

Loss Models
From Data to Decisions
Second Edition



Loss Models
From Data to Decisions
Second Edition

Stuart A. Klugman
Drake University

Harry H. Panjer
University of Waterloo

Gordon E. Willmot
University of Waterloo

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan,
David W. Scott, Adrian F. M. Smith, Jozef L. Teugels*

Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

A complete list of the titles in this series appears at the end of this volume.

 **WILEY-
INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2004 by John Wiley & Sons, Inc. All rights reserved.

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

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

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

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Klugman, Stuart A., 1949–

Loss models : from data to decisions / Stuart A. Klugman, Harry H. Panjer, Gordon E.

Willmot.—2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-21577-5 (cloth)

1. Insurance—Statistical methods. 2. Insurance—Mathematical models. I. Panjer, Harry H., 1946–. II. Willmot, Gordon E., 1957–. III. Title.

HG8781.K583 2004

368'.01—dc22

AM2NCA

PO-21420

MAR.28/06

12004048066

2004043066

Printed in the United States of America.

10 9 8 7 5 4 3

Contents

Preface xvii

Acknowledgments xxi

Part I Introduction

1	Modeling	3
1.1	The model-based approach	3
1.1.1	The modeling process	3
1.1.2	The modeling advantage	5
1.2	Organization of this book	6

Part II Actuarial models

2	Random variables	11
2.1	Introduction	11
2.2	Key functions and four models	13
2.2.1	Exercises	23

3	Basic distributional quantities	25
3.1	Moments	25
3.1.1	Exercises	33
3.2	Percentiles	34
3.2.1	Exercises	35
3.3	Generating functions and sums of random variables	36
3.3.1	Exercises	37
4	Classifying and creating distributions	39
4.1	Introduction	39
4.2	The role of parameters	40
4.2.1	Parametric and scale distributions	41
4.2.2	Parametric distribution families	42
4.2.3	Finite mixture distributions	43
4.2.4	Data-dependent distributions	45
4.2.5	Exercises	46
4.3	Tail weight	48
4.3.1	Existence of moments	48
4.3.2	Limiting ratios	50
4.3.3	Hazard rate and mean residual life patterns	50
4.3.4	Exercises	54
4.4	Creating new distributions	56
4.4.1	Introduction	56
4.4.2	Multiplication by a constant	56
4.4.3	Raising to a power	57
4.4.4	Exponentiation	58
4.4.5	Mixing	59
4.4.6	Frailty models	62
4.4.7	Splicing	64
4.4.8	Exercises	65
4.5	Selected distributions and their relationships	69
4.5.1	Introduction	69
4.5.2	Two parametric families	69
4.5.3	Limiting distributions	70
4.5.4	Exercises	72
4.6	Discrete distributions	72
4.6.1	Introduction	72

4.6.2	The Poisson distribution	73
4.6.3	The negative binomial distribution	76
4.6.4	The binomial distribution	79
4.6.5	The $(a, b, 0)$ class	81
4.6.6	Truncation and modification at zero	83
4.6.7	Compound frequency models	88
4.6.8	Further properties of the compound Poisson class	95
4.6.9	Mixed frequency models	101
4.6.10	Poisson mixtures	103
4.6.11	Effect of exposure on frequency	108
4.6.12	An inventory of discrete distributions	109
4.6.13	Exercises	109
5	Frequency and severity with coverage modifications	115
5.1	Introduction	115
5.2	Deductibles	115
5.2.1	Exercises	120
5.3	The loss elimination ratio and the effect of inflation for ordinary deductibles	121
5.3.1	Exercises	123
5.4	Policy limits	124
5.4.1	Exercises	126
5.5	Coinsurance, deductibles, and limits	126
5.5.1	Exercises	128
5.6	The impact of deductibles on claim frequency	129
5.6.1	Exercises	133
6	Aggregate loss models	135
6.1	Introduction	135
6.1.1	Exercise	138
6.2	Model choices	139
6.2.1	Exercises	140
6.3	The compound model for aggregate claims	140
6.3.1	Exercises	148
6.4	Analytic results	153
6.4.1	Exercises	156
6.5	Computing the aggregate claims distribution	159

6.6	The recursive method	161
6.6.1	Applications to compound frequency models	162
6.6.2	Underflow/overflow problems	165
6.6.3	Numerical stability	166
6.6.4	Continuous severity	166
6.6.5	Constructing arithmetic distributions	167
6.6.6	Exercises	170
6.7	The impact of individual policy modifications on aggregate payments	174
6.7.1	Exercises	177
6.8	Calculations with approximate distributions	178
6.8.1	Arithmetic distributions	178
6.8.2	Empirical distributions	181
6.8.3	Piecewise linear cdf	182
6.8.4	Exercises	184
6.9	Inversion methods	184
6.9.1	Fast Fourier transform	185
6.9.2	Direct numerical inversion	188
6.9.3	Exercise	190
6.10	Comparison of methods	190
6.11	The individual risk model	192
6.11.1	Parametric approximation	192
6.11.2	Exact calculation of the aggregate distribution	195
6.11.3	Compound Poisson approximation	201
6.11.4	Exercises	205
7	Discrete-time ruin models	209
7.1	Introduction	209
7.2	Process models for insurance	210
7.2.1	Processes	210
7.2.2	An insurance model	212
7.2.3	Ruin	213
7.3	Discrete, finite-time ruin probabilities	215
7.3.1	The discrete-time process	215
7.3.2	Evaluating the probability of ruin	216
7.3.3	Exercises	222

8	Continuous-time ruin models	223
8.1	Introduction	223
8.1.1	The Poisson process	223
8.1.2	The continuous-time problem	225
8.2	The adjustment coefficient and Lundberg's inequality	225
8.2.1	The adjustment coefficient	226
8.2.2	Lundberg's inequality	230
8.2.3	Exercises	232
8.3	An integrodifferential equation	234
8.3.1	Exercises	239
8.4	The maximum aggregate loss	239
8.4.1	Exercises	242
8.5	Cramer's asymptotic ruin formula and Tijms' approximation	244
8.5.1	Exercises	250
8.6	The Brownian motion risk process	252
8.7	Brownian motion and the probability of ruin	256
Part III Construction of empirical models		
9	Review of mathematical statistics	265
9.1	Introduction	265
9.2	Point estimation	266
9.2.1	Introduction	266
9.2.2	Measures of quality	267
9.2.3	Exercises	273
9.3	Interval estimation	275
9.3.1	Exercise	277
9.4	Tests of hypotheses	277
9.4.1	Exercise	281
10	Estimation for complete data	283
10.1	Introduction	283
10.2	The empirical distribution for complete, individual data	288
10.2.1	Exercises	291

10.3 Empirical distributions for grouped data	292	12.5.2 Negative binomial	386
10.3.1 Exercises	295	12.5.3 Binomial	389
11 Estimation for modified data	297	12.5.4 The $(a, b, 1)$ class	392
11.1 Point estimation	297	12.5.5 Compound models	396
11.1.1 Exercises	304	12.5.6 Effect of exposure on maximum likelihood estimation	398
11.2 Means, variances, and interval estimation	305	12.5.7 Exercises	399
11.2.1 Exercises	315	12.6 Bivariate models	402
11.3 Kernel density models	316	12.6.1 Introduction	402
11.3.1 Exercises	321	12.6.2 Copulas	403
11.4 Approximations for large data sets	322	12.6.3 Exercise	405
11.4.1 Introduction	322	12.7 Models with covariates	405
11.4.2 Kaplan-Meier type approximations	323	12.7.1 Introduction	405
11.4.3 Multiple-decrement tables	324	12.7.2 Proportional hazards models	407
11.4.4 Exercises	325	12.7.3 The generalized linear and accelerated failure time models	413
		12.7.4 Exercises	416
Part IV Parametric statistical methods			
12 Parameter estimation	331	13 Model selection	419
12.1 Method of moments and percentile matching	331	13.1 Introduction	419
12.1.1 Exercises	335	13.2 Representations of the data and model	420
12.2 Maximum likelihood estimation	337	13.3 Graphical comparison of the density and distribution functions	421
12.2.1 Introduction	337	13.3.1 Exercises	426
12.2.2 Complete, individual data	340	13.4 Hypothesis tests	427
12.2.3 Complete, grouped data	341	13.4.1 Kolmogorov-Smirnov test	428
12.2.4 Truncated or censored data	341	13.4.2 Anderson-Darling test	430
12.2.5 Exercises	345	13.4.3 Chi-square goodness-of-fit test	432
12.3 Variance and interval estimation	351	13.4.4 Likelihood ratio test	436
12.3.1 Exercises	358	13.4.5 Exercises	438
12.4 Bayesian estimation	360	13.5 Selecting a model	440
12.4.1 Definitions and Bayes' theorem	360	13.5.1 Introduction	440
12.4.2 Inference and prediction	364	13.5.2 Judgment-based approaches	441
12.4.3 Conjugate prior distributions and the linear exponential family	371	13.5.3 Score-based approaches	442
12.4.4 Computational issues	374	13.5.4 Exercises	449
12.4.5 Exercises	377		
12.5 Estimation for discrete distributions	383	14 Five examples	455
12.5.1 Poisson	383	14.1 Introduction	455
		14.2 Time to death	455

14.2.1 The data	455
14.2.2 Some calculations	457
14.2.3 Exercise	459
14.3 Time from incidence to report	459
14.3.1 The problem and some data	460
14.3.2 Analysis	460
14.4 Payment amount	461
14.4.1 The data	462
14.4.2 The first model	463
14.4.3 The second model	465
14.5 An aggregate loss example	466
14.6 Another aggregate loss example	471
14.6.1 Distribution for a single policy	471
14.6.2 One hundred policies—excess of loss	471
14.6.3 One hundred policies—aggregate stop-loss	473
14.6.4 Numerical convolutions	474
14.7 Comprehensive exercises	476
Part V Adjusted estimates and simulation	
15 Interpolation and smoothing	483
15.1 Introduction	483
15.2 Polynomial interpolation and smoothing	485
15.2.1 Exercises	488
15.3 Cubic spline interpolation	489
15.3.1 Construction of cubic splines	491
15.3.2 Exercises	499
15.4 Approximating functions with splines	500
15.4.1 Exercise	504
15.5 Extrapolating with splines	504
15.5.1 Exercise	505
15.6 Smoothing splines	505
15.6.1 Exercise	514
16 Credibility	515
16.1 Introduction	515
16.2 Statistical concepts	517

16.2.1 Conditional distributions	518
16.2.2 Conditional expectation	520
16.2.3 Nonparametric unbiased estimators	523
16.2.4 Exercises	529
16.3 Limited fluctuation credibility theory	530
16.3.1 Full credibility	532
16.3.2 Partial credibility	535
16.3.3 Problems with the approach	539
16.3.4 Notes and References	539
16.3.5 Exercises	539
16.4 Greatest accuracy credibility theory	542
16.4.1 Introduction	542
16.4.2 The Bayesian methodology	545
16.4.3 The credibility premium	553
16.4.4 The Buhlmann model	557
16.4.5 The Buhlmann–Straub model	560
16.4.6 Exact credibility	566
16.4.7 Linear versus Bayesian versus no credibility	569
16.4.8 Notes and References	577
16.4.9 Exercises	578
16.5 Empirical Bayes parameter estimation	589
16.5.1 Nonparametric estimation	592
16.5.2 Semiparametric estimation	600
16.5.3 Parametric estimation	602
16.5.4 Notes and References	607
16.5.5 Exercises	607
17 Simulation	611
17.1 Basics of simulation	611
17.1.1 The simulation approach	612
17.1.2 Exercises	617
17.2 Examples of simulation in actuarial modeling	618
17.2.1 Aggregate loss calculations	618
17.2.2 Examples of lack of independence or identical distributions	619
17.2.3 Simulation analysis of the two examples	620
17.2.4 Statistical analyses	622
17.2.5 Exercises	625

<i>Appendix A An inventory of continuous distributions</i>	627
A.1 <i>Introduction</i>	627
A.2 <i>Transformed beta family</i>	631
A.2.1 <i>Four-parameter distribution</i>	631
A.2.2 <i>Three-parameter distributions</i>	631
A.2.3 <i>Two-parameter distributions</i>	633
A.3 <i>Transformed gamma family</i>	635
A.3.1 <i>Three-parameter distributions</i>	635
A.3.2 <i>Two-parameter distributions</i>	636
A.3.3 <i>One-parameter distributions</i>	638
A.4 <i>Other distributions</i>	638
A.5 <i>Distributions with finite support</i>	640
 <i>Appendix B An inventory of discrete distributions</i>	643
B.1 <i>Introduction</i>	643
B.2 <i>The (a, b, 0) class</i>	644
B.3 <i>The (a, b, 1) class</i>	645
B.3.1 <i>The zero-truncated subclass</i>	645
B.3.2 <i>The zero-modified subclass</i>	647
B.4 <i>The compound class</i>	648
B.4.1 <i>Some compound distributions</i>	648
B.5 <i>A hierarchy of discrete distributions</i>	650
 <i>Appendix C Frequency and severity relationships</i>	651
 <i>Appendix D The recursive formula</i>	653
 <i>Appendix E Discretization of the severity distribution</i>	655
E.1 <i>The method of rounding</i>	655
E.2 <i>Mean preserving</i>	656
E.3 <i>Undiscretization of a discretized distribution</i>	657
 <i>Appendix F Numerical optimization and solution of systems of equations</i>	659
F.1 <i>Maximization using Solver</i>	660
F.2 <i>The simplex method</i>	664
F.3 <i>Using Excel® to solve equations</i>	665

References

671

Index

681

Preface

The preface to the first edition of this text explained our mission as follows:

This textbook is organized around the principle that much of actuarial science consists of the construction and analysis of mathematical models which describe the process by which funds flow into and out of an insurance system. An analysis of the entire system is beyond the scope of a single text, so we have concentrated our efforts on the loss process, that is, the outflow of cash due to the payment of benefits.

We have not assumed that the reader has any substantial knowledge of insurance systems. Insurance terms are defined when they are first used. In fact, most of the material could be disassociated from the insurance process altogether, and this book could be just another applied statistics text. What we have done is kept the examples focused on insurance, presented the material in the language and context of insurance, and tried to avoid getting into statistical methods that would have little use in actuarial practice.

In particular, the first edition of this text was published in 1998 to achieve three goals:

1. Update the distribution fitting material from *Loss Distributions* [59] by Robert Hogg and Stuart Klugman, published in 1984.
2. Update material on discrete distributions and collective risk model calculations from *Insurance Risk Models* [106] by Harry Panjer and Gordon Willmot, published in 1992.

3. Integrate the material the three authors had developed for the Society of Actuaries' Intensive Seminar 152, Applied Risk Theory.

Shortly after publication, the Casualty Actuarial Society and the Society of Actuaries altered their examination syllabus to include our first edition. The good news was that the first edition was selected as source material for the new third and fourth examinations. The bad news was that the subject matter was split between the examinations in a manner that was not compatible with the organization of the text. By itself, that is sufficient reason to produce a revision. But there are others.

1. The first edition was written with an assumption that readers would be familiar with the subject of mathematical statistics. This had been part of the actuarial examination process at the time the book was written but was subsequently removed. Some background material on mathematical statistics is now presented in Chapter 9.
2. For a long time, actuarial education has included the subject of survival models. This is the study of determining probability models for time to death, failure, or disability. It is not much different from the study of determining probability models for the amount or number of claims. This edition integrates that subject and in doing so adds an emphasis on building empirical models. This is covered in Chapters 10 and 11.
3. There were two items that had been removed from the actuarial syllabus over the years that we wanted to see returned, at least in brief. One is graduation, the smoothing of and interpolation of sequences of numbers. This is covered in Chapter 15. The other is the adjustment of estimation formulas when dealing with the large amounts of data in mortality studies. This is covered in Section 11.4.
4. While simulation was briefly covered in the first edition, the material has been slightly expanded and now appears in Chapter 17.

With regard to continuing material, besides the rearrangement of the material on models and modeling, other substantive changes are a significant rewrite of the ruin theory material (Chapters 7 and 8), and a better explanation of the limited fluctuation credibility formulas (Chapter 16).

While we have attempted to integrate the material into a single, logical development of actuarial model building, various sections stand alone, thus the division of the text into various parts.

Since the publication of the first edition, computational power continues to increase. For that edition, specialized DOS programs were made available for obtaining maximum likelihood estimates and performing aggregate loss calculations. Those programs continue to be available at the Wiley ftp site:

ftp://ftp.wiley.com/public/sci_tech_med/loss_models/

In addition, files containing the data for examples and exercises are also available. However, it is more likely that users will be calculating using a spreadsheet program such as Microsoft Excel®.¹ At various places in the text we indicate how Excel® commands may help. This is not an endorsement by the authors, but rather a recognition of the pervasiveness of this tool.

