$$p(x_1, x_2) = \frac{n!}{x_1!x_2!}(p_1)^{x_1}(p_2)^{x_2}, \tag{1.26b}$$

where $x_1 + x_2 = n$ and $p_1 + p_2 = 1.$ It is easy to see that the right sides of Equations (1.26a) and (1.26b) are the same.

Now we generalize the model to one where there are *k* possible distinct outcomes for each of the *n* trials. The random variable counting the number of outcomes of Type *i* is $X_i,$ for

$i = 1, 2, \cdots, k,$ with probability p_i. Then the set of random variables $\{X_1, X_2, \cdots, X_k\}$ has a joint probability function given by

$$p(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!}(p_1)^{x_1}(p_2)^{x_2}\ldots(p_k)^{x_k}, \tag{1.26c}$$

where $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} p_i = 1$. This is called the *multinomial distribution.*[5]

The expected value of X_i is

$$E[X_i] = n \cdot p_i, \tag{1.27b}$$

its variance is

$$Var(X_i) = n \cdot p_i(1-p_i), \tag{1.28b}$$

and the covariance of any pair of random variables in the set is

$$Cov(X_i, X_j) = -n \cdot p_i \cdot p_j. \tag{1.28c}$$

1.2.3 THE NEGATIVE BINOMIAL DISTRIBUTION

Note that in the binomial distribution the random variable was the number of successes out of a fixed number of trials, *n*, where *n* is a fixed parameter of the distribution. In the *negative binomial distribution* the number of successes, denoted *r*, is a fixed parameter of the distribution and the random variable *X* represents the number of failures that occur before the r^{th} success is obtained.[6] The probability function is

$$p(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \tag{1.30a}$$

for $x = 0,1,2,\cdots$, the expected value is

$$E[X] = \frac{rq}{p}, \tag{1.31a}$$

the variance is

$$Var(X) = \frac{rq}{p^2}, \tag{1.32a}$$

[5] The multinomial distribution is a special case of a multivariate distribution, considered more generally in Section 1.4. We include it here because of its close relationship to the univariate binomial distribution.

[6] Note that if the number of failures, denoted *X*, is random, then the total number of *trials* needed to obtain *r* successes, denoted *Y*, is also random, since we would have $Y = X + r$. Some textbooks (see, for example, Hassett and Stewart [12]), discuss both the "*X*-meaning" and the "*Y*-meaning" of the negative binomial distribution.

and the moment generating function is

$$M_X(t) = \left(\frac{p}{1-qe^t}\right)^r, \tag{1.33a}$$

where, in all cases, $q = 1-p$.

Note that the description of the negative binomial distribution given here would require that the parameter r be a nonnegative integer. In Section 5.1.3, where we consider an important use of the negative binomial random variable as a model for the number of insurance claims, we will show that the requirement of an integer value for r can be relaxed.[7]

1.2.4 THE GEOMETRIC DISTRIBUTION

The *geometric distribution* is simply the special case of the negative binomial with $r = 1$. The random variable, X, now denotes the number of failures that occur before the first success is obtained.[8] Its probability function is

$$p(x) = p(1-p)^x, \tag{1.30b}$$

for $x = 0,1,2,\cdots$, the expected value is

$$E[X] = \frac{q}{p}, \tag{1.31b}$$

the variance is

$$Var(X) = \frac{q}{p^2}, \tag{1.32b}$$

and the moment generating function is

$$M_X(t) = \frac{p}{1-qe^t}, \tag{1.33b}$$

where $q = 1-p$ in all cases.

1.2.5 THE POISSON DISTRIBUTION

The *Poisson distribution* is a one-parameter discrete distribution with probability function given by

---

[7] Some textbooks use an alternate parameterization for the negative binomial distribution with $p = \frac{1}{1+\beta}$ and $q = \frac{\beta}{1+\beta}$. This is further explored in Exercise 6-38 and summarized in Appendix B.

[8] As with the negative binomial, some textbooks define the geometric random variable to be the number of trials, Y, needed to obtain the first success. In that case the probability function is $p(y) = p(1-p)^{y-1}$, for $y = 1,2,\cdots$.