As in the first edition, many of the exercises are taken from examinations of the Casualty Actuarial Society and the Society of Actuaries. They have been reworded to fit the terminology and notation of this text and the five answer choices from the original questions are not provided. Such exercises are indicated with an asterisk (*). Of course, these questions may not be representative of those asked on examinations given in the future.

Finally, a word about our cover picture. In the summer of 1993 the Des Moines and Raccoon Rivers flooded, putting sections of Des Moines, Iowa under water. At the left center of the picture is the Des Moines Water Works. Contamination knocked out the water supply for 12 days. While living through adverse events can be interesting, it is not necessary to do so to build probability models for their occurrence, timing, and severity. Our thanks to Melissa Sharer of the Des Moines Water Works for providing the picture.

S. A. KLUGMAN
H. H. PANJER
G. E. WILLMOT

*Des Moines, Iowa
Waterloo, Ontario*

¹Microsoft® and Excel® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Acknowledgments

Stuart Klugman:

While producing a second edition is not as daunting or grueling a task as creating the first edition, there was still a lot to be done and several people who helped make it possible. While many individuals provided corrections and suggestions to improve the first edition, Elias Shiu and Don Minassian were especially prolific in pointing out our shortcomings. Clive Keatinge provided many corrections to this text prior to publication. When the actuarial societies determined that the presentation needed to be altered to conform with the allocation of material to the two examinations, a committee was convened under the leadership of Clive Keatinge. This group laid out the plan that is reflected in the presentation in Chapters 2–5 and 9–13. We have added a few subjects to make the presentation more complete but have relied on the committee's expertise to set the agenda for these chapters.

Special thanks to my coauthors for meeting the deadline.

I am revising this acknowledgement shortly after my wife Marie died from a long struggle with multiple diseases. Both when healthy and throughout her illness she was an inspiration to me and all who knew her. The final draft was submitted to the publisher on time not because I was excited to be working but because meeting commitments and keeping promises was something she valued.

Harry Panjer:

Revising and editing a book is not much fun. However, in doing this second edition, we did have some fun in adding a lot of new material. Each coauthor made major contributions to several chapters of completely new material. I particularly enjoyed drafting the chapter on spline interpolation and smoothing which provides a modern treatment of the classical actuarial problem of graduation. I am indebted to Stuart Klugman and students at Drake University who field-tested this material and discovered several errors (thus saving me potential future embarrassment).

Stuart Klugman provided the visionary, spiritual, and practical leadership for this project. I am also indebted to Gordon Willmot who led the proof-reading team at the University of Waterloo. Whatever errors remain are a tiny fraction of the total from the first draft. They did a fantastic job. And finally, thanks to my wife Joanne Coyle who tolerated my many weekends and evenings at the office. Our sons Lucas D. and Lucas R., who saw the first edition being written, have grown up and moved away since then. They will only discover this second edition when someone points out to them that they are mentioned in this acknowledgement.

Gordon Willmot:

The second edition has involved relatively substantial reorganization and updating. Many of these changes are a reflection of the somewhat novel nature of the combination of topics that are presented in order to attempt to provide an integrated approach to the use of data to calibrate models and the subsequent use of these models for pricing and related purposes. As such, the present treatment has undoubtedly benefitted from the input of many individuals who are clearly too numerous to mention. Special thanks, however, do go to Catherine Donnelly, Steve Drekić, Mary Lou Dufton, Jessica Ling-Wai Lam, and Claire Xiao-Dan Yang for their invaluable assistance on this second edition.

I also wish to express my sincere gratitude to my coauthors, and in particular to Stuart Klugman for spearheading this project, keeping us on track while leading by example, and for his nearly infinite patience in dealing with all related issues in spite of difficult circumstances. I also wish to again express my thanks to my wife Deborah and my daughters Rachel, Lauren, and Kristen for both the implicit and explicit sacrifices that they have made in order that I devote time and energy to this work.

S.A.K., H.H.P., G.E.W.

Part I

Introduction

1

Modeling

1.1 THE MODEL-BASED APPROACH

The model-based approach should be considered in the context of the objectives of any given problem. Many problems in actuarial science involve the building of a mathematical model that can be used to forecast or predict insurance costs in the future.

A model is a simplified mathematical description which is constructed based on the knowledge and experience of the actuary combined with data from the past. The data guide the actuary in selecting the form of the model as well as in calibrating unknown quantities, usually called **parameters**. The model provides a balance between simplicity and conformity to the available data.

The simplicity is measured in terms of such things as the number of unknown parameters (the fewer the simpler); the conformity to data is measured in terms of the discrepancy between the data and the model. Model selection is based on a balance between the two criteria, namely, fit and simplicity.

1.1.1 The modeling process

The modeling process is illustrated in Figure 1.1, which describes six stages.

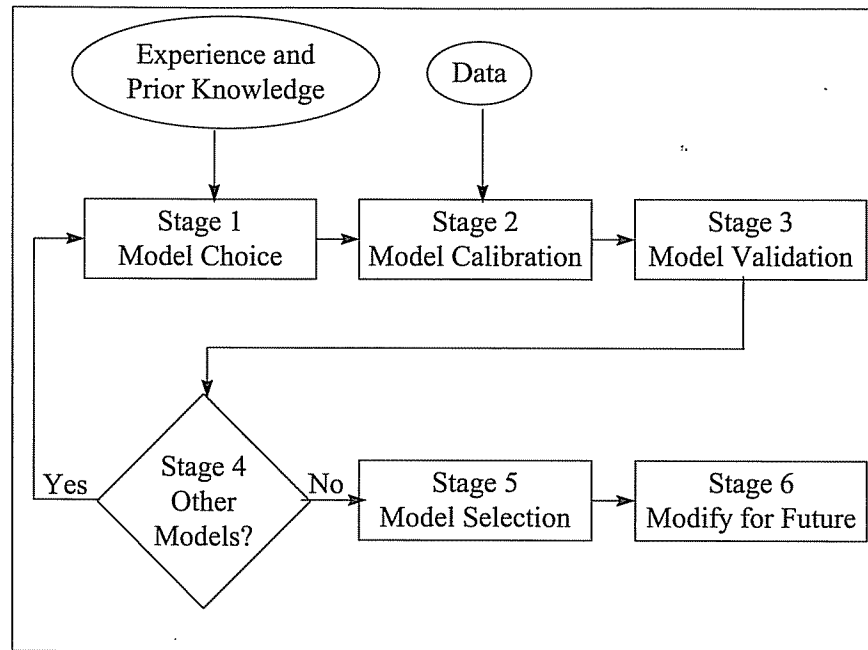


Fig. 1.1 The modeling process.

Stage 1 One or more models are selected based on the analyst's prior knowledge and experience and possibly on the nature and form of available data. For example, in studies of mortality, models may contain covariate information such as age, sex, duration, policy type, medical information, and lifestyle variables. In studies of the size of insurance loss, a statistical distribution (e.g., lognormal, gamma, or Weibull) may be chosen.

Stage 2 The model is calibrated based on available data. In mortality studies, these data may be information on a set of life insurance policies. In studies of property claims, the data may be information about each of a set of actual insurance losses paid under a set of property insurance policies.

Stage 3 The fitted model is validated to determine if it adequately conforms to the data. Various diagnostic tests can be used. These may be well-known statistical tests, such as the chi-square goodness-of-fit test or the Kolmogorov-Smirnov test, or may be more qualitative in nature. The choice of test may relate directly to the ultimate purpose of the modeling exercise. In insurance-related studies, the fitted model is often required to replicate in total the losses actually experienced in the data. In insurance practice this is often referred to as **unbiasedness** of a model.

Stage 4 An opportunity is provided to consider other possible models. This is particularly useful if Stage 3 revealed that all models were inadequate. It is also possible that more than one valid model will be under consideration at this stage.

Stage 5 All valid models considered in Stages 1-4 are compared using some criteria to select between them. This may be done by using the test results previously obtained or may be done by using another criterion. Once a winner is selected, the losers may be retained for sensitivity analyses.

Stage 6 Finally, the selected model is adapted for application to the future. This could involve adjustment of parameters to reflect anticipated inflation from the time the data were collected to the period of time to which the model will be applied.

As new data are collected or the environment changes, the six stages will need to be repeated to improve the model.

1.1.2 The modeling advantage

Determination of the advantages of using models requires us to consider the alternative, decision making based strictly upon empirical evidence. The empirical approach assumes that the future can be expected to be exactly like a sample from the past, perhaps adjusted for trends such as inflation. Consider the following illustration.

Example 1.1 A portfolio of group life insurance certificates consists of 1,000 employees of various ages and death benefits. Over the past five years, 14 employees died and received a total of 580,000 in benefits (adjusted for inflation because the plan relates benefits to salary). Determine the empirical estimate of next year's expected benefit payment.

The empirical estimate for next year is then 116,000 (one-fifth of the total), which would need to be further adjusted for benefit increases. The danger, of course, is that it is unlikely that the experience of the past five years accurately reflects the future of this portfolio as there can be considerable fluctuation in such short-term results. □

It seems much more reasonable to build a model, in this case a mortality table. This table would be based on the experience of many lives, not just the 1,000 in our group. With this model we not only can estimate the expected payment for next year, but we can also measure the risk involved by calculating the standard deviation of payments or perhaps various percentiles from the distribution of payments. This is precisely the problem covered in great detail in *Actuarial Mathematics* [16].

This approach was codified by the Society of Actuaries Committee on Actuarial Principles. In the publication "Principles of Actuarial Science" [123], p. 571, Principle 3.1 states that "Actuarial risks can be stochastically modeled based on assumptions regarding the probabilities that will apply to the actuarial risk variables in the future, including assumptions regarding the future environment." The actuarial risk variables referred to are occurrence, timing, and severity—that is, the chances of a claim event, the time at which the event occurs if it does, and the cost of settling the claim.

1.2 ORGANIZATION OF THIS BOOK

This text takes the reader through the modeling process, but not in the order presented in the previous section. There is a difference between how models are best applied and how they are best learned. In this text we first learn about the models and how to use them. This is followed by instruction in how to determine which model to use. The reason is that it is difficult to select models in a vacuum. Unless the analyst has a thorough knowledge of the set of available models, it is difficult to narrow the choice to the ones worth considering. With that in mind, the organization of the text is as follows:

1. Review of probability—Almost by definition, contingent events imply probability models. Chapters 2 and 3 review random variables and some of the basic calculations that may be done with such models, including moments and percentiles.
2. Understanding probability distributions—In order to select a probability model, the analyst should possess a reasonably large collection of such models. In addition, in order to make a good a priori model choice, characteristics of these models should be available. In Chapter 4 a variety of distributional models are introduced and their characteristics explored. This includes both continuous and discrete distributions.
3. Coverage modifications—Insurance contracts often do not provide full payment. For example, there may be a deductible (the insurance policy does not pay the first \$250, for example) or a limit (the insurance policy does not pay more than \$10,000 for any one loss event, for example). Such modifications alter the probability distribution and affect related calculations such as moments. Chapter 5 shows how this is done.
4. Aggregate losses and ruin—To this point the models are either for the amount of a single payment or for the number of payments. Of interest when modeling a portfolio, line of business, or entire company is the total amount paid. A model that combines the probabilities concerning the number of payments and the amounts of each payment is called an aggregate loss model. Calculations for such models are covered in

Chapter 6. Usually, the payments arrive sequentially through time. It is possible, if the payments turn out to be large, that at some point the entity will run out of money. This state of affairs is called ruin. In Chapters 7 and 8 models are established that allow for the calculation of the probability this will happen.

5. Review of mathematical statistics—Because most of the models being considered are probability models, techniques of mathematical statistics are needed to estimate model specifications and make choices. While Chapter 9 is not a replacement for a thorough text or course in mathematical statistics, it does contain the essential items needed later in this text.
6. Construction of empirical models—Sometimes it is appropriate to work with the empirical distribution of the data. It may be because the volume of data is sufficient or because a good portrait of the data is needed. Chapters 10 and 11 cover this for the simple case of straightforward data, adjustments for truncated and censored data, and modifications suitable for large data sets, particularly those encountered in mortality studies.
7. Construction of parametric models—Often it is valuable to smooth the data and thus represent the population by a probability distribution. Chapter 12 provides methods for parameter estimation for the models introduced earlier. Model selection is covered in Chapter 13.
8. Chapter 14 contains examples that summarize and integrate the topics discussed to this point.
9. Adjustment of estimates—At times, further adjustment of the results is needed. Two such adjustments are covered in this text. The first is interpolation and smoothing (also called graduation) and is covered in Chapter 15, where the emphasis is on cubic splines. There are situations, such as time to death, where no simple probability distribution is known to describe the observations. An empirical approach is likely to produce results that are not as smooth as the population is known to be. Graduation methods can provide the needed adjustment. A second situation occurs when there are one or more estimates based on a small number of observations. Accuracy could be improved by adding other, related observations, but care must be taken if the additional data are from a different population. Credibility methods, covered in Chapter 16, provide a mechanism for making the appropriate adjustment when additional data are to be included.
10. Simulation—When analytical results are difficult to obtain, simulation (use of random numbers) may provide the needed answer. A brief introduction to this technique is provided in Chapter 17.

Part II

Actuarial models

2

Random variables

2.1 INTRODUCTION

An actuarial model is an attempt to represent an uncertain stream of future payments. The uncertainty may be with respect to any or all of occurrence (is there a payment?), timing (when is the payment made?), and severity (how much is paid?). Because the most useful means of representing uncertainty is through probability, we concentrate on probability models. In all cases, the relevant probability distributions are assumed to be known. Determining appropriate distributions is covered in Chapters 10–13. In this part, the following aspects of actuarial probability models will be covered:

1. Definition of random variable, important functions, and some examples.
2. Basic calculations from probability models.
3. Specific probability distributions and their properties.
4. More advanced calculations using severity models.
5. Models incorporating the possibility of a random number of payments each of random amount.
6. Models that track a company's surplus through time.

Loss Models: From Data to Decisions, Second Edition.

By Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot
ISBN 0-471-21577-5 Copyright © 2004 John Wiley & Sons, Inc.

There are two important models that are absent from this text. The first is a model for investment earnings in future years. While such techniques are within the scope of this text, the additional finance background needed to motivate these models would detract from the primary purpose of this text. The second is a model that combines the earning of interest (whether random or not) with the timing of the payment. While simple models of this type may show up in examples, thorough coverage of the use of these models for life insurance and annuities is sufficiently specialized to require a separate text, such as [16].

The commonality we seek here is that all models for random phenomena have similar elements. For each, there is a set of possible outcomes. The particular outcome that occurs will determine the success of our enterprise. Attaching probabilities to the various outcomes allows us to quantify our expectations and the risk of not meeting them. In this spirit, the underlying random variable will almost always be denoted X or Y . The context will provide a name and some likely characteristics. Of course, there are actuarial models that do not look like those covered here. For example, in life insurance a *model office* is a list of cells containing policy type, age range, gender, and so on.

To expand on this concept, consider the following definitions from the latest working draft of "Joint Principles of Actuarial Science"¹:

Phenomena are occurrences that can be observed. An *experiment* is an observation of a given phenomenon under specified conditions. The result of an experiment is called an *outcome*; an *event* is a set of one or more possible outcomes. A stochastic phenomenon is a phenomenon for which an associated experiment has more than one possible outcome. An event associated with a *stochastic phenomenon* is said to be *contingent*. *Probability* is a measure of the likelihood of the occurrence of an event. It is measured on a scale of increasing likelihood from zero to one. A *random variable* is a function that assigns a numerical value to every possible outcome.

The following list contains a number of random variables encountered in actuarial work:

1. The age at death of a randomly selected birth. (Model 1)
2. The time to death from when insurance was purchased for a randomly selected insured life.
3. The time from occurrence of a disabling event to recovery or death for a randomly selected workers compensation claimant.

¹This document is a work in progress of a joint committee from the Casualty Actuarial Society and the Society of Actuaries. Key principles are that models exist that represent actuarial phenomena and that given sufficient data it is possible to calibrate models.

4. The time from the incidence of a randomly selected claim to its being reported to the insurer.
5. The time from the reporting of a randomly selected claim to its settlement.
6. The number of dollars paid on a randomly selected life insurance claim.
7. The number of dollars paid on a randomly selected automobile bodily injury claim. (Model 2)
8. The number of automobile bodily injury claims in one year from a randomly selected insured automobile. (Model 3)
9. The total dollars in medical malpractice claims paid in one year owing to events at a randomly selected hospital. (Model 4)
10. The time to default or prepayment on a randomly selected insured home loan that terminates early.
11. The amount of money paid at maturity on a randomly selected high-yield bond.

Because all of these phenomena can be expressed as random variables, the machinery of probability and mathematical statistics is at our disposal both to create and to analyze models for them. The following paragraphs discuss the five key functions used in describing a random variable. They will be illustrated with four ongoing models as identified in the list above plus two more to be introduced later.

2.2 KEY FUNCTIONS AND FOUR MODELS

Definition 2.1 *The cumulative distribution function, also called the distribution function and usually denoted $F_X(x)$ or $F(x)$,² for a random variable X is the probability that X is less than or equal to a given number. That is, $F_X(x) = \Pr(X \leq x)$. The abbreviation *cdf* is often used.*

The distribution function must satisfy a number of requirements³:

- $0 \leq F(x) \leq 1$ for all x .

²When denoting functions associated with random variables, it is common to identify the random variable through a subscript on the function. Here, subscripts will be used only when needed to distinguish one random variable from another. In addition, for the six models to be introduced shortly, rather than write the distribution function for random variable 2 as $F_{X_2}(x)$, it will simply be denoted $F_2(x)$.

³The first point follows from the last three.

- $F(x)$ is nondecreasing.
- $F(x)$ is right-continuous.⁴
- $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.

Because it need not be left-continuous, it is possible for the distribution function to jump. When it jumps, the value is assigned to the top of the jump.

Here are possible distribution functions for each of the four models.

Model 1⁵ This random variable could serve as a model for the age at death. All ages between 0 and 100 are possible. While experience suggests that there is an upper bound for human lifetime, models with no upper limit may be useful if they assign extremely low probabilities to extreme ages. This allows the modeler to avoid setting a specific maximum age.

$$F_1(x) = \begin{cases} 0, & x < 0, \\ 0.01x, & 0 \leq x < 100, \\ 1, & x \geq 100. \end{cases}$$

□

Model 2 This random variable could serve as a model for the number of dollars paid on an automobile insurance claim. All positive values are possible. As with mortality, there is more than likely an upper limit (all the money in the world comes to mind), but this model illustrates that in modeling, correspondence to reality need not be perfect.

$$F_2(x) = \begin{cases} 0, & x < 0, \\ 1 - \left(\frac{2000}{x + 2000} \right)^3, & x \geq 0. \end{cases}$$

□

Example 2.2 Draw graphs of the distribution function for Models 1 and 2 (graphs for the other models are requested in Exercise 2.2).

The graphs appear in Figures 2.1 and 2.2.

□

Model 3 This random variable could serve as a model for the number of claims on one policy in one year. Probability is concentrated at the five points (0,1,2,3,4) and the probability at each is given by the size of the jump in the distribution function. While this model places a maximum on the number of claims, models with no limit (such as the Poisson distribution) could also be used.

⁴Right-continuous means that at any point x_0 the limiting value of $F(x)$ as x approaches x_0 from the right is equal to $F(x_0)$. This need not be true as x approaches x_0 from the left.

⁵The six models (four introduced here and two later) will be identified by the numbers 1-6. Other examples will use the traditional numbering scheme as used for Definitions, etc.

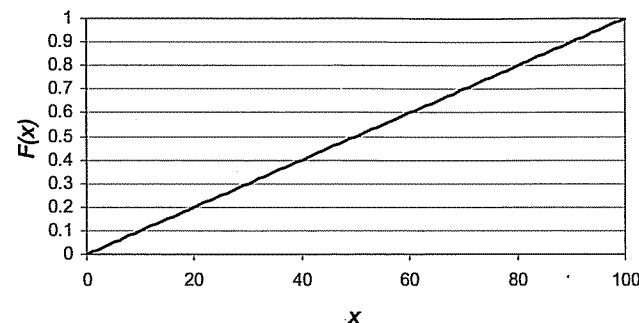


Fig. 2.1 Distribution function for Model 1.

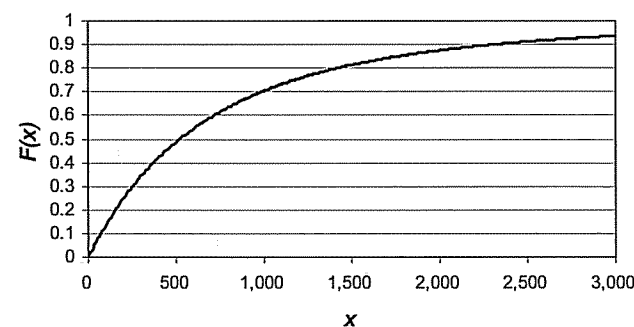


Fig. 2.2 Distribution function for Model 2.

$$F_3(x) = \begin{cases} 0, & x < 0, \\ 0.5, & 0 \leq x < 1, \\ 0.75, & 1 \leq x < 2, \\ 0.87, & 2 \leq x < 3, \\ 0.95, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

□

Model 4 This random variable could serve as a model for the total dollars paid on a malpractice policy in one year. Most of the probability is at zero (0.7) because in most years nothing is paid. The remaining 0.3 of probability is distributed over positive values.

$$F_4(x) = \begin{cases} 0, & x < 0, \\ 1 - 0.3e^{-0.00001x}, & x \geq 0. \end{cases} \quad \square$$

Definition 2.3 The support of a random variable is the set of numbers that are possible values of the random variable.

Definition 2.4 A random variable is called *discrete* if the support contains at most a countable number of values. It is called *continuous* if the distribution function is continuous and is differentiable everywhere with the possible exception of a countable number of values. It is called *mixed* if it is not discrete and is continuous everywhere with the exception of at least one value and at most a countable number of values.

These three definitions do not exhaust all possible random variables but will cover all cases encountered in this text. The distribution function for a discrete variable will be constant except for jumps at the values with positive probability. A mixed distribution will have at least one jump. Requiring continuous variables to be differentiable allows the variable to have a density function (defined later) at almost all values.

Example 2.5 For each of the four models, determine the support and indicate which type of random variable it is.

The distribution function for Model 1 is continuous and is differentiable except at 0 and 100 and therefore is a continuous distribution. The support is values from 0 to 100 with it not being clear if 0 or 100 are included. The distribution function for Model 2 is continuous and is differentiable except at 0 and therefore is a continuous distribution. The support is all positive numbers and perhaps 0. The random variable for Model 3 places probability only at 0, 1, 2, 3, and 4 (the support) and thus is discrete. The distribution function for Model 4 is continuous except at 0, where it jumps. It is a mixed distribution with support on nonnegative numbers. \square

These four models illustrate the most commonly encountered forms of the distribution function. For the remainder of this text, values of functions like the distribution function will be presented only for values in the range of the support of the random variable.

Definition 2.6 The survival function, usually denoted $S_X(x)$ or $S(x)$, for a random variable X is the probability that X is greater than a given number. That is, $S_X(x) = \Pr(X > x) = 1 - F_X(x)$.

As a result:

- $0 \leq S(x) \leq 1$ for all x .

- $S(x)$ is nonincreasing.
- $S(x)$ is right-continuous.
- $\lim_{x \rightarrow -\infty} S(x) = 1$ and $\lim_{x \rightarrow \infty} S(x) = 0$.

Because the survival function need not be left-continuous, it is possible for it to jump (down). When it jumps, the value is assigned to the bottom of the jump.

Because the survival function is the complement of the distribution function, knowledge of one implies knowledge of the other. Historically, when the random variable is measuring time, the survival function is presented, while when it is measuring dollars, the distribution function is presented.

Example 2.7 For completeness, here are the survival functions for the four models.

$$S_1(x) = 1 - 0.01x, \quad 0 \leq x < 100,$$

$$S_2(x) = \left(\frac{2,000}{x + 2,000} \right)^3, \quad x \geq 0,$$

$$S_3(x) = \begin{cases} 0.5, & 0 \leq x < 1, \\ 0.25, & 1 \leq x < 2, \\ 0.13, & 2 \leq x < 3, \\ 0.05, & 3 \leq x < 4, \\ 0, & x \geq 4 \end{cases}$$

$$S_4(x) = 0.3e^{-0.00001x}, \quad x \geq 0. \quad \square$$

Example 2.8 Graph the survival function for Models 1 and 2.

The graphs appear in Figures 2.3 and 2.4. \square

Either the distribution or survival function can be used to determine probabilities. Let $F(b-) = \lim_{x \nearrow b} F(x)$ and let $S(b-)$ be similarly defined. That is, we want the limit as x approaches b from below. We have $\Pr(a < X \leq b) = F(b) - F(a) = S(a) - S(b)$ and $\Pr(X = b) = F(b) - F(b-) = S(b-) - S(b)$. When the distribution function is continuous at x , $\Pr(X = x) = 0$; otherwise the probability is the size of the jump. The next two functions are more directly related to the probabilities. The first is for continuous distributions, the second for discrete distributions.

Definition 2.9 The probability density function, also called the density function, usually denoted $f_X(x)$ or $f(x)$, is the derivative of the distribution function or, equivalently, the negative of the derivative of the survival function.

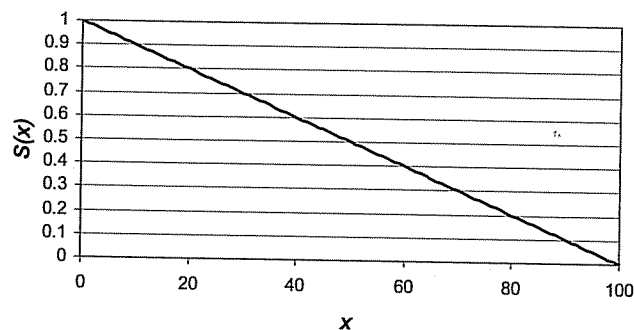


Fig. 2.3 Survival function for Model 1.

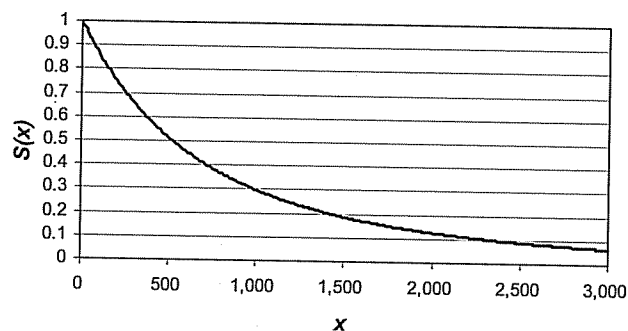


Fig. 2.4 Survival function for Model 2.

That is, $f(x) = F'(x) = -S'(x)$. The density function is defined only at those points where the derivative exists. The abbreviation *pdf* is often used.

While the density function does not directly provide probabilities, it does provide relevant information. Values of the random variable in regions with higher density values are more likely to occur than those in regions with lower values. Probabilities for intervals and the distribution and survival functions can be recovered by integration. That is, when the density function is defined over the relevant interval, $\Pr(a < X \leq b) = \int_a^b f(x)dx$, $F(b) = \int_{-\infty}^b f(x)dx$, and $S(b) = \int_b^{\infty} f(x)dx$.

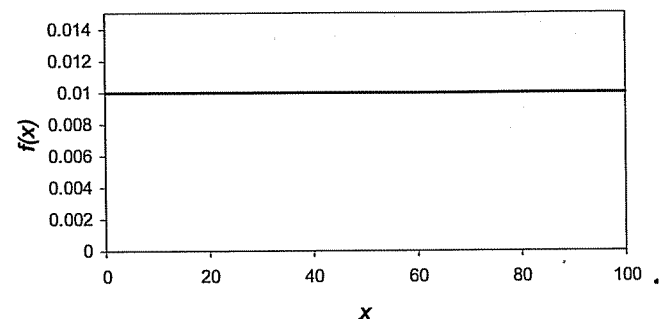


Fig. 2.5 Density function for Model 1.

Example 2.10 For our models,

$$f_1(x) = 0.01, \quad 0 < x < 100,$$

$$f_2(x) = \frac{3(2,000)^3}{(x + 2,000)^4}, \quad x > 0,$$

$$f_3(x) \text{ is not defined,}$$

$$f_4(x) = 0.000003e^{-0.00001x}, \quad x > 0.$$

It should be noted that for Model 4 the density function does not completely describe the probability distribution. As a mixed distribution, there is also discrete probability at 0. \square

Example 2.11 Graph the density function for Models 1 and 2.

The graphs appear in Figures 2.5 and 2.6. \square

Definition 2.12 The *probability function*, also called the *probability mass function*, usually denoted $p_X(x)$ or $p(x)$, describes the probability at a distinct point when it is not 0. The formal definition is $p_X(x) = \Pr(X = x)$.

For discrete random variables, the distribution and survival functions can be recovered as $F(x) = \sum_{y \leq x} p(y)$ and $S(x) = \sum_{y > x} p(y)$.

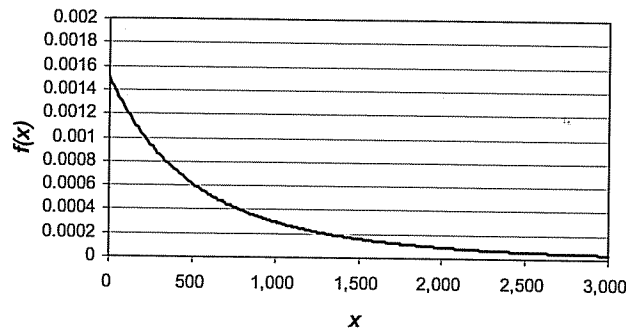


Fig. 2.6 Density function for Model 2.

Example 2.13 For our models,

$p_1(x)$ is not defined,

$p_2(x)$ is not defined,

$$p_3(x) = \begin{cases} 0.50, & x = 0, \\ 0.25, & x = 1, \\ 0.12, & x = 2, \\ 0.08, & x = 3, \\ 0.05, & x = 4, \end{cases}$$

$$p_4(0) = 0.7.$$

It is again noted that the distribution in Model 4 is mixed, so the above describes only the discrete portion of that distribution. There is no easy way to present probabilities/densities for a mixed distribution. On the other hand, they tend to be more revealing of the mixed nature of the distribution. For Model 4 we would present the probability density function as

$$f_4(x) = \begin{cases} 0.7, & x = 0, \\ 0.000003e^{-0.00001x}, & x > 0, \end{cases}$$

realizing that, technically, it is not a probability density function at all. When the density function is assigned a value at a specific point, as opposed to being defined on an interval, it is understood to be a discrete probability mass. \square

Definition 2.14 The hazard rate, also known as the force of mortality and the failure rate and usually denoted $h_X(x)$ or $h(x)$, is the ratio of the density and survival functions when the density function is defined. That is, $h_X(x) = f_X(x)/S_X(x)$.

When called the force of mortality, the hazard rate is often denoted $\mu(x)$, and when called the failure rate, it is often denoted $\lambda(x)$. Regardless, it may be interpreted as the probability density at x given that the argument will be at least x . We also have $h_X(x) = -S'(x)/S(x) = -d \ln S(x)/dx$. The survival function can be recovered from $S(b) = e^{-\int_0^b h(x)dx}$. Though not necessary, this formula implies that the support is on nonnegative numbers. In mortality terms, the force of mortality is the annualized probability⁶ that a person age x will die in the next instant, expressed as a death rate per year. In this text we will always use $h(x)$ to denote the hazard rate, although one of the alternative names may be used.

Example 2.15 For our models,

$$h_1(x) = \frac{0.01}{1 - 0.01x}, \quad 0 < x < 100,$$

$$h_2(x) = \frac{3}{x + 2,000}, \quad x > 0,$$

$h_3(x)$ is not defined,

$$h_4(x) = 0.00001, \quad x > 0.$$

Once again, note that for the mixed distribution the hazard rate is only defined over part of the random variable's support. This is different from the problem above where both a probability density function and a probability function are involved. Where there is a discrete probability mass, the hazard rate is not defined. \square

Example 2.16 Graph the hazard rate function for Models 1 and 2.

The graphs appear in Figures 2.7 and 2.8. \square

The following model illustrates a situation in which there is a point where the density and hazard rate functions are not defined.

Model 5 An alternative to the simple lifetime distribution in Model 1 is given below. Note that it is piecewise linear and the derivative at 50 is not defined. Therefore, neither the density function nor the hazard rate function is defined at 50. Unlike the mixed model of Model 4, there is no discrete probability mass at this point. Because the probability of 50 occurring is zero, the density or hazard rate at 50 could be arbitrarily defined with no effect on subsequent

⁶Note that the force of mortality is not a probability (in particular, it can be greater than 1) although it does no harm to visualize it as a probability.