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \tag{1.34}$$

for $x = 0,1,2,\cdots$, where $\lambda > 0$. Its expected value is

$$E[X] = \lambda, \tag{1.35}$$

its variance is also

$$Var(X) = \lambda, \tag{1.36}$$

and its moment generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}. \tag{1.37}$$

The Poisson distribution has several delightful properties that make it a convenient one to use in various actuarial and other stochastic applications. These will be reviewed in conjunction with the discussion of the Poisson process and the use of the Poisson distribution as a model for number of insurance losses or insurance claims in Chapter 5.

1.3 SURVEY OF PARTICULAR CONTINUOUS DISTRIBUTIONS

In this section we review four standard continuous probability distributions with which the reader should be familiar from a prior study of probability. Additional continuous distributions are introduced later in the text as claims severity distributions (in Chapter 6).

1.3.1 THE CONTINUOUS UNIFORM DISTRIBUTION

As its name suggests, the *uniform distribution* is characterized by a constant probability density at all points in its domain. If the random variable is defined on the interval $a < X < b$, and if the density function is constant, then it follows that the density function must be

$$f(x) = \frac{1}{b-a}, \tag{1.38}$$

for $a < x < b$. That is, the constant density function is the reciprocal of the length of the interval on which the random variable is defined. The mean of the uniform distribution is

$$E[X] = \frac{a+b}{2}, \tag{1.39}$$

the variance is

$$Var(X) = \frac{(b-a)^2}{12}, \tag{1.40}$$

the moment generating function is

$$M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}, \tag{1.41}$$

for $t \neq 0,$ and the cumulative distribution function is

$$F(x) = \frac{x-a}{b-a}. \tag{1.42}$$

As a consequence of the constant density function, the median is the same as the mean and the distribution is equimodal, since all points have the same probability density.

1.3.2 THE NORMAL DISTRIBUTION

The *normal distribution* will have use in the models developed later in the text. For now the reader should recall that the density function for this distribution is based on the two parameters μ and $\sigma,$ where $\sigma > 0,$ which are also the mean and standard deviation, respectively, of the distribution. Specifically,

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{1.43a}$$

for $-\infty < x < \infty,$ where, as mentioned,

$$E[X] = \mu \tag{1.44}$$

and

$$Var(X) = \sigma^2. \tag{1.45}$$

The moment generating function is

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}. \tag{1.46}$$

An extremely important property of the normal distribution is that any linear transformation of the random variable will also have a normal distribution. In particular, the random variable Z derived from the normal random variable X by the linear transformation

$$Z = \frac{X-\mu}{\sigma} \tag{1.47}$$

will have a normal distribution with mean

$$E[Z] = \frac{1}{\sigma} \cdot E[X] - \frac{\mu}{\sigma} = 0, \tag{1.48}$$

since $E[X] = \mu,$ and variance

$$Var(Z) = \frac{Var(X)}{\sigma^2} = 1, \tag{1.49}$$

since $Var(X) = \sigma^2$ and $Var(\mu/\sigma) = 0$. The random variable Z is called the *unit normal random variable* or the *standard normal random variable*. Its probability density function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{1.43b}$$

does not have a closed form antiderivative, so probability values are not found by integrating $f(z)$.[9] Rather, values of the cumulative distribution function $F_Z(z)$ are determined by approximate integration and stored in a table for look-up as needed.[10] Probability values for the normal random variable X are likewise looked up in the table of standard values after making the appropriate linear transformation. A modern alternative to table look-up is that values can be determined by numerical integration using appropriate computer software or even a pocket calculator.