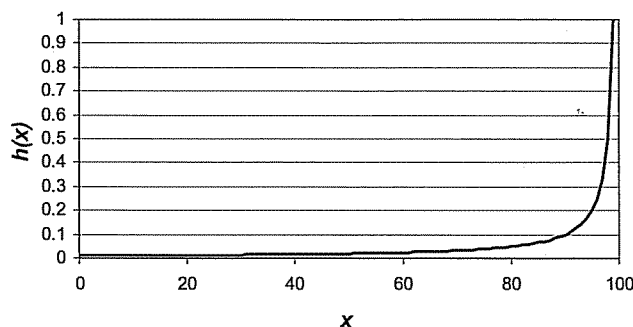


Fig. 2.7 Hazard rate function for Model 1.

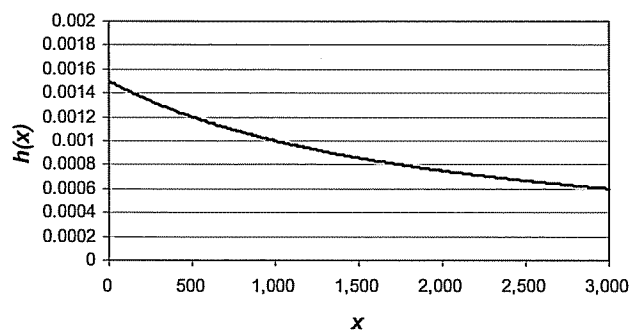


Fig. 2.8 Hazard rate function for Model 2.

calculations. In this text such values will be arbitrarily defined so that the function is right continuous.⁷ See the solution to Exercise 2.1 for an example.

$$S_5(x) = \begin{cases} 1 - 0.01x, & 0 \leq x < 50, \\ 1.5 - 0.02x, & 50 \leq x < 75. \end{cases}$$

□

⁷By arbitrarily defining the value of the density or hazard rate function at such a point, it is clear that using either of them to obtain the survival function will work. If there is discrete probability at this point (in which case these functions are left undefined), then the density and hazard functions are not sufficient to completely describe the probability distribution.

A variety of commonly used continuous distributions are presented in Appendix A and many discrete distributions are presented in Appendix B. An interesting feature of a random variable is the value that is most likely to occur.

Definition 2.17 The *mode* of a random variable is the most likely value. For a discrete variable it is the value with the largest probability. For a continuous variable it is the value for which the density function is largest. If there are local maxima, these points are also considered to be modes.

Example 2.18 Where possible, determine the mode for Models 1–5.

For Model 1, the density function is constant. All values from 0 to 100 could be the mode, or equivalently, it could be said that there is no mode. For Model 2, the density function is strictly decreasing and so the mode is at 0. For Model 3, the probability is highest at 0. As a mixed distribution, it is not possible to define a mode for Model 4. Model 5 has a density that is constant over two intervals, with higher values from 50 to 75. These values are all modes. □

2.2.1 Exercises

2.1 Determine the distribution, density, and hazard rate functions for Model 5.

2.2 Construct graphs of the distribution function for Models 3–5. Also graph the density or probability function as appropriate and the hazard rate function, where it exists.

2.3 (*) A random variable X has density function $f(x) = 4x(1 + x^2)^{-3}$, $x > 0$. Determine the mode of X .

2.4 (*) A nonnegative random variable has a hazard rate function of $h(x) = A + e^{2x}$, $x \geq 0$. You are also given $S(0.4) = 0.5$. Determine the value of A .

2.5 (*) X has a Pareto distribution with parameters $\alpha = 2$ and $\theta = 10,000$. Y has a Burr distribution with parameters $\alpha = 2$, $\gamma = 2$, and $\theta = \sqrt{20,000}$. Let r be the ratio of $\Pr(X > d)$ to $\Pr(Y > d)$. Determine $\lim_{d \rightarrow \infty} r$.

3

Basic distributional quantities

3.1 MOMENTS

There are a variety of interesting calculations that can be done from the models described in the previous chapter. Examples are the average amount paid on a claim that is subject to a deductible or policy limit or the average remaining lifetime of a person age 40.

Definition 3.1 *The k th raw moment of a random variable is the expected (average) value of the k th power of the variable, provided it exists. It is denoted by $E(X^k)$ or by μ'_k . The first raw moment is called the mean of the random variable and is usually denoted by μ .*

Note that μ is not related to $\mu(x)$, the force of mortality as mentioned on page 21. For random variables that take on only nonnegative values [that is, $\Pr(X \geq 0) = 1$], k may be any real number. When presenting formulas for calculating this quantity, a distinction between continuous and discrete variables needs to be made. Formulas will be presented for random variables that are either everywhere continuous or everywhere discrete. For mixed models, evaluate the formula by integrating with respect to its density function wherever the random variable is continuous and by summing with respect to its probability function wherever the random variable is discrete and adding

the results. The formula for the k th raw moment is

$$\begin{aligned}\mu'_k = E(X^k) &= \int_{-\infty}^{\infty} x^k f(x) dx \text{ if the random variable is continuous} \\ &= \sum_j x_j^k p(x_j) \text{ if the random variable is discrete,}\end{aligned}\quad (3.1)$$

where the sum is to be taken over all x_j with positive probability. Finally, note that it is possible that the integral or sum will not converge, in which case the moment is said not to exist.

Example 3.2 Determine the first two raw moments for each of the five models.

The subscripts on the random variable X indicate which model is being used.

$$\begin{aligned}E(X_1) &= \int_0^{100} x(0.01)dx = 50, \\ E(X_1^2) &= \int_0^{100} x^2(0.01)dx = 3,333.33, \\ E(X_2) &= \int_0^{\infty} x \frac{3(2,000)^3}{(x+2,000)^4} dx = 1,000, \\ E(X_2^2) &= \int_0^{\infty} x^2 \frac{3(2,000)^3}{(x+2,000)^4} dx = 4,000,000, \\ E(X_3) &= 0(0.5) + 1(0.25) + 2(0.12) + 3(0.08) + 4(0.05) = 0.93, \\ E(X_3^2) &= 0(0.5) + 1(0.25) + 4(0.12) + 9(0.08) + 16(0.05) = 2.25, \\ E(X_4) &= 0(0.7) + \int_0^{\infty} x(0.000003)e^{-0.00001x} dx = 30,000, \\ E(X_4^2) &= 0^2(0.7) + \int_0^{\infty} x^2(0.000003)e^{-0.00001x} dx = 6,000,000,000, \\ E(X_5) &= \int_0^{50} x(0.01)dx + \int_{50}^{75} x(0.02)dx = 43.75, \\ E(X_5^2) &= \int_0^{50} x^2(0.01)dx + \int_{50}^{75} x^2(0.02)dx = 2,395.83.\end{aligned}\quad \square$$

Before proceeding further, an additional model will be introduced. This one looks similar to Model 3, but with one key difference. It is discrete, but with the added requirement that all of the probabilities must be integral multiples of some number. In addition, the model must be related to sample data in a particular way.

Definition 3.3 The empirical model is a discrete distribution based on a sample of size n which assigns probability $1/n$ to each data point.

Model 6 Consider a sample of size 8 in which the observed data points were 3, 5, 6, 6, 6, 7, 7, and 10. The empirical model then has probability function

$$p_6(x) = \begin{cases} 0.125, & x = 3, \\ 0.125, & x = 5, \\ 0.375, & x = 6, \\ 0.25, & x = 7, \\ 0.125, & x = 10. \end{cases} \quad \square$$

Alert readers will note that many discrete models with finite support look like empirical models. Model 3 could have been the empirical model for a sample of size 100 that contained 50 zeros, 25 ones, 12 twos, 8 threes, and 5 fours. Regardless, we will use the term empirical model only when there is an actual sample behind it. The two moments for Model 6 are

$$E(X_6) = 6.25, \quad E(X_6^2) = 42.5$$

using the same approach as in Model 3. It should be noted that the mean of this random variable is equal to the sample arithmetic average (also called the sample mean).

Definition 3.4 The k th central moment of a random variable is the expected value of the k th power of the deviation of the variable from its mean. It is denoted by $E[(X - \mu)^k]$ or by μ_k . The second central moment is usually called the variance and denoted σ^2 and its square root, σ , is called the standard deviation. The ratio of the standard deviation to the mean is called the coefficient of variation. The ratio of the third central moment to the cube of the standard deviation, $\gamma_1 = \mu_3/\sigma^3$, is called the skewness. The ratio of the fourth central moment to the fourth power of the standard deviation, $\gamma_2 = \mu_4/\sigma^4$, is called the kurtosis.¹

The continuous and discrete formulas for calculating central moments are

$$\begin{aligned}\mu_k &= E[(X - \mu)^k] \\ &= \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx \text{ if the random variable is continuous} \\ &= \sum_j (x_j - \mu)^k p(x_j) \text{ if the random variable is discrete.}\end{aligned}\quad (3.2)$$

¹It would be more accurate to call these items the "coefficient of skewness" and "coefficient of kurtosis" because there are other quantities that also measure asymmetry and flatness. The simpler expressions will be used in this text.

In reality, the integral need be taken only over those x values where $f(x)$ is positive. The standard deviation is a measure of how much the probability is spread out over the random variable's possible values. It is measured in the same units as the random variable itself. The coefficient of variation measures the spread relative to the mean. The skewness is a measure of asymmetry. A symmetric distribution has a skewness of zero, while a positive skewness indicates that probabilities to the right tend to be assigned to values further from the mean than those to the left. The kurtosis measures flatness of the distribution relative to a normal distribution (which has a kurtosis of 3). Kurtosis values above 3 indicate that (keeping the standard deviation constant), relative to a normal distribution, more probability tends to be at points away from the mean than at points near the mean. The coefficients of variation, skewness, and kurtosis are all dimensionless.

There is a link between raw and central moments. The following equation indicates the connection between second moments. The development uses the continuous version from (3.1) and (3.2), but the result applies to all random variables.

$$\begin{aligned}\mu_2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} (x^2 - 2x\mu + \mu^2) f(x) dx \\ &= E(X^2) - 2\mu E(X) + \mu^2 = \mu'_2 - \mu^2.\end{aligned}\quad (3.3)$$

Example 3.5 The density function of the gamma distribution appears to be positively skewed. Demonstrate that this is true and illustrate with graphs.

From Appendix A, the first three raw moments of the gamma distribution are $\alpha\theta$, $\alpha(\alpha+1)\theta^2$, and $\alpha(\alpha+1)(\alpha+2)\theta^3$. From (3.3) the variance is $\alpha\theta^2$ and from the solution to Exercise 3.1 the third central moment is $2\alpha\theta^3$. Therefore, the skewness is $2\alpha^{-1/2}$. Because α must be positive, the skewness is always positive. Also, as α decreases, the skewness increases.

Consider the following two gamma distributions. One has parameters $\alpha = 0.5$ and $\theta = 100$ while the other has $\alpha = 5$ and $\theta = 10$. These have the same mean, but their skewness coefficients are 2.83 and 0.89, respectively. Figure 3.1 demonstrates the difference. \square

Note that when calculating the standard deviation for Model 6 in Exercise 3.2 the result is the sample standard deviation using n (as opposed to the more commonly used $n-1$) in the denominator. Finally, it should be noted that when calculating moments it is possible that the integral or sum will not exist (as is the case for the third and fourth moments for Model 2). For the models we typically encounter, the integrand and summand are nonnegative and so failure to exist implies that the required limit that gives the integral or sum is infinity. See Example 4.15 for an illustration.

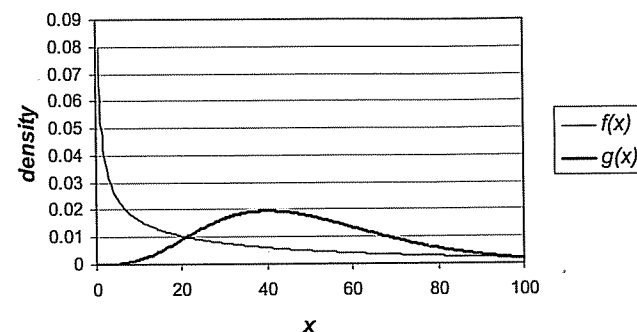


Fig. 3.1 Densities of $f(x) \sim \text{gamma}(0.5, 100)$ and $g(x) \sim \text{gamma}(5, 10)$.

Definition 3.6 For a given value of d with $\Pr(X > d) > 0$, the excess loss variable is $Y = X - d$ given that $X > d$. Its expected value,

$$e_X(d) = e(d) = E(Y) = E(X - d | X > d),$$

is called the mean excess loss function. Other names for this expectation are mean residual life function and complete expectation of life. When the latter terminology is used, the commonly used symbol is \bar{e}_d .

This variable could also be called a left truncated and shifted variable. It is left truncated because observations below d are discarded. It is shifted because d is subtracted from the remaining values. When X is a payment variable, the mean excess loss is the expected amount paid given that there has been a payment in excess of a deductible of d . When X is the age at death, the mean excess loss is the expected remaining time until death given that the person is alive at age d . The k th moment of the excess loss variable is determined from

$$\begin{aligned}e_X^k(d) &= \frac{\int_d^{\infty} (x - d)^k f(x) dx}{1 - F(d)} \text{ if the variable is continuous} \\ &= \frac{\sum_{x_j > d} (x_j - d)^k p(x_j)}{1 - F(d)} \text{ if the variable is discrete.}\end{aligned}\quad (3.4)$$

Here, $e_X^k(d)$ is defined only if the integral or sum converges. There is a particularly convenient formula for calculating the first moment. The development is given below for the continuous version, but the result holds for all random variables. The second line is based on an integration by parts where the

antiderivative of $f(x)$ is taken as $-S(x)$.

$$\begin{aligned}
 e_X(d) &= \frac{\int_d^\infty (x-d)f(x)dx}{1-F(d)} \\
 &= \frac{-(x-d)S(x)|_d^\infty + \int_d^\infty \dot{S}(x)dx}{S(d)} \\
 &= \frac{\int_d^\infty S(x)dx}{S(d)}.
 \end{aligned} \tag{3.5}$$

Definition 3.7 The left censored and shifted variable is

$$Y = (X-d)_+ = \begin{cases} 0, & X < d, \\ X-d, & X \geq d. \end{cases}$$

It is left censored because values below d are not ignored but are set equal to 0. There is no standard name or symbol for the moments of this variable. For dollar events, the distinction between the excess loss variable and the left censored and shifted variable is one of *per payment* versus *per loss*. In the former situation, the variable exists only when a payment is made. The latter variable takes on the value 0 whenever a loss produces no payment. The moments can be calculated from

$$\begin{aligned}
 E[(X-d)_+^k] &= \int_d^\infty (x-d)^k f(x)dx \text{ if the variable is continuous,} \\
 &= \sum_{x_j > d} (x_j - d)^k p(x_j) \text{ if the variable is discrete.}
 \end{aligned} \tag{3.6}$$

It should be noted that

$$E[(X-d)_+^k] = e^k(d)[1-F(d)]. \tag{3.7}$$

Example 3.8 Construct graphs to illustrate the difference between the excess loss variable and the left censored and shifted variable.

The two graphs in Figures 3.2 and 3.3 plot the modified variable Y as a function of the unmodified variable X . The only difference is that for X values below 100 the variable is undefined while for the left censored and shifted variable it is set equal to zero. □

The next definition provides a complementary function to the excess loss.

Definition 3.9 The limited loss variable is

$$Y = X \wedge u = \begin{cases} X, & X < u, \\ u, & X \geq u. \end{cases}$$

Its expected value, $E[X \wedge u]$, is called the *limited expected value*.

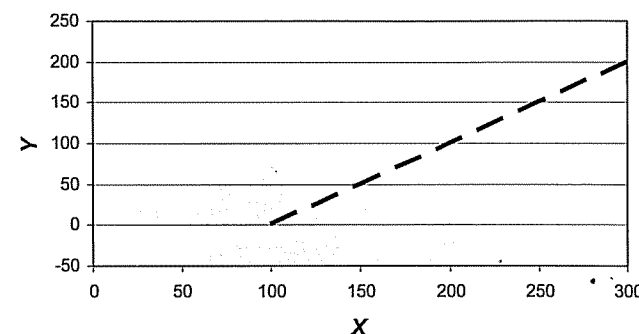


Fig. 3.2 Excess loss variable.

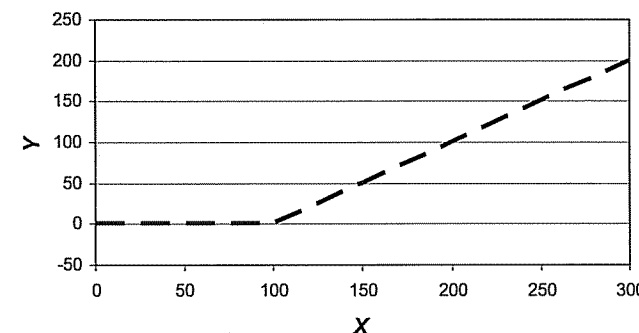


Fig. 3.3 Left censored and shifted variable.

This variable could also be called the **right censored variable**. It is right censored because values above u are set equal to u . An insurance phenomenon that relates to this variable is the existence of a policy limit that sets a maximum on the benefit to be paid. Note that $(X-d)_+ + (X \wedge d) = X$. That is, buying one policy with a limit of d and another with a deductible of d is equivalent to buying full coverage. This is illustrated in Figure 3.4.

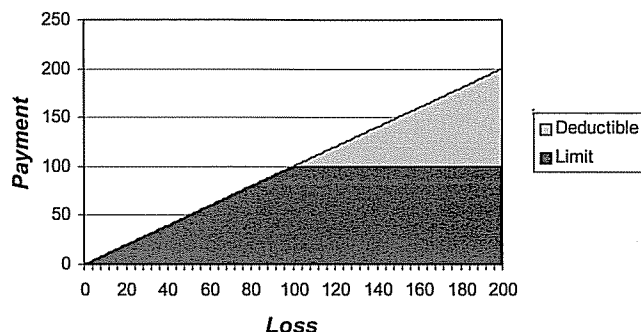


Fig. 3.4 Limit of 100 plus deductible of 100 equals full coverage.

The most direct formulas for the k th moment of the limited loss variable are

$$\begin{aligned}
 E[(X \wedge u)^k] &= \int_{-\infty}^u x^k f(x) dx + u^k [1 - F(u)] \\
 &\quad \text{if the random variable is continuous} \\
 &= \sum_{x_j \leq u} x_j^k p(x_j) + u^k [1 - F(u)] \\
 &\quad \text{if the random variable is discrete.}
 \end{aligned} \tag{3.8}$$

Another interesting formula is derived as follows:

$$\begin{aligned}
 E[(X \wedge u)^k] &= \int_{-\infty}^0 x^k f(x) dx + \int_0^u x^k f(x) dx + u^k [1 - F(u)] \\
 &= x^k F(x) \Big|_{-\infty}^0 - \int_{-\infty}^0 kx^{k-1} F(x) dx \\
 &\quad - x^k S(x) \Big|_0^u + \int_0^u kx^{k-1} S(x) dx + u^k S(u) \\
 &= - \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^u kx^{k-1} S(x) dx,
 \end{aligned} \tag{3.9}$$

where the second line uses integration by parts. For $k = 1$, we have

$$E(X \wedge u) = - \int_{-\infty}^0 F(x) dx + \int_0^u S(x) dx.$$

The corresponding formula for discrete random variables is not particularly interesting. The limited expected value also represents the expected dollar

saving per incident when a deductible is imposed. The k th limited moment of many common continuous distributions is presented in Appendix A. Exercise 3.8 asks you to develop a relationship between the three first moments introduced previously.

3.1.1 Exercises

3.1 Develop formulas similar to (3.3) for μ_3 and μ_4 .

3.2 Calculate the standard deviation, skewness, and kurtosis for each of the six models. It may help to note that Model 2 is a Pareto distribution and the density function in the continuous part of Model 4 is an exponential distribution. Formulas that may help with calculations for these models appear in Appendix A.

3.3 (*) A random variable has a mean and a coefficient of variation of 2. The third raw moment is 136. Determine the skewness.

3.4 (*) Determine the skewness of a gamma distribution that has a coefficient of variation of 1.

3.5 Determine the mean excess loss function for Models 1–4. Compare the functions for Models 1, 2, and 4.

3.6 (*) For two random variables, X and Y , $e_Y(30) = e_X(30) + 4$. Let X have a uniform distribution on the interval from 0 to 100 and let Y have a uniform distribution on the interval from 0 to w . Determine w .

3.7 (*) A random variable has density function $f(x) = \lambda^{-1} e^{-x/\lambda}$, $x, \lambda > 0$. Determine $e(\lambda)$, the mean residual life function evaluated at λ .

3.8 Show that the following relationship holds:

$$E(X) = e(d)S(d) + E(X \wedge d). \tag{3.10}$$

3.9 Determine the limited expected value function for Models 1–4. Do this using both (3.8) and (3.10). For Models 1 and 2 also obtain the function using (3.9).

3.10 (*) Which of the following statements are true?

- (a) The mean residual life function for an empirical distribution is continuous.
- (b) The mean residual life function for an exponential distribution is constant.
- (c) If it exists, the mean residual life function for a Pareto distribution is decreasing.



3.11 (*) Losses have a Pareto distribution with $\alpha = 0.5$ and $\theta = 10,000$. Determine the mean residual life at 10,000.

3.12 Define a right truncated variable and provide a formula for its k th moment.

3.13 (*) The severity distribution of individual claims has pdf

$$f(x) = 2.5x^{-3.5}, \quad x \geq 1.$$

Determine the coefficient of variation.

3.14 (*) Claim sizes are for 100, 200, 300, 400, or 500. The true probabilities for these values are 0.05, 0.20, 0.50, 0.20, and 0.05, respectively. Determine the skewness and kurtosis for this distribution.

3.15 (*) Losses follow a Pareto distribution with $\alpha > 1$ and θ unspecified. Determine the ratio of the mean excess loss function at $x = 2\theta$ to the mean excess loss function at $x = \theta$.

3.2 PERCENTILES

One other value of interest that may be derived from the distribution function is the percentile function. It is the inverse of the distribution function, but because this quantity is not well defined, an arbitrary definition must be created.

Definition 3.10 The 100 p th percentile of a random variable is any value π_p such that $F(\pi_p-) \leq p \leq F(\pi_p)$. The 50th percentile, $\pi_{0.5}$ is called the median.

If the distribution function has a value of p for one and only one x value, then the percentile is uniquely defined. In addition, if the distribution function jumps from a value below p to a value above p , then the percentile is at the location of the jump. The only time the percentile is not uniquely defined is when the distribution function is constant at a value of p over a range of values. In that case, any value in that range can be used as the percentile.

Example 3.11 Determine the 50th and 80th percentiles for Models 1 and 3.

For Model 1, the p th percentile can be obtained from $p = F(\pi_p) = 0.01\pi_p$ and so $\pi_p = 100p$, and in particular, the requested percentiles are 50 and 80 (see Figure 3.5). For Model 3 the distribution function equals 0.5 for all $0 \leq x < 1$ and so all such values can be the 50th percentile. For the 80th percentile, note that at $x = 2$ the distribution function jumps from 0.75 to 0.87 and so $\pi_{0.8} = 2$ (see Figure 3.6). \square

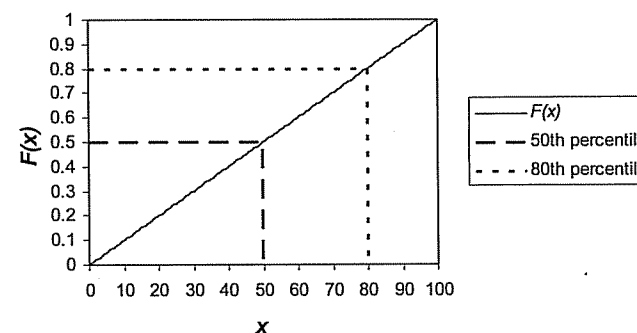


Fig. 3.5 Percentiles for Model 1.

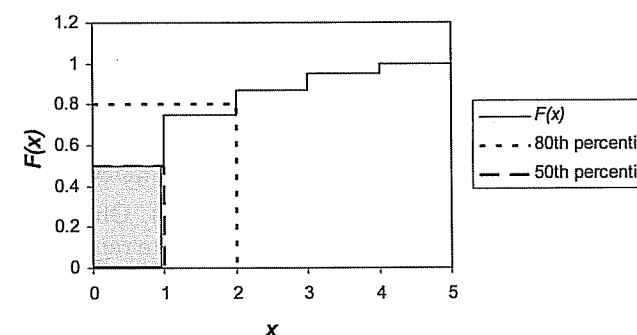


Fig. 3.6 Percentiles for Model 2.

3.2.1 Exercises

3.16 (*) The cdf of a random variable is $F(x) = 1 - x^{-2}$, $x \geq 1$. Determine the mean, median, and mode of this random variable.

3.17 Determine the 50th and 80th percentiles for Models 2, 4, 5, and 6.

3.18 (*) Losses have a Pareto distribution with parameters α and θ . The 10th percentile is $\theta - k$. The 90th percentile is $5\theta - 3k$. Determine the value of α .

3.19 (*) Losses have a Weibull distribution with parameters τ and θ . The 25th percentile is 1,000 and the 75th percentile is 100,000. Determine the value of τ .

3.3 GENERATING FUNCTIONS AND SUMS OF RANDOM VARIABLES

An insurance company rarely insures only one person. The total claims paid on all policies is the sum of all payments. Thus it is useful to be able to determine properties of $S_k = X_1 + \cdots + X_k$. The first result is a version of the central limit theorem.

Theorem 3.12 For a random variable S_k as defined above, $E(S_k) = E(X_1) + \cdots + E(X_k)$. Also, if X_1, \dots, X_k are independent, $\text{Var}(S_k) = \text{Var}(X_1) + \cdots + \text{Var}(X_k)$. If the random variables X_1, X_2, \dots are independent and their first two moments meet certain conditions, $\lim_{k \rightarrow \infty} [S_k - E(S_k)] / \sqrt{\text{Var}(S_k)}$ has a normal distribution with mean 0 and variance 1.

Obtaining the distribution or density function of S_k is usually very difficult. However, there are a few cases where it is simple. The key to this simplicity is the generating function.

Definition 3.13 For a random variable X , the *moment generating function (mgf)* is $M_X(t) = E(e^{tX})$ for all t for which the expected value exists. The *probability generating function (pgf)* is $P_X(z) = E(z^X)$ for all z for which the expectation exists.

Note that $M_X(t) = P_X(e^t)$ and $P_X(z) = M_X(\ln z)$. Often the mgf is used for continuous random variables and the pgf for discrete random variables. For us, the value of these functions is not so much that they generate moments or probabilities but that there is a one-to-one correspondence between a random variable's distribution function and its mgf and pgf (i.e., two random variables with different distribution functions cannot have the same mgf or pgf). The following result aids in working with sums of random variables.

Theorem 3.14 Let $S_k = X_1 + \cdots + X_k$, where the random variables in the sum are independent. Then $M_{S_k}(t) = \prod_{j=1}^k M_{X_j}(t)$ and $P_{S_k}(z) = \prod_{j=1}^k P_{X_j}(z)$ provided all the component mgfs and pgfs exist.

Proof: We use the fact that the expected product of independent random variables is the product of the individual expectations. Then,

$$\begin{aligned} M_{S_k}(t) &= E(e^{tS_k}) = E[e^{t(X_1 + \cdots + X_k)}] \\ &= \prod_{j=1}^k E(e^{tX_j}) = \prod_{j=1}^k M_{X_j}(t). \end{aligned}$$

A similar argument can be used for the pgf. \square

Example 3.15 Show that the sum of independent gamma random variables, each with same value of θ , has a gamma distribution.

The moment generating function of a gamma variable is

$$\begin{aligned} E(e^{tX}) &= \frac{\int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\theta} dx}{\Gamma(\alpha) \theta^\alpha} \\ &= \frac{\int_0^\infty x^{\alpha-1} e^{-x(-t+1/\theta)} dx}{\Gamma(\alpha) \theta^\alpha} \\ &= \frac{\int_0^\infty y^{\alpha-1} (-t+1/\theta)^{-\alpha} e^{-y} dy}{\Gamma(\alpha) \theta^\alpha} \\ &= \frac{\Gamma(\alpha) (-t+1/\theta)^{-\alpha}}{\Gamma(\alpha) \theta^\alpha} = \left(\frac{1}{1-\theta t} \right)^\alpha, \quad t < 1/\theta. \end{aligned}$$

Now let X_j have a gamma distribution with parameters α_j and θ . Then the moment generating function of the sum is

$$M_{S_k}(t) = \prod_{j=1}^k \left(\frac{1}{1-\theta t} \right)^{\alpha_j} = \left(\frac{1}{1-\theta t} \right)^{\alpha_1 + \cdots + \alpha_k}$$

which is the moment generating function of a gamma distribution with parameters $\alpha_1 + \cdots + \alpha_k$ and θ . \square

Example 3.16 Obtain the mgf and pgf for the Poisson distribution.

The pgf is

$$P_X(z) = \sum_{x=0}^{\infty} z^x \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(z\lambda)^x}{x!} = e^{-\lambda} e^{z\lambda} = e^{\lambda(z-1)}.$$

Then the mgf is $M_X(t) = P_X(e^t) = \exp[\lambda(e^t - 1)]$. \square

3.3.1 Exercises

3.20 (*) A portfolio contains 16 independent risks, each with a gamma distribution with parameters $\alpha = 1$ and $\theta = 250$. Give an expression using the incomplete gamma function for the probability that the sum of the losses exceeds 6,000. Then approximate this probability using the central limit theorem.

3.21 (*) The severities of individual claims have the Pareto distribution with parameters $\alpha = 8/3$, and $\theta = 8,000$. Use the central limit theorem to approximate the probability that the sum of 100 independent claims will exceed 600,000.

3.22 (*) The severities of individual claims have the gamma distribution (see Appendix A) with parameters $\alpha = 5$ and $\theta = 1,000$. Use the central limit

theorem to approximate the probability that the sum of 100 independent claims exceeds 525,000.

3.23 A sample of 1,000 health insurance contracts on adults produced a sample mean of 1,300 for the annual benefits paid with a standard deviation of 400. It is expected that 2,500 contracts will be issued next year. Use the central limit theorem to estimate the probability that benefit payments will be more than 101% of the expected amount.

3.24 Show that the mgf for the inverse Gaussian distribution with θ replaced by $\beta = \mu^2/\theta$ is given by

$$M(t) = \exp \left[\frac{\mu}{\beta} \left(1 - \sqrt{1 - 2\beta t} \right) \right], \quad t < 1/(2\beta).$$

4

Classifying and creating distributions

4.1 INTRODUCTION

The set of all possible distribution functions is too large to comprehend. Therefore, when searching for a distribution function to use as a model for a random phenomenon, it can be helpful if the field can be narrowed. One division that has already been discussed is the separation into discrete, continuous, and mixed distributions. In most situations it will be obvious which of the three applies. Beyond this, we need more artificial distinctions. The next section describes a split based on the complexity of the model. The following section then looks at the shape of the distribution to distinguish one from another. After that, a few methods of creating additional distributions are introduced. This is followed by a listing of some commonly used continuous distributions. The final section is an extensive treatment of discrete distributions. By the end of this chapter, most of the distributions in Appendices A and B will have been introduced. While this chapter is more about differences from one distribution to another, these distributions have some common elements that are desirable for actuarial models. Among them are:

- The support is a subset of the nonnegative real numbers. Most actuarial phenomena are counts or are measurements of time or money and as such are rarely negative, although if the random variable of interest

is financial gain, negative outcomes are certainly possible. However, financial gain is often the result of the realization of several nonnegative variables, some measuring income and some measuring expenses.

- Some distributions are special cases of others. This allows the modeler to choose from models of varying complexity.
- They can have one or more modes and a mode may or may not be at zero.

4.2 THE ROLE OF PARAMETERS

This split has to do with how much information is needed to specify the model. The number of quantities (parameters) needed to do so gives some indication of how complex a model is, in the sense that many items are needed to describe it. Arguments for a simple model include the following:

- With few items required in its specification, it is more likely that each item can be determined more accurately.
- It is more likely to be stable across time and across settings. That is, if the model does well today, it (perhaps with small changes to reflect inflation or similar phenomena) will probably do well tomorrow and will also do well in other, similar, situations.
- Because data can often be irregular, a simple model may provide necessary smoothing.

Of course, complex models also have advantages.

- With many items required in its specification, it can more closely match reality.
- With many items required in its specification, it can more closely match irregularities in the data.

Another way to express the difference is that simpler models can be estimated more accurately, but the model itself may be too superficial. The principle of parsimony states that the simplest model that adequately reflects reality should be used. The definition of “adequately” will depend on the purpose for which the model is to be used.