1.3.3 THE EXPONENTIAL DISTRIBUTION

Another standard continuous distribution with some convenient properties is the one-parameter *exponential distribution*. It is defined over all positive values of x by the density function

$$f(x) = \beta \cdot e^{-\beta x}, \tag{1.50a}$$

for $x > 0$ and $\beta > 0$. The expected value is

$$E[X] = \frac{1}{\beta}, \tag{1.51}$$

the variance is

$$Var(X) = \frac{1}{\beta^2}, \tag{1.52}$$

the moment generating function is

$$M_X(t) = \frac{\beta}{\beta - t}, \tag{1.53}$$

for $t < \beta$, and the cumulative distribution function is

$$F(x) = 1 - e^{-\beta x}. \tag{1.54}$$

[9] The PDF of the unit normal random variable Z is sometimes denoted by $\phi(z)$ instead of $f(z)$.

[10] Similarly, the CDF is sometimes denoted by $\Phi(z)$.

(The reader should note that some textbooks prefer the notation

$$f(x) = \frac{1}{\theta} \cdot e^{-x/\theta}, \tag{1.50b}$$

so that $E[X] = \theta, Var(X) = \theta^2,$ and $M_X(t) = (1-\theta t)^{-1}$.)

1.3.4 THE GAMMA DISTRIBUTION

The two-parameter *gamma distribution* is defined by the density function

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} e^{-\beta x}, \tag{1.55}$$

for $x > 0, \alpha > 0,$ and $\beta > 0,$[11] where $\Gamma(\alpha)$ is the *gamma function* defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx. \tag{1.56}$$

By substituting $\alpha = 1$ into the gamma density given by Equation (1.55), and noting that $\Gamma(1) = 1,$ we obtain the exponential density given by Equation (1.50a). Thus the exponential is a special case of the gamma with $\alpha = 1.$ The mean of the gamma distribution is

$$E[X] = \frac{\alpha}{\beta}, \tag{1.57a}$$

the variance is

$$Var(X) = \frac{\alpha}{\beta^2,} \tag{1.57b}$$

and the moment generating function is

$$M_X(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha, \tag{1.57c}$$

for $t < \beta.$ The cumulative distribution function is given by

$$F(x) = \int_0^x f(y)\, dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x y^{\alpha-1} e^{-\beta y}\, dy. \tag{1.58a}$$

[11] Some texts prefer to use $\frac{1}{\theta}$ in place of $\beta,$ as already mentioned for the exponential distribution.

If we let $\beta y = t,$ so $y = t/\beta$ and $dy = \frac{1}{\beta} \cdot dt,$ then the integral becomes

$$\begin{aligned} F(x) &= \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} \beta^{\alpha} \left(\frac{t}{\beta}\right)^{\alpha-1} \cdot e^{-t} \cdot \frac{1}{\beta} \cdot dt \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} t^{\alpha-1} e^{-t}\, dt \\ &= \Gamma(\alpha; \beta x), \end{aligned} \tag{1.58b}$$

where $\Gamma(\alpha; \beta x)$ is the *incomplete gamma function* defined by

$$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^{x} t^{\alpha-1} e^{-t}\, dt. \tag{1.58c}$$

Further insight into the relationship of the gamma to the exponential distribution will be provided in Section 1.5, and applications of the incomplete gamma function in actuarial models will arise in Chapters 6 and 7. It is further explored in Appendix K, including its evaluation by simulation.

1.4 MULTIVARIATE PROBABILITY

Whenever two or more random variables are involved in the same model we find ourselves dealing with a case of *multivariate probability*. In this section we will review the fundamental aspects of multivariate probability, including the interrelationships among the *joint*, *marginal*, and *conditional distributions*, in both the discrete and continuous cases.

One of the most important aspects of multivariate probability is the process for finding the unconditional mean and variance of a random variable from the associated conditional means and variances. The formulas relating the unconditional and conditional means and variances are given by the *double expectation theorem*. Although this is a result with which the reader might be familiar from prior study, it has so many important applications later in this text that we wish to review it in some detail at this time. We will do this by example, separately for the discrete and continuous cases. An example does not establish the general result, of course; for that purpose the reader is referred to Section 7.4 of Ross [35].

1.4.1 THE DISCRETE CASE

We illustrate the key components of discrete multivariate probability with a numerical example. Suppose the discrete random variable X can assume the values $x = 0, 1, 2$ and the discrete random variable Y can assume the values $y = 1, 2.$ Let X and Y have the joint distribution given by the following table, and let $p(x, y)$ denote the joint probability function.