In the following subsections, we will move from simpler models to more complex models. There is some difficulty in naming the various classifications because there is not universal agreement on the definitions. With the exception of parametric distributions, the other category names have been created by the authors. It should also be understood that these categories do not cover the universe of possible models nor will every model be easy to categorize. These should be considered as qualitative descriptions.

4.2.1 Parametric and scale distributions

These models are simple enough to be specified by a few key numbers.

Definition 4.1 A *parametric distribution* is a set of distribution functions, each member of which is determined by specifying one or more values called *parameters*. The number of parameters is fixed and finite.

The most familiar parametric distribution is the normal distribution with parameters μ and σ^2 . When values for these two parameters are specified, the distribution function is completely known.

These are the simplest distributions in this subsection, because typically only a small number of values need to be specified. All of the individual distributions in Appendices A and B are parametric. Within this class, distributions with fewer parameters are simpler than those with more parameters.

For much of actuarial modeling work, it is especially convenient if the name of the distribution is unchanged when the random variable is multiplied by a constant. The most common uses for this phenomenon are to model the effect of inflation and to accommodate a change in the monetary unit.

Definition 4.2 A *parametric distribution* is a *scale distribution* if, when a random variable from that set of distributions is multiplied by a positive constant, the resulting random variable is also in that set of distributions.

Example 4.3 Demonstrate that the exponential distribution is a scale distribution.

According to Appendix A, the distribution function is $F_X(x) = 1 - e^{-x/\theta}$. Let $Y = cX$, where $c > 0$. Then,

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y), \\ &= \Pr(cX \leq y), \\ &= \Pr\left(X \leq \frac{y}{c}\right), \\ &= 1 - e^{-y/c\theta}. \end{aligned}$$

But this is an exponential distribution with parameter $c\theta$. □

Definition 4.4 For random variables with nonnegative support, a *scale parameter* is a parameter for a scale distribution that meets two conditions. First, when a member of the scale distribution is multiplied by a positive constant, the scale parameter is multiplied by the same constant. Second, when a member of the scale distribution is multiplied by a positive constant, all other parameters are unchanged.

Example 4.5 Demonstrate that the gamma distribution, as defined in Appendix A, has a scale parameter.

Let X have the gamma distribution and $Y = cX$. Then, using the incomplete gamma notation in Appendix A,

$$\begin{aligned} F_Y(y) &= \Pr\left(X \leq \frac{y}{c}\right) \\ &= \Gamma\left(\alpha; \frac{y}{c\theta}\right) \end{aligned}$$

indicating that Y has a gamma distribution with parameters α and $c\theta$. Therefore, the parameter θ is a scale parameter. \square

Many textbooks write the density function for the gamma distribution as

$$f(x) = \frac{x^{\alpha-1} e^{-\beta x} \beta^\alpha}{\Gamma(\alpha)}.$$

We have chosen to use the version of the density function that has a scale parameter. When the alternative version is multiplied by c , the parameters become α and β/c . As well, the mean is proportional to θ in our version, while it is proportional to $1/\beta$ in the alternative version. Our version makes it easier to get ballpark estimates of this parameter, although, for the alternative definition, one need only keep in mind that the parameter is inversely proportional to the mean.

It is often possible to recognize a scale parameter from looking at the distribution or density function. In particular, the distribution function would have x always appear as x/θ .

4.2.2 Parametric distribution families

A slightly more complex version of a parametric distribution is one in which the number of parameters is finite but not fixed in advance.

Definition 4.6 A *parametric distribution family* is a set of parametric distributions that are related in some meaningful way.

The most common type of parametric distribution family is described in the following example.

Example 4.7 One type of parametric distribution family is based on a specified parametric distribution. Other members of the family are obtained by setting one or more parameters from the specified distribution equal to a pre-set value or to each other. Demonstrate that the transformed beta family as defined in Appendix A is a parametric distribution family.

The transformed beta distribution has four parameters. Each of the other named distributions in the family is a transformed beta distribution with certain parameters set equal to 1 (for example, the Pareto distribution has $\gamma = \tau = 1$) or to each other (the paralogistic distribution has $\tau = 1$ and $\gamma = \alpha$). Note that the number of parameters (ranging from two to four) is not known in advance. There is a subtle difference in definitions. A modeler who uses the transformed beta distribution looks at all four parameters over their range of possible values. A modeler who uses the transformed beta family pays particular attention to the possibility of using special cases such as the Burr distribution. For example, if the former modeler collects some data and decides that $\tau = 1.01$, that will be the value to use. The latter modeler will note that $\tau = 1$ gives a Burr distribution and will likely use that model instead. \square

4.2.3 Finite mixture distributions

By themselves, mixture distributions are no more complex, but later in this subsection we will find a way to increase the complexity level. One motivation for mixing is that the underlying phenomenon may actually be several phenomena that occur with unknown probabilities. For example, a randomly selected dental claim may be from a check-up, from a filling, from a repair (such as a crown), or from a surgical procedure. Because of the differing modes for these possibilities, a mixture model may work well.

Definition 4.8 A random variable Y is a *k-point mixture*¹ of the random variables X_1, X_2, \dots, X_k if its cdf is given by

$$F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \dots + a_k F_{X_k}(y), \quad (4.1)$$

where all $a_j > 0$ and $a_1 + a_2 + \dots + a_k = 1$.

This essentially assigns probability a_j to the outcome that Y is a realization of the random variable X_j . Note that, if we have 20 choices for a given random variable, a two-point mixture allows us to create over 200 new distributions.² This may be sufficient for most modeling situations. Nevertheless, these are still parametric distribution, though perhaps with many parameters.

Example 4.9 For models involving general liability insurance, actuaries at the Insurance Services Office has had some success with a mixture of two

¹The words "mixed" and "mixture" have been used interchangeably to refer to the type of distribution described here as well as distributions that are partly discrete and partly continuous. This text will not attempt to resolve that confusion. The context will make clear which type of distribution is being considered.

²There are actually $\binom{20}{2} + 20 = 210$ choices. The extra 20 represent the cases where both distributions are of the same type but with different parameters.

Pareto distributions. They also found that five parameters were not necessary. The distribution they selected has cdf

$$F(x) = 1 - a \left(\frac{\theta_1}{\theta_1 + x} \right)^\alpha - (1 - a) \left(\frac{\theta_2}{\theta_2 + x} \right)^{\alpha+2}.$$

Note that the shape parameters in the two Pareto distributions differ by 2. The second distribution places more probability on smaller values. This might be a model for frequent, small claims while the first distribution covers large, but infrequent claims. This distribution has only four parameters, bringing some parsimony to the modeling process. \square

Suppose we do not know how many distributions should be in the mixture. Then the value of k becomes a parameter, as indicated in the following definition.

Definition 4.10 A variable-component mixture distribution has a distribution function that can be written as

$$F(x) = \sum_{j=1}^K a_j F_j(x), \quad \sum_{j=1}^K a_j = 1, \quad a_j > 0, \quad j = 1, \dots, K, \quad K = 1, 2, \dots$$

These models have been called *semiparametric* because in complexity they are between parametric models and nonparametric models (see the next subsection). This distinction becomes more important when model selection is discussed in Chapter 13. When the number of parameters is to be estimated from data, hypothesis tests to determine the appropriate number of parameters become more difficult. When all of the components have the same parametric distribution (but different parameters), the resulting distribution is called a "variable mixture of gs" distribution, where g stands for the name of the component distribution.

Example 4.11 Determine the distribution, density, and hazard rate functions for the variable mixture of exponentials distribution.

A combination of exponential distribution functions can be written

$$\begin{aligned} F(x) &= 1 - a_1 e^{-x/\theta_1} - a_2 e^{-x/\theta_2} - \dots - a_K e^{-x/\theta_K}, \\ \sum_{j=1}^K a_j &= 1, \quad a_j, \theta_j > 0, \quad j = 1, \dots, K, \quad K = 1, 2, \dots \end{aligned}$$

and then the other functions are

$$\begin{aligned} f(x) &= a_1 \theta_1^{-1} e^{-x/\theta_1} + a_2 \theta_2^{-1} e^{-x/\theta_2} + \dots + a_K \theta_K^{-1} e^{-x/\theta_K}, \\ h(x) &= \frac{a_1 \theta_1^{-1} e^{-x/\theta_1} + a_2 \theta_2^{-1} e^{-x/\theta_2} + \dots + a_K \theta_K^{-1} e^{-x/\theta_K}}{a_1 e^{-x/\theta_1} + a_2 e^{-x/\theta_2} + \dots + a_K e^{-x/\theta_K}}. \end{aligned}$$

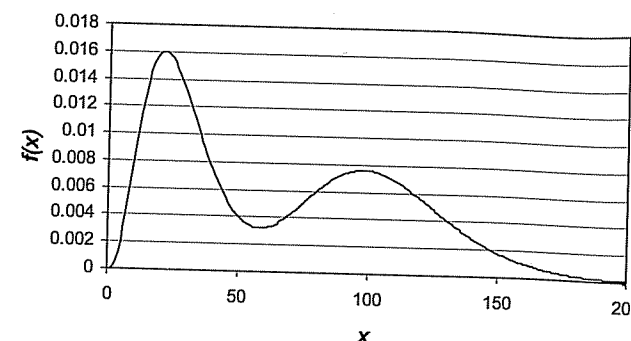


Fig. 4.1 Two-point mixture of gammas distribution.

The number of parameters is not fixed nor is it even limited. For example, when $K = 2$ there are three parameters (a_1, θ_1, θ_2), noting that a_2 is not a parameter because once a_1 is set the value of a_2 is determined. However, when $K = 4$ there are seven parameters. \square

Example 4.12 Illustrate how a two-point mixture of gamma variables can create a bimodal distribution.

Consider a fifty-fifty mixture of two gamma distributions. One has parameters $\alpha = 4$ and $\theta = 7$ (for a mode of 21) and the other has parameters $\alpha = 15$ and $\theta = 7$ (for a mode of 98). The density function is

$$f(x) = 0.5 \frac{x^3 e^{-x/7}}{3!7^4} + 0.5 \frac{x^{14} e^{-x/7}}{14!7^{15}}$$

and a graph appears in Figure 4.1. \square

4.2.4 Data-dependent distributions

Models 1–5 and many of the examples rely on an associated phenomenon (the random variable) but not on observations of that phenomenon. For example, without having observed any dental claims, we could postulate a lognormal distribution with parameters $\mu = 5$ and $\sigma = 1$. Our model may be a poor description of dental claims, but that is a different matter. On the other hand, it is possible to construct models that require data. These models also have parameters but are often called nonparametric.

Definition 4.13 A data-dependent distribution is at least as complex as the data or knowledge that produced it, and the number of "parameters" increases as the number of data points or amount of knowledge increases.

Essentially, these models have as many (or more) "parameters" than observations in the data set. The empirical distribution as illustrated by Model 6 on page 27 is a data-dependent distribution. Each data point contributes probability $1/n$ to the probability function, so the n parameters are the n observations in the data set that produced the empirical distribution.

Another example of a data-dependent model is the kernel smoothing model, which is covered in more detail in Section 11.3. Rather than place a spike of probability $1/n$ at each data point, a continuous density function with area $1/n$ replaces the data point. This piece is centered at the data point so that this model follows the data, but not perfectly. It provides some smoothing versus the empirical distribution. A simple example is given below.

Example 4.14 Construct a kernel smoothing model from Model 6 using the uniform kernel and a bandwidth of 2.

The probability density function is

$$f(x) = \sum_{j=1}^5 p_6(x_j) K_j(x),$$

$$K_j(x) = \begin{cases} 0, & |x - x_j| > 2, \\ 0.25, & |x - x_j| \leq 2, \end{cases}$$

where the sum is taken over the five points where the original model has positive probability. For example, the first term of the sum is the function

$$p_6(x_1) K_1(x) = \begin{cases} 0, & x < 1, \\ 0.03125, & 1 \leq x \leq 5, \\ 0, & x > 5. \end{cases}$$

The complete density function is the sum of five such functions, which are illustrated in Figure 4.2. \square

Note that both the kernel smoothing model and the empirical distribution can also be written as mixture distributions. The reason these models are classified separately is that the number of components relates to the sample size rather than to the phenomenon and its random variable.

4.2.5 Exercises

4.1 Demonstrate that the lognormal distribution as parameterized in Appendix A is a scale distribution but has no scale parameter. Display an alternative parametrization of this distribution that does have a scale parameter.

4.2 Which of Models 1–6 could be considered as members of a parametric distribution? For those that are, name or describe the distribution.

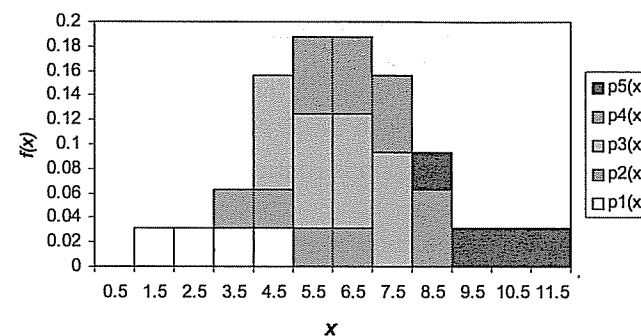


Fig. 4.2 Kernel density distribution.

4.3 (*) Claims have a Pareto distribution with $\alpha = 2$ and θ unknown. Claims the following year experience 6% uniform inflation. Let r be the ratio of the proportion of claims that will exceed d next year to the proportion of claims that exceed d this year. Determine the limit of r as d goes to infinity.

4.4 Determine the mean and second moment of the two-point mixture distribution in Example 4.9. The solution to this exercise provides general formulas for raw moments of a mixture distribution.

4.5 Determine expressions for the mean and variance of the mixture of gammas distribution.

4.6 Which of Models 1–6 could be considered to be from parametric distribution families? Which could be considered to be from variable-component mixture distributions?

4.7 (*) Seventy-five percent of claims have a normal distribution with a mean of 3,000 and a variance of 1,000,000. The remaining 25% have a normal distribution with a mean of 4,000 and a variance of 1,000,000. Determine the probability that a randomly selected claim exceeds 5,000.

4.8 (*) Let X has a Burr distribution with parameters $\alpha = 1$, $\gamma = 2$, and $\theta = \sqrt{1,000}$ and let Y has a Pareto distribution with parameters $\alpha = 1$ and $\theta = 1,000$. Let Z be a mixture of X and Y with equal weight on each component. Determine the median of Z . Let $W = 1.1Z$. Demonstrate that W is also a mixture of a Burr and a Pareto distribution and determine the parameters of W .

4.9 (*) Consider three random variables: X is a mixture of a uniform distribution on the interval 0 to 2 and a uniform distribution on the interval 0 to 3;

Y is the sum of two random variables, one is uniform on 0 to 2 and the other is uniform on 0 to 3; Z is a normal distribution that has been right censored at 1. Match these random variables with the following descriptions:

- (a) Both the distribution and density functions are continuous.
- (b) The distribution function is continuous but the density function is discontinuous.
- (c) The distribution function is discontinuous.

4.10 Demonstrate that the model in Example 4.14 is a mixture of uniform distributions.

4.11 Show that the inverse Gaussian distribution as parameterized in Appendix A is a scale family but does not have a scale parameter.

4.12 Show that the Weibull distribution has a scale parameter.

4.3 TAIL WEIGHT

The *tail* of a distribution (more properly, the right tail) is that part that reveals probabilities about large values. It is of interest to actuaries because it is the occurrence (or lack thereof) of large values that is most influential on profits. Risky types of insurance such as medical malpractice feature more large claims (relative to the mean) than less risky insurances such as automobile physical damage. Random variables that tend to assign higher probabilities to large values are said to be heavy tailed. Tail weight can be a relative concept (model A has a heavier tail than model B) or an absolute concept (distributions with a certain property are classified as heavy tailed). When choosing models, tail weight can help narrow the choices or can confirm someone else's choice. For example, when someone models medical malpractice payments with a Pareto distribution, it seems reasonable as the Pareto distribution is regarded as having a heavy tail. Conversely, the light-tailed lognormal distribution may be a reasonable model for dental insurance payments. However, it should be noted that various measures of tail weight need not agree.

4.3.1 Existence of moments

Recall that in the continuous case the k th raw moment for a random variable that takes on only positive values (like most insurance payment variables) is given by $\int_0^\infty x^k f(x) dx$. Depending on the density function and the value of k , this integral may not exist. If the density function is too large for large values of x , then, when multiplied by the large number x^k , the values will be too large for the integral to converge. Thus, existence of all positive moments indicates a light right tail, while existence of only positive moments up to a

certain value (or existence of no positive moments at all) indicates a heavy right tail.³

Example 4.15 Demonstrate that for the gamma distribution all positive raw moments exist but for the Pareto distribution they do not.

For the gamma distribution

$$\begin{aligned}\mu'_k &= \int_0^\infty x^k \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha} dx \\ &= \int_0^\infty (y\theta)^k \frac{(y\theta)^{\alpha-1} e^{-y}}{\Gamma(\alpha)\theta^\alpha} \theta dy, \text{ making the substitution } y = x/\theta \\ &= \frac{\theta^k}{\Gamma(\alpha)} \Gamma(\alpha + k) < \infty \text{ for all } k > 0.\end{aligned}$$

while for the Pareto distribution

$$\begin{aligned}\mu'_k &= \int_0^\infty x^k \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} dx \\ &= \int_\theta^\infty (y-\theta)^k \frac{\alpha\theta^\alpha}{y^{\alpha+1}} dy, \text{ making the substitution } y = x + \theta \\ &= \alpha\theta^\alpha \int_\theta^\infty \sum_{j=0}^k \binom{k}{j} y^{j-\alpha-1} (-\theta)^{k-j} dy \text{ for integer values of } k.\end{aligned}$$

The integral exists only if all of the exponents on y in the sum are less than -1 . That is, $j - \alpha - 1 < -1$ for all j , or, equivalently, $k < \alpha$. Therefore, only some moments exist. \square

By this reckoning, the Pareto distribution is said to have a heavier tail than the gamma distribution. A look at the moment formulas in Appendix

³In the same manner, existence of all negative moments indicates a light left tail. A feel for the left tail may aid in choosing an appropriate distributional model. The following statements apply to density functions that are monotonic and differentiable near zero.

In particular, $f(0)$ and the slope of the density function near zero are related to the existence of negative moments [that is, $E(X^k)$ where k is negative]. Suppose that negative moments exist only for $k > -r$. If $r < 1$, $f(x)$ goes to infinity as $x \rightarrow 0$. If $r = 1$, $f(0)$ is a nonnegative number. If $1 < r < 2$, $f(0) = 0$ and the slope goes to infinity as $x \rightarrow 0$. If $r = 2$, $f(0) = 0$ and the slope at 0 is a nonnegative number. If $r > 2$, $f(0) = 0$ and the slope at 0 is 0, so only a small portion of the distribution is near zero.

As an example, the Weibull distribution with $r = 0.2$ has been used for workers compensation insurance. The value of r is 0.2, which means that a lot of probability is near 0. Setting θ to produce a mean of 30,000 gives a 28% chance of a claim being less than 1 and a 1% chance of it exceeding 500,000. This may be a reasonable model for large losses (see Section 4.4.7 for a way to use part of a model). In contrast, the lognormal distribution that has the same mean and variance has less than 0.1% probability of being less than 1 and also about 1% probability of exceeding 500,000. The lognormal distribution has all negative moments and thus has $f(x)$ go to zero as x goes to zero.

A reveals which distributions have heavy tails and which do not, as indicated by the existence of moments.

4.3.2 Limiting ratios

An indication that one distribution has a heavier tail than another is that the ratio of the two survival functions should diverge to infinity (with the heavier tailed distribution in the numerator). This implies that the numerator distribution puts significantly more probability on large values. It is equivalent to examine the ratio of density functions. The limit will be the same, as can be seen by an application of L'Hôpital's rule:

$$\lim_{x \rightarrow \infty} \frac{S_1(x)}{S_2(x)} = \lim_{x \rightarrow \infty} \frac{S_1'(x)}{S_2'(x)} = \lim_{x \rightarrow \infty} \frac{-f_1(x)}{-f_2(x)}.$$

Example 4.16 Demonstrate that the Pareto distribution has a heavier tail than the gamma distribution using the limit of the ratio of their density functions.

To avoid confusion, the letters τ and λ will be used for the parameters of the gamma distribution instead of the customary α and θ . Then the required limit is

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{f_{\text{Pareto}}(x)}{f_{\text{gamma}}(x)} &= \lim_{x \rightarrow \infty} \frac{\alpha \theta^\alpha (x + \theta)^{-\alpha-1}}{x^{\tau-1} e^{-x/\lambda} \lambda^{-\tau} \Gamma(\tau)^{-1}} \\ &= c \lim_{x \rightarrow \infty} \frac{e^{x/\lambda}}{(x + \theta)^{\alpha+1} x^{\tau-1}} \\ &> c \lim_{x \rightarrow \infty} \frac{e^{x/\lambda}}{(x + \theta)^{\alpha+\tau}} \end{aligned}$$

and, either by application of L'Hôpital's rule or by remembering that exponentials go to infinity faster than polynomials, the limit is infinity. Figure 4.3 shows a portion of the density functions for a Pareto distribution with parameters $\alpha = 3$ and $\theta = 10$ and a gamma distribution with parameters $\alpha = \frac{1}{3}$ and $\theta = 15$. Both distributions have a mean of 5 and a variance of 75. The graph is consistent with the algebraic derivation. \square

4.3.3 Hazard rate and mean residual life patterns

The nature of the hazard rate function also reveals information about the tail of the distribution. If the hazard rate function is decreasing, then at large values the chance of that value becomes small and the chance of larger values becomes greater. Thus the distribution will have a heavier tail. Conversely, if the hazard rate function is increasing, a lighter tail is expected.

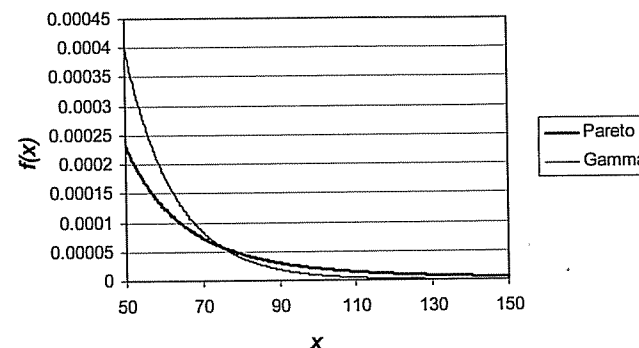


Fig. 4.3 Tails of gamma and Pareto distributions.

Example 4.17 Compare the tails of the Pareto and gamma distributions by looking at their hazard rate functions.

The hazard rate function for the Pareto distribution is

$$h(x) = \frac{f(x)}{S(x)} = \frac{\alpha \theta^\alpha (x + \theta)^{-\alpha-1}}{\theta^\alpha (x + \theta)^{-\alpha}} = \frac{\alpha}{x + \theta}$$

which is decreasing. For the gamma distribution we need to be a bit more clever because there is no closed form expression for $S(x)$. Observe that

$$\frac{1}{h(x)} = \frac{\int_x^\infty f(t) dt}{f(x)} = \frac{\int_0^\infty f(x+y) dy}{f(x)}$$

and so, if $f(x+y)/f(x)$ is an increasing function of x for any fixed y , then $1/h(x)$ will be increasing in x and so the random variable will have a decreasing hazard rate. Now, for the gamma distribution

$$\frac{f(x+y)}{f(x)} = \frac{(x+y)^{\alpha-1} e^{-(x+y)/\theta}}{x^{\alpha-1} e^{-x/\theta}} = \left(1 + \frac{y}{x}\right)^{\alpha-1} e^{-y/\theta},$$

which is strictly increasing in x provided $\alpha < 1$ and strictly decreasing in x if $\alpha > 1$. By this measure, some gamma distributions have a heavy tail (those with $\alpha < 1$) and some have a light tail. Note that when $\alpha = 1$ we have the exponential distribution and a constant hazard rate. Also, even though $h(x)$ is complicated in the gamma case, we know what happens for large x . Because $f(x)$ and $S(x)$ both go to 0 as $x \rightarrow \infty$, L'Hôpital's rule yields

$$\begin{aligned} \lim_{x \rightarrow \infty} h(x) &= \lim_{x \rightarrow \infty} \frac{f(x)}{S(x)} = - \lim_{x \rightarrow \infty} \frac{f'(x)}{f(x)} = - \lim_{x \rightarrow \infty} \left[\frac{d}{dx} \ln f(x) \right] \\ &= - \lim_{x \rightarrow \infty} \frac{d}{dx} \left[(\alpha - 1) \ln x - \frac{x}{\theta} \right] = \lim_{x \rightarrow \infty} \left(\frac{1}{\theta} - \frac{\alpha - 1}{x} \right) = \frac{1}{\theta}. \end{aligned}$$

That is, $h(x) \rightarrow 1/\theta$ as $x \rightarrow \infty$. \square

The mean residual life also gives information about tail weight. If the mean residual life function is increasing in d , then at large values the expected outcome is much larger and thus probability is moved to the right, indicating a heavier tail than a model where the mean residual life function is decreasing or is increasing at a slower rate. In fact, the mean residual life function and the hazard rate are closely related in several ways. First, note that

$$\begin{aligned} \frac{S(y+d)}{S(d)} &= \frac{\exp\left[-\int_0^{y+d} h(x)dx\right]}{\exp\left[-\int_0^d h(x)dx\right]} = \exp\left[-\int_d^{y+d} h(x)dx\right] \\ &= \exp\left[-\int_0^y h(d+t)dt\right]. \end{aligned}$$

Therefore, if the hazard rate is decreasing, then for fixed y it follows that $\int_0^y h(d+t)dt$ is a decreasing function of d , and from the above $S(y+d)/S(d)$ is an increasing function of d . But from (3.5), the mean residual life function may be expressed as

$$e(d) = \frac{\int_d^\infty S(x)dx}{S(d)} = \int_0^\infty \frac{S(y+d)}{S(d)} dy.$$

Thus, if the hazard rate is a decreasing function, then the mean residual life function $e(d)$ is an increasing function of d because the same is true of $S(y+d)/S(d)$ for fixed y . Similarly, if the hazard rate is an increasing function, then the mean residual life function is a decreasing function. It is worth noting (and is perhaps counterintuitive), however, that the converse implication is not true. Exercise 4.16 gives an example of a distribution which has a decreasing mean residual life function, but the hazard rate is not increasing for all values. Nevertheless, the implications described above are generally consistent with the above discussions of heaviness of the tail.

There is a second relationship between the mean residual life function and the hazard rate. As $d \rightarrow \infty$, $S(d)$ and $\int_d^\infty S(x)dx$ go to 0. Thus, the limiting behavior of the mean residual life function as $d \rightarrow \infty$ may be ascertained via L'Hôpital's rule because (3.5) holds. We have

$$\lim_{d \rightarrow \infty} e(d) = \lim_{d \rightarrow \infty} \frac{\int_d^\infty S(x)dx}{S(d)} = \lim_{d \rightarrow \infty} \frac{-S(d)}{-f(d)} = \lim_{d \rightarrow \infty} \frac{1}{h(d)}$$

as long as the indicated limits exist. These limiting relationships are useful if $S(x)$ [and hence also $h(x)$ and $e(d)$] is complicated.

Example 4.18 *Examine the behavior of the mean residual life function of the gamma distribution.*

Because $e(d) = \int_d^\infty S(x)dx/S(d)$ and $S(x)$ is complicated, $e(d)$ is complicated. But $e(0) = E(X) = \alpha\theta$ from Appendix A, and, using Example 4.17, we have

$$\lim_{x \rightarrow \infty} e(x) = \lim_{x \rightarrow \infty} \frac{1}{h(x)} = \frac{1}{\lim_{x \rightarrow \infty} h(x)} = \theta.$$

Also, from Example 4.17, $h(x)$ is strictly decreasing in x for $\alpha < 1$ and strictly increasing in x for $\alpha > 1$, implying that $e(d)$ is strictly increasing from $e(0) = \alpha\theta$ to $e(\infty) = \theta$ for $\alpha < 1$ and strictly decreasing from $e(0) = \alpha\theta$ to $e(\infty) = \theta$ for $\alpha > 1$. For $\alpha = 1$, we have the exponential distribution for which $e(d) = \theta$. \square

Further insight into the mean residual lifetime and the heaviness of the tail may be obtained by introducing the so-called equilibrium distribution, which has important applications in connection with the continuous time ruin model of Chapter 8. For positive random variables with $S(0) = 1$, it follows from Definition 3.6 and (3.5) with $d = 0$ that $E(X) = \int_0^\infty S(x)dx$, or equivalently $1 = \int_0^\infty [S(x)/E(X)]dx$, so that

$$f_e(x) = \frac{S(x)}{E(X)}, \quad x \geq 0, \quad (4.2)$$

is a probability density function. The corresponding survival function is

$$S_e(x) = \int_x^\infty f_e(t)dt = \frac{\int_x^\infty S(t)dt}{E(X)}, \quad x \geq 0.$$

The hazard rate corresponding to the equilibrium distribution is

$$h_e(x) = \frac{f_e(x)}{S_e(x)} = \frac{S(x)}{\int_x^\infty S(t)dt} = \frac{1}{e(x)}$$

using (3.5). Thus, the reciprocal of the mean residual life function is itself a hazard rate, and this fact may be used to show that the mean residual life function uniquely characterizes the original distribution. We have

$$f_e(x) = h_e(x)S_e(x) = h_e(x)e^{-\int_0^x h_e(t)dt},$$

or equivalently

$$S(x) = \frac{e(0)}{e(x)} e^{-\int_0^x \{1/e(t)\} dt}$$

using $e(0) = E(X)$.

The equilibrium distribution also provides further insight into the relationship between the hazard rate, the mean residual life function, and heaviness of the tail. Assuming that $S(0) = 1$, and thus $e(0) = E(X)$, we have

$\int_x^\infty S(t)dt = e(0)S_e(x)$, and from (3.5), $\int_x^\infty S(t)dt = e(x)S(x)$. Equating these two expressions results in

$$\frac{e(x)}{e(0)} = \frac{S_e(x)}{S(x)}.$$

If the mean residual life function is increasing (implied if the hazard rate is decreasing), then $e(x) \geq e(0)$, which is obviously equivalent to $S_e(x) \geq S(x)$ from the above equality. This in turn implies that

$$\int_0^\infty S_e(x)dx \geq \int_0^\infty S(x)dx.$$

But $E(X) = \int_0^\infty S(x)dx$ from Definition 3.6 and (3.5) if $S(0) = 1$. Also, $\int_0^\infty S_e(x)dx = \int_0^\infty xf_e(x)dx$ since both sides represent the mean of the equilibrium distribution. This may be evaluated using (3.9) with $u = \infty, k = 2$, and $F(0) = 0$ to give the equilibrium mean, that is,

$$\int_0^\infty S_e(x)dx = \int_0^\infty xf_e(x)dx = \frac{1}{E(X)} \int_0^\infty xS(x)dx = \frac{E(X^2)}{2E(X)}.$$

The inequality may thus be expressed as

$$\frac{E(X^2)}{2E(X)} \geq E(X),$$

or using $\text{Var}(X) = E(X^2) - \{E(X)\}^2$ as $\text{Var}(X) \geq \{E(X)\}^2$. That is, the squared coefficient of variation, and hence the coefficient of variation itself, is at least 1 if $e(x) \geq e(0)$. Reversing the inequalities implies that the coefficient of variation is at most 1 if $e(x) \leq e(0)$, in turn implied if the mean residual life function is decreasing or the hazard rate is increasing. These values of the coefficient of variation are consistent with the comments made here about the heaviness of the tail.

4.3.4 Exercises

4.13 Using the methods in this section (except for the mean residual life), compare the tail weight of the Weibull and inverse Weibull distributions.

4.14 Arguments as in Example 4.16 place the lognormal distribution between the gamma and Pareto distributions with regard to tail weight. To reinforce this conclusion, consider a gamma distribution with parameters $\alpha = 0.2, \theta = 500$; a lognormal distribution with parameters $\mu = 3.709290, \sigma = 1.338566$; and a Pareto distribution with parameters $\alpha = 2.5, \theta = 150$. First demonstrate that all three distributions have the same mean and variance. Then numerically demonstrate that there is a value such that the gamma pdf is smaller than the lognormal and Pareto pdfs for all arguments above that

value and that there is another value such that the lognormal pdf is smaller than the Pareto pdf for all arguments above that value.

4.15 Let Y be a random variable that has the equilibrium density from (4.2). That is, $f_Y(y) = f_e(y) = S_X(y)/E(X)$ for some random variable X . Use integration by parts to show that

$$M_Y(t) = \frac{M_X(t) - 1}{tE(X)}$$

whenever $M_X(t)$ exists.

4.16 You are given that the random variable X has probability density function $f(x) = (1 + 2x^2)e^{-2x}, x \geq 0$.

- Determine the survival function $S(x)$.
- Determine the hazard rate $h(x)$.
- Determine the survival function $S_e(x)$ of the equilibrium distribution.
- Determine the mean residual life function $e(x)$.
- Determine $\lim_{x \rightarrow \infty} h(x)$ and $\lim_{x \rightarrow \infty} e(x)$.
- Prove that $e(x)$ is strictly decreasing but $h(x)$ is not strictly increasing.