Y \ X	0	1	2
1	.10	.20	.30
2	.10	.10	.20

The marginal distribution of X is given by

$$Pr(X=0) = .10+.10 = .20,$$

$$Pr(X=1) = .20+.10 = .30,$$

and

$$Pr(X=2) = .30+.20 = .50.$$

The moments of X can be calculated directly from the marginal distribution. We have

$$E[X] = (0)(.20)+(1)(.30)+(2)(.50) = 1.30,$$
$$E[X^2] = (0)(.20)+(1)(.30)+(4)(.50) = 2.30,$$

and

$$Var(X) = 2.30-(1.30)^2 = .61.$$

Now we consider an alternative, but longer (at least this time), way to find $E[X]$ and $Var(X)$. First we find the marginal distribution of Y as

$$Pr(Y=1) = .10+.20+.30 = .60$$

and

$$Pr(Y=2) = .10+.10+.20 = .40.$$

Next we find both conditional distributions for X, one given $Y=1$ and the other given $Y=2$. We have

$$Pr(X=0\,|\,Y=1) = \frac{.10}{.60} = \frac{1}{6},$$

$$Pr(X=1\,|\,Y=1) = \frac{.20}{.60} = \frac{2}{6},$$

and

$$Pr(X=2\,|\,Y=1) = \frac{.30}{.60} = \frac{3}{6}.$$

From this conditional distribution we find the conditional moments of X, given $Y=1$. We have

$$E[X\,|\,Y=1] = (0)\left(\frac{1}{6}\right)+(1)\left(\frac{2}{6}\right)+(2)\left(\frac{3}{6}\right) = \frac{8}{6},$$

$$E[X^2\,|\,Y=1] = (0)\left(\frac{1}{6}\right)+(1)\left(\frac{2}{6}\right)+(4)\left(\frac{3}{6}\right) = \frac{14}{6},$$

and

$$Var(X\,|\,Y=1) = \frac{14}{6}-\left(\frac{8}{6}\right)^2 = \frac{20}{36}.$$

Similarly we find the conditional distribution

$$Pr(X=0\,|\,Y=2) = \frac{.10}{.40} = \frac{1}{4},$$

$$Pr(X=1\,|\,Y=2) = \frac{.10}{.40} = \frac{1}{4},$$

and

$$Pr(X=2\,|\,Y=2) = \frac{.20}{.40} = \frac{2}{4},$$

and its associated conditional moments

$$E[X\,|\,Y=2] = (0)\left(\frac{1}{4}\right)+(1)\left(\frac{1}{4}\right)+(2)\left(\frac{2}{4}\right) = \frac{5}{4},$$

$$E[X^2\,|\,Y=2] = (0)\left(\frac{1}{4}\right)+(1)\left(\frac{1}{4}\right)+(4)\left(\frac{2}{4}\right) = \frac{9}{4},$$

and

$$Var(X\,|\,Y=2) = \frac{9}{4}-\left(\frac{5}{4}\right)^2 = \frac{11}{16}.$$

We now come to the key part of the operation. We recognize that the conditional expected value of X, denoted $E_X[X\,|\,Y]$, is a random variable because it is a function of the random variable Y. It can take on the two possible values $\frac{8}{6}$ and $\frac{5}{4}$, and does so with probability .60 and .40, respectively, the probabilities associated with the two possible values of Y. We can find the moments of this random variable as

$$E_Y[E_X[X\,|\,Y]] = \left(\frac{8}{6}\right)(.60)+\left(\frac{5}{4}\right)(.40) = \frac{13}{10},$$

$$E_Y[(E_X[X\,|\,Y])^2] = \left(\frac{8}{6}\right)^2\cdot(.60)+\left(\frac{5}{4}\right)^2\cdot(.40) = \frac{203}{120},$$

and

$$Var_Y(E_X[X\,|\,Y]) = \frac{203}{120}-\left(\frac{13}{10}\right)^2 = \frac{1}{600}.$$

Similarly the conditional variance of X given Y, denoted $Var_X(X\,|\,Y)$, is a random variable because it too is a function of Y. Its two possible values are $\frac{20}{36}$ and $\frac{11}{16}$, so its expected value is

$$E_Y[Var_X(X\,|\,Y)] = \left(\frac{20}{36}\right)(.60)+\left(\frac{11}{16}\right)(.40) = \frac{73}{120}.$$

Finally we observe that

$$E_Y\left[E_X[X \mid Y]\right] = \frac{13}{10} = E[X],$$

which states that the expected value of the conditional expectation is the unconditional expected value of X. This constitutes the first part of the double expectation theorem. The second part states that

$$E_Y[Var_X(X \mid Y)] + Var_Y(E_X[X \mid Y]) = \frac{73}{120} + \frac{1}{600} = \frac{366}{600} = \frac{61}{100} = Var(X),$$

which says that the expected value of the conditional variance plus the variance of the conditional expectation is the unconditional variance of X.

1.4.2 THE CONTINUOUS CASE

Multivariate probability in the continuous case is handled more compactly than in the discrete case. We cannot list all possible pairs of (x, y) in the continuous joint domain; instead we specify the joint density at the point (x, y) in the form of a joint density function denoted $f(x, y)$. Recall that the marginal density of X is then found by integrating the joint density over all values of Y, and the marginal density of Y is found by integrating the joint density over all values of X. The conditional density of X, given Y, is then found by dividing the joint density by the marginal density of Y, and, similarly, the conditional density of Y, given X, is found by dividing the joint density by the marginal density of X. These basic relationships are illustrated in the following example.

Let the continuous random variable X have a uniform distribution on the interval $0 < x < 12$, and let the continuous random variable Y have a conditional distribution, given $X = x$, that is uniform on the interval $0 < y < x$. We seek the unconditional expected value and variance of Y.

We could, of course, proceed by first finding the marginal distribution of Y and then finding the unconditional expected value and variance of Y directly from this marginal distribution. Since X is uniform we have $f_X(x) = \frac{1}{12}$, and since Y is conditionally uniform we have $f_{Y|X}(y \mid x) = \frac{1}{x}$. Then the joint density is $f(x, y) = \frac{1}{12x}$, and the marginal density of Y is

$$f_Y(y) = \int_y^{12} f(x, y)\, dx = \int_y^{12} \left(\frac{1}{12x}\right) dx = \frac{1}{12}[\ln 12 - \ln y].$$

To then find the first and second moments of Y directly from the marginal density of Y is a bit of a calculus challenge. Instead, we will find the unconditional expected value and variance of Y from its conditional moments by using the double expectation theorem. We have, since Y is conditionally uniform, $E_Y[Y \mid X] = \frac{X}{2}$ and $Var_Y(Y \mid X) = \frac{X^2}{12}$. Then, directly from the double expectation theorem, we have

$$E[Y] = E_X\left[E_Y[Y \mid X]\right] = E_X\left[\frac{X}{2}\right] = \frac{1}{2} \cdot E[X] = 3,$$

since, being uniform on $0 < x < 12$, we have $E[X] = 6$. Similarly,

$$\begin{aligned} Var(Y) &= E_X[Var_Y(Y \mid X)] + Var_X(E_Y[Y \mid X]) \\ &= E_X\left[\frac{X^2}{12}\right] + Var_X\left(\frac{X}{2}\right) \\ &= \frac{1}{12} \cdot E[X^2] + \frac{1}{4} \cdot Var(X) \\ &= \left(\frac{1}{12}\right)(48) + \left(\frac{1}{4}\right)(12) = 7, \end{aligned}$$

since $Var(X) = 12$ and $E[X^2] = Var(X) + (E[X])^2 = 48$.

In the discrete case example, presented in Section 1.4.1, the unconditional mean and variance were found more easily from the marginal distribution than via the double expectation theorem. In this continuous example, however, the opposite is true; the mean and variance of Y are found more easily via the double expectation theorem than from the marginal distribution of Y.

The double expectation theorem will have several applications throughout this text. For future reference we restate it here as

$$E[X] = E_Y[E_X[X \mid Y]] \tag{1.59a}$$

and

$$Var(X) = E_Y[Var_X(X \mid Y)] + Var_Y(E_X[X \mid Y]). \tag{1.59b}$$

1.5 SUMS OF INDEPENDENT RANDOM VARIABLES

Consider the random variable S (chosen to stand for "sum"), defined as

$$S = X_1 + X_2 + \cdots + X_n, \tag{1.60}$$

where the X_i's are all mutually independent. This model arises often in actuarial science, and is referred to there as the *individual risk model*. We will explore this model within the actuarial context more fully in Chapter 7. At this point we wish to review what we already know about the random variable S from a prior study of probability.

1.5.1 THE MOMENTS OF *S*

The expected value of S is simply the sum of the expected values of the X_i's, so we have

$$E[S] = \sum_{i=1}^{n} E[X_i]. \tag{1.61a}$$

If the X_i's all have the same distribution, so that we may use $E[X]$ to denote the common $E[X_i]$, we then have

$$E[S] = n \cdot E[X]. \tag{1.61b}$$

Because of the assumption of independence, it is also true that

$$Var(S) = \sum_{i=1}^{n} Var(X_i), \tag{1.62a}$$

which, if the X_i's all have the same distribution so they have common $Var(X)$, can be written as

$$Var(S) = n \cdot Var(X). \tag{1.62b}$$

More generally, the moments of S can be found from its moment generating function $M_S(t)$, which is itself found from the MGF's of the X_i's as

$$M_S(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t). \tag{1.63a}$$

That is, the MGF of the sum of independent random variables is the *product* of the MGF's of the several random variables in the sum. If the X_i's all have the same distribution, and therefore the same MGF, denoted $M_X(t)$, then we have

$$M_S(t) = \left[M_X(t)\right]^n. \tag{1.63b}$$

Can we go beyond knowing only the moments of the random variable S and find its actual distribution? The answer is that yes we can, in certain cases; this is pursued in the remaining subsections of this section.

1.5.2 DISTRIBUTIONS CLOSED UNDER CONVOLUTION

If the X_i random variables belong to certain families of distributions, then the random variable S will belong to that same family, albeit with different parameter values. Distributions for which this property holds are said to be *closed under convolution*.

For example, if each X_i is a Poisson random variable with parameter λ_i, then S is a Poisson random variable with parameter $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. The result follows from Equation (1.63a). We have

$$\begin{aligned} M_S(t) &= e^{\lambda_1(e^t-1)} \cdot e^{\lambda_2(e^t-1)} \cdot \cdots \cdot e^{\lambda_n(e^t-1)} \\ &= e^{\lambda_1(e^t-1)+\lambda_2(e^t-1)+\cdots+\lambda_n(e^t-1)} \\ &= e^{\lambda(e^t-1)}, \end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$, which is the MGF for a Poisson random variable with parameter λ.

Again by using Equation (1.63a) we can show that if each X_i is a binomial random variable, with common value of the parameter p but each with its own parameter value n_i, then S is a binomial random variable with parameters $n = n_1 + n_2 + \cdots + n_n$ and the common value of p.

Similarly we can show that if each X_i is a negative binomial random variable, with common parameter p but each with its own parameter value r_i, then S is a negative binomial random variable with parameters $r = r_1 + r_2 + \cdots + r_n$ and the common value of p.

If each X_i is a geometric random variable with common parameter p, then S is a negative binomial random variable with parameters $r = n$ and the common value of p.

On the continuous side, if each X_i is a normal random variable with parameters μ_i and σ_i, then S is a normal random variable with parameters $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. (Note that it is *not* true that $\sigma = \sigma_1 + \sigma_2 + \cdots + \sigma_n$.)

If each X_i is an exponential random variable with common parameter β, then S is a gamma random variable with parameters $\alpha = n$ and the common value of β.

Finally if each X_i is a gamma random variable, with common parameter β but each with its own parameter α_i, then S is a gamma random variable with parameters $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ and the common value of β.

Note that the uniform random variable is not closed under convolution.

The property of being closed under convolution is a convenient one for the X_i random variables to have. For example, in the continuous case, it is convenient if we are justified in assuming a gamma distribution (including the special case exponential) for the X_i's so that we can then analyze the distribution of S as another gamma random variable. This will be revisited in our further discussion of the individual risk model in Chapter 7.

1.5.3 THE METHOD OF CONVOLUTIONS

Consider the individual risk model in the special case with $n = 2$, so that $S = X_1 + X_2$. Suppose the distributions of X_1 and X_2 are discrete, with only a few points in each distribution. For example, suppose we have the distributions $p_1(1) = .40$ and $p_1(2) = .60$ for X_1, and $p_2(1) = .50$, $p_2(2) = .40$ and $p_2(3) = .10$ for X_2. Then we can simply tabulate the discrete distribution of S. We observe that S can take on the values 2, 3, 4, and 5. We can then easily calculate the probability values associated with each possible value of S:

$$Pr(S=2) = Pr(X_1 = 1 \text{ and } X_2 = 1) = (.40)(.50) = .20$$

$$Pr(S=3) = Pr(X_1 = 1 \text{ and } X_2 = 2) + Pr(X_1 = 2 \text{ and } X_2 = 1) = (.40)(.40) + (.60)(.50) = .46$$

$$Pr(S{=}4) \;=\; Pr(X_1 = 1 \text{ and } X_2 = 3) + Pr(X_1 = 2 \text{ and } X_2 = 2) \;=\; (.40)(.10) + (.60)(.40) \;=\; .28$$

$$Pr(S{=}5) \;=\; Pr(X_1 = 2 \text{ and } X_2 = 3) \;=\; (.60)(.10) \;=\; .06$$

(Note, as a check on our arithmetic, that the probability values for S sum to unity.) This process for tabulating the distribution of S is called the *method of convolutions*.

If the model has $n = 3$ terms in the sum, we can use the method of convolutions recursively. With $S = X_1 + X_2 + X_3$, we first let $R = X_1 + X_2$ and find the distribution of R by convoluting the distributions of X_1 and X_2, as just discussed, and then we convolute the distributions of R and X_3 to find the distribution of S. This process can then be extended to any number of terms in the sum, noting that we only ever convolute two distributions at one time. Thus if there are n terms in the sum, we would perform $n - 1$ sequential convolutions to reach the distribution of S.

The method of convolutions to find the distribution of S in either the individual or collective risk model will be further investigated in Chapter 7 and again in Chapter 8, including the case where each X_i has a continuous distribution.

1.5.4 APPROXIMATING THE DISTRIBUTION OF *S*

If the X_i's do not belong to a family of distributions that is closed under convolution, and if the method of convolutions is too unwieldy (because the domain of each X_i is quite large and/or because we have a large value of n), so that the methods of Sections 1.5.2 and 1.5.3 are either impossible or impractical to use, we can always resort to approximating the distribution of S. If n is "sufficiently large," then we can invoke the Central Limit Theorem to conclude that S has an approximately normal distribution. The mean and variance of S are available by summing the means and variances of the X_i's, and these two values completely define the normal distribution of S. This approach will be utilized at several points throughout the text.

1.6 COMPOUND DISTRIBUTIONS

In this section we consider a variation on the sum of independent random variables model of Section 1.5, in which the number of terms in the sum is itself random. In this case we write

$$S \;=\; X_1 + X_2 + \cdots + X_N, \tag{1.64}$$

where N is used as the subscript of the last term, in contrast to the use of n in Equation (1.60), to remind us that the number of terms in the sum is a random variable. We again assume that the X_i random variables are mutually independent, and that they all have the same distribution. We further assume that each X_i is also independent of the random variable N. In prob-

ability theory, S is said to have a *compound distribution*. The distribution of N is called the *primary distribution* and the common distribution of X_i is called the *secondary distribution*.

In the context of actuarial science, the compound distribution is called the *collective risk model*. We will further discuss this model, and contrast it with the simpler individual risk model, in Chapter 7.

1.6.1 THE MOMENTS OF *S*

To find the mean and variance of S when it has the compound distribution defined by Equation (1.64) we make use of the double expectation theorem reviewed in Section 1.4. We consider S to be conditional on the random variable N. Then, conditional on having N terms in the sum, each with expected value $E[X]$ and variance $Var(X)$, we have

$$E[S \mid N] = N \cdot E[X] \tag{1.61c}$$

and

$$Var(S \mid N) = N \cdot Var(X). \tag{1.62c}$$

(We can use the notation $E[X]$ and $Var(X)$, rather than $E[X_i]$ and $Var(X_i)$, because all X_i are assumed to have the same distribution.)

By the first part of the double expectation theorem we then have

$$E[S] = E_N\big[E[S \mid N]\big] = E_N\big[N \cdot E[X]\big] = E[N] \cdot E[X], \tag{1.65}$$

since $E[X]$ is a constant with respect to the random variable N. Similarly, by the second part of the double expectation theorem we have

$$\begin{aligned} Var(S) &= E_N\big[Var(S \mid N)\big] + Var_N\big(E[S \mid N]\big) \\ &= E_N\big[N \cdot Var(X)\big] + Var_N\big(N \cdot E[X]\big) \\ &= E[N] \cdot Var(X) + Var(N) \cdot \big(E[X]\big)^2, \end{aligned} \tag{1.66}$$

since both $E[X]$ and $Var(X)$ are constants with respect to the random variable N.

We can also express the moment generating function of S as a function of the MGF's of N (the primary distribution) and X (the secondary distribution), using the first part of the double expectation theorem, since the MGF of S is just the expected value of e^{tS}. Again conditioning on N we have

$$M_S(t) = E[e^{tS}] = E_N[E[e^{tS} \mid N]].$$

But $E[e^{tS} \mid N]$ is the MGF of S, given that there are N terms in the sum. From Equation (1.63b) we know that this conditional MGF is

$$M_S(t) \mid N = E[e^{tS} \mid N] = [M_X(t)]^N,$$

which can also be written as

$$E[e^{tS} \mid N] = e^{\ln[M_X(t)]^N} = e^{N \cdot \ln M_X(t)}.$$

Let $r = \ln M_X(t)$. Then we have

$$M_S(t) = E_N\left[E[e^{tS} \mid N]\right] = E_N[e^{rN}] = M_N(r).$$

This tells us that the MGF of S is equal to the MGF of the primary distribution N, evaluated at $r = \ln M_X(t)$. Thus we can conclude that

$$M_S(t) = M_N[\ln M_X(t)]. \tag{1.67}$$

For example, suppose N has a Poisson distribution with MGF given by $M_N(t) = e^{\lambda(e^t - 1)}$, as given by Equation (1.37). Then by Equation (1.67) the MGF of S is

$$M_S(t) = M_N[\ln M_X(t)] = e^{\lambda[M_X(t) - 1]}, \tag{1.68}$$

in terms of the MGF of the random variable X.

1.6.2 THE COMPOUND POISSON DISTRIBUTION

If N has a Poisson distribution, then S is said to have a *compound Poisson distribution*. In this case the mean of S, by Equation (1.65), can be written as

$$E[S] = \lambda \cdot E[X], \tag{1.69}$$

since $E[N] = \lambda$. Further, since $Var(N) = \lambda$ as well, Equation (1.66) for the variance of S simplifies to

$$\begin{aligned} Var(S) &= \lambda \cdot Var(X) + \lambda \cdot (E[X])^2 \\ &= \lambda[Var(X) + (E[X])^2] \\ &= \lambda \cdot E[X^2], \end{aligned} \tag{1.70}$$

since $Var(X) = E[X^2] - (E[X])^2$. Note that the moments of S can also be obtained from the MGF of S, given by Equation (1.68).

The compound Poisson distribution will be extensively used in our further discussion of the collective risk model in Chapter 7 and the ruin model in Chapter 8.