4.17 Assume that X has probability density function $f(x), x \geq 0$.

- Prove that

$$S_e(x) = \frac{\int_x^\infty (y-x)f(y)dy}{E(X)}.$$

- Use (a) to show that

$$\int_x^\infty yf(y)dy = xS(x) + E(X)S_e(x).$$

- Prove that (b) may be rewritten as

$$S(x) = \frac{\int_x^\infty yf(y)dy}{x + e(x)}$$

and that this in turn implies that

$$S(x) \leq \frac{E(X)}{x + e(x)}.$$

(d) Use (c) to prove that, if $e(x) \geq e(0)$, then

$$S(x) \leq \frac{E(X)}{x + E(X)}$$

and that this in turn implies that the mean is at least as large as the (smallest) median.

(e) Prove that (b) may be rewritten as

$$S_e(x) = \frac{e(x)}{x + e(x)} \cdot \frac{\int_x^\infty y f(y) dy}{E(X)}$$

and thus that

$$S_e(x) \leq \frac{e(x)}{x + e(x)}.$$

4.4 CREATING NEW DISTRIBUTIONS

4.4.1 Introduction

This section indicates how new parametric distributions can be created from existing ones. Many of the distributions in Appendix A were created this way.

4.4.2 Multiplication by a constant

This transformation is equivalent to applying inflation uniformly across all loss levels and is known as a change of scale. For example, if this year's losses are given by the random variable X , then uniform inflation of 5% indicates that next year's losses can be modeled with the random variable $Y = 1.05X$.

Theorem 4.19 Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$. Let $Y = \theta X$ with $\theta > 0$. Then

$$F_Y(y) = F_X\left(\frac{y}{\theta}\right), \quad f_Y(y) = \frac{1}{\theta} f_X\left(\frac{y}{\theta}\right).$$

Proof:

$$F_Y(y) = \Pr(Y \leq y) = \Pr(\theta X \leq y) = \Pr\left(X \leq \frac{y}{\theta}\right) = F_X\left(\frac{y}{\theta}\right)$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\theta} f_X\left(\frac{y}{\theta}\right). \quad \square$$

Corollary 4.20 The parameter θ is a scale parameter for the random variable Y .

The following example illustrates this process.

Example 4.21 Let X have pdf $f(x) = e^{-x}$, $x > 0$. Determine the cdf and pdf of $Y = \theta X$.

$$\begin{aligned} F_X(x) &= 1 - e^{-x}, & F_Y(y) &= 1 - e^{-y/\theta}, \\ f_Y(y) &= \frac{1}{\theta} e^{-y/\theta}. \end{aligned}$$

We recognize this as the exponential distribution. \square

4.4.3 Raising to a power

Theorem 4.22 Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $F_X(0) = 0$. Let $Y = X^{1/\tau}$. Then, if $\tau > 0$,

$$F_Y(y) = F_X(y^\tau), \quad f_Y(y) = \tau y^{\tau-1} f_X(y^\tau), \quad y > 0$$

while, if $\tau < 0$,

$$F_Y(y) = 1 - F_X(y^\tau), \quad f_Y(y) = -\tau y^{\tau-1} f_X(y^\tau). \quad (4.3)$$

Proof: If $\tau > 0$

$$F_Y(y) = \Pr(X \leq y^\tau) = F_X(y^\tau),$$

while if $\tau < 0$

$$F_Y(y) = \Pr(X \geq y^\tau) = 1 - F_X(y^\tau).$$

The pdf follows by differentiation. \square

It is more common to keep parameters positive and so, when τ is negative, create a new parameter $\tau^* = -\tau$. Then (4.3) becomes

$$F_Y(y) = 1 - F_X(y^{-\tau^*}), \quad f_Y(y) = \tau^* y^{-\tau^*-1} f_X(y^{-\tau^*}).$$

Drop the asterisk for future use of this positive parameter.

Definition 4.23 When raising a distribution to a power, if $\tau > 0$ the resulting distribution is called **transformed**, if $\tau = -1$ it is called **inverse**, and if $\tau < 0$ (but is not -1) it is called **inverse transformed**. To create the distributions in Appendix A and to retain θ as a scale parameter, the base distribution should be raised to a power before being multiplied by θ .

Example 4.24 Suppose X has the exponential distribution. Determine the cdf of the inverse, transformed, and inverse transformed exponential distributions.

The inverse exponential distribution with no scale parameter has cdf

$$F(y) = 1 - [1 - e^{-1/y}] = e^{-1/y}.$$

With the scale parameter added it is $F(y) = e^{-\theta/y}$.

The transformed exponential distribution with no scale parameter has cdf

$$F(y) = 1 - \exp(-y^\tau).$$

With the scale parameter added it is $F(y) = 1 - \exp[-(y/\theta)^\tau]$. This distribution is more commonly known as the **Weibull distribution**.

The inverse transformed exponential distribution with no scale parameter has cdf

$$F(y) = 1 - [1 - \exp(-y^{-\tau})] = \exp(-y^{-\tau}).$$

With the scale parameter added it is $F(y) = \exp[-(\theta/y)^\tau]$. This distribution is the **inverse Weibull**. \square

Another base distribution has pdf $f(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$. When a scale parameter is added, this becomes the **gamma distribution**. It has inverse and transformed versions that can be created using the results in this section. Unlike the distributions introduced to this point, this one does not have a closed form cdf. The best we can do is define notation for the function.

Definition 4.25 The *incomplete gamma function* with parameter $\alpha > 0$ is denoted and defined by

$$\Gamma(\alpha; x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt$$

while the *gamma function* is denoted and defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

In addition, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ and for positive integer values of n , $\Gamma(n) = (n-1)!$. Appendix A provides details on numerical methods of evaluating these quantities. Furthermore, these functions are built into most spreadsheet programs and many statistical and numerical analysis programs.

4.4.4 Exponentiation

Theorem 4.26 Let X be a continuous random variable with pdf $f_X(x)$ and cdf $F_X(x)$ with $f_X(x) > 0$ for all real x . Let $Y = \exp(X)$. Then, for $y > 0$,

$$F_Y(y) = F_X(\ln y), \quad f_Y(y) = \frac{1}{y} f_X(\ln y).$$

Proof: $F_Y(y) = \Pr(e^X \leq y) = \Pr(X \leq \ln y) = F_X(\ln y)$. \square

Example 4.27 Let X have the normal distribution with mean μ and variance σ^2 . Determine the cdf and pdf of $Y = e^X$.

$$F_Y(y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$$

$$f_Y(y) = \frac{1}{y\sigma} \phi\left(\frac{\ln y - \mu}{\sigma}\right) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2\right]. \quad \square$$

We could try to add a scale parameter by creating $W = \theta Y$, but this adds no value, as is demonstrated in Exercise 4.22. This example created the **lognormal** distribution (the name has stuck even though “expnormal” would seem more descriptive).

4.4.5 Mixing

The concept of mixing can be extended from mixing a finite number of random variables to mixing an uncountable number. In the following theorem, the pdf $f_\Lambda(\lambda)$ plays the role of the discrete probabilities a_j in the k -point mixture.

Theorem 4.28 Let X have pdf $f_{X|\Lambda}(x|\lambda)$ and cdf $F_{X|\Lambda}(x|\lambda)$, where λ is a parameter of X . While X may have other parameters, they are not relevant. Let λ be a realization of the random variable Λ with pdf $f_\Lambda(\lambda)$. Then the unconditional pdf of X is

$$f_X(x) = \int f_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda) d\lambda, \quad (4.4)$$

where the integral is taken over all values of λ with positive probability. The resulting distribution is a **mixture distribution**. The distribution function can be determined from

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x \int f_{X|\Lambda}(y|\lambda) f_\Lambda(\lambda) d\lambda dy \\ &= \int \int_{-\infty}^x f_{X|\Lambda}(y|\lambda) f_\Lambda(\lambda) dy d\lambda \\ &= \int F_{X|\Lambda}(x|\lambda) f_\Lambda(\lambda) d\lambda. \end{aligned}$$

Moments of the mixture distribution can be found from

$$E(X^k) = E[E(X^k|\Lambda)]$$

and, in particular,

$$\text{Var}(X) = E[\text{Var}(X|\Lambda)] + \text{Var}[E(X|\Lambda)].$$

Proof: The integrand is, by definition, the joint density of X and Λ . The integral is then the marginal density. For the expected value (assuming the order of integration can be reversed),

$$\begin{aligned} E(X^k) &= \int \int x^k f_{X|\Lambda}(x|\lambda) f_{\Lambda}(\lambda) d\lambda dx \\ &= \int \left[\int x^k f_{X|\Lambda}(x|\lambda) dx \right] f_{\Lambda}(\lambda) d\lambda \\ &= \int E(X^k|\lambda) f_{\Lambda}(\lambda) d\lambda \\ &= E[E(X^k|\Lambda)]. \end{aligned}$$

For the variance,

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= E[E(X^2|\Lambda)] - \{E[E(X|\Lambda)]\}^2 \\ &= E\{\text{Var}(X|\Lambda) + [E(X|\Lambda)]^2\} - \{E[E(X|\Lambda)]\}^2 \\ &= E[\text{Var}(X|\Lambda)] + \text{Var}[E(X|\Lambda)]. \end{aligned}$$

Note that, if $f_{\Lambda}(\lambda)$ is discrete, the integrals must be replaced with sums. An alternative way to write the results is $f_X(x) = E_{\Lambda}[f_{X|\Lambda}(x|\Lambda)]$ and $F_X(x) = E_{\Lambda}[F_{X|\Lambda}(x|\Lambda)]$, where the subscript on E indicates that the random variable is Λ .

An interesting phenomenon is that mixture distributions tend to be heavy-tailed so this method is a good way to generate such a model. In particular, if $f_{X|\Lambda}(x|\lambda)$ has a decreasing hazard rate function for all λ , then the mixture distribution will also have a decreasing hazard rate function (see Ross [114], pp. 407–409 for details). The following example shows how a familiar heavy-tailed distribution may be obtained by mixing.

Example 4.29 Let $X|\Lambda$ have an exponential distribution with parameter $1/\Lambda$. Let Λ have a gamma distribution. Determine the unconditional distribution of X .

We have (note that the parameter θ in the gamma distribution has been replaced by its reciprocal)

$$\begin{aligned} f_X(x) &= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \lambda e^{-\lambda x} \lambda^{\alpha-1} e^{-\theta \lambda} d\lambda \\ &= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} \lambda^{\alpha} e^{-\lambda(x+\theta)} d\lambda \\ &= \frac{\theta^{\alpha}}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(x+\theta)^{\alpha+1}} \\ &= \frac{\alpha \theta^{\alpha}}{(x+\theta)^{\alpha+1}}. \end{aligned}$$

This is a Pareto distribution. □

The following example provides an illustration useful in Chapter 16.

Example 4.30 Suppose that, given $\Theta = \theta$, X is normally distributed with mean θ and variance v , so that

$$f_{X|\Theta}(x|\theta) = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{1}{2v}(x-\theta)^2\right], \quad -\infty < x < \infty,$$

and Θ is itself normally distributed with mean μ and variance a , that is,

$$f_{\Theta}(\theta) = \frac{1}{\sqrt{2\pi a}} \exp\left[-\frac{1}{2a}(\theta-\mu)^2\right], \quad -\infty < \theta < \infty.$$

Determine the marginal pdf of X .

The marginal pdf of X is

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{1}{2v}(x-\theta)^2\right] \frac{1}{\sqrt{2\pi a}} \exp\left[-\frac{1}{2a}(\theta-\mu)^2\right] d\theta \\ &= \frac{1}{2\pi\sqrt{va}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2v}(x-\theta)^2 - \frac{1}{2a}(\theta-\mu)^2\right] d\theta. \end{aligned}$$

We leave as an exercise for the reader the verification of the algebraic identity

$$\frac{(x-\theta)^2}{v} + \frac{(\theta-\mu)^2}{a} = \frac{a+v}{va} \left(\theta - \frac{ax+v\mu}{a+v} \right)^2 + \frac{(x-\mu)^2}{a+v}$$

obtained by completion of the square in θ . Thus,

$$f_X(x) = \frac{\exp\left[-\frac{(x-\mu)^2}{2(a+v)}\right]}{\sqrt{2\pi(a+v)}} \int_{-\infty}^{\infty} \sqrt{\frac{a+v}{2\pi va}} \exp\left[-\frac{a+v}{2va} \left(\theta - \frac{ax+v\mu}{a+v} \right)^2\right] d\theta.$$

We recognize the integrand as the pdf (as a function of θ) of a normal distribution with mean $(ax+v\mu)/(a+v)$ and variance $(va)/(a+v)$. Thus the integral is 1 and so

$$f_X(x) = \frac{\exp\left[-\frac{(x-\mu)^2}{2(a+v)}\right]}{\sqrt{2\pi(a+v)}}, \quad -\infty < x < \infty;$$

that is, X is normal with mean μ and variance $a+v$. □

The following example is taken from Hayne [50]. It illustrates how this type of mixture distribution can arise. In particular, continuous mixtures are

often used to provide a model for parameter uncertainty. That is, the exact value of a parameter is not known, but a probability density function can be elucidated to describe possible values of that parameter.

Example 4.31 *In the valuation of warranties on automobiles it is important to recognize that the number of miles driven varies from driver to driver. It is also the case that for a particular driver the number of miles varies from year to year. Suppose the number of miles for a randomly selected driver has the inverse Weibull distribution but that the year-to-year variation in the scale parameter has the transformed gamma distribution with the same value for τ . Determine the distribution for the number of miles driven in a randomly selected year by a randomly selected driver.*

Using the parameterizations from Appendix A, the inverse Weibull for miles driven in a year has parameters Λ (in place of Θ) and τ while the transformed gamma distribution for the scale parameter Λ has parameters τ , θ , and α . The marginal density is

$$\begin{aligned} f(x) &= \int_0^\infty \frac{\tau \lambda^\tau}{x^{\tau+1}} e^{-(\lambda/x)^\tau} \frac{\tau \lambda^{\tau\alpha-1}}{\theta^{\tau\alpha} \Gamma(\alpha)} e^{-(\lambda/\theta)^\tau} d\lambda \\ &= \frac{\tau^2}{\theta^{\tau\alpha} \Gamma(\alpha) x^{\tau+1}} \int_0^\infty \lambda^{\tau+\tau\alpha-1} \exp[-\lambda^\tau (x^{-\tau} + \theta^{-\tau})] d\lambda \\ &= \frac{\tau^2}{\theta^{\tau\alpha} \Gamma(\alpha) x^{\tau+1}} \int_0^\infty [y^{1/\tau} (x^{-\tau} + \theta^{-\tau})^{-1/\tau}]^{\tau+\tau\alpha-1} e^{-y} \\ &\quad \times y^{\tau^{-1}-1} \tau^{-1} (x^{-\tau} + \theta^{-\tau})^{-1/\tau} dy \\ &= \frac{\tau}{\theta^{\tau\alpha} \Gamma(\alpha) x^{\tau+1} (x^{-\tau} + \theta^{-\tau})^{\alpha+1}} \int_0^\infty y^\alpha e^{-y} dy \\ &= \frac{\tau \Gamma(\alpha+1)}{\theta^{\tau\alpha} \Gamma(\alpha) x^{\tau+1} (x^{-\tau} + \theta^{-\tau})^{\alpha+1}} \\ &= \frac{\tau \alpha \theta^\tau x^{\tau\alpha-1}}{(x^\tau + \theta^\tau)^{\alpha+1}}. \end{aligned}$$

The third line is obtained by the transformation $y = \lambda^\tau (x^{-\tau} + \theta^{-\tau})$. The final line uses the fact that $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$. The result is an inverse Burr distribution. Note that this distribution applies to a particular driver. Another driver may have a different Weibull shape parameter τ and as well that driver's Weibull scale parameter Θ may have a different distribution and, in particular, a different mean. \square

4.4.6 Frailty models

An important type of mixture distribution is a frailty model. Although the physical motivation for this particular type of mixture is originally from the

analysis of lifetime distributions in survival analysis, the resulting mathematical convenience implies that the approach may also be viewed as a useful way to generate new distributions by mixing.

We begin by introducing a *frailty* random variable $\Lambda > 0$ and define the conditional hazard rate (given $\Lambda = \lambda$) of X to be $h_{X|\Lambda}(x|\lambda) = \lambda a(x)$, where $a(x)$ is a known function of x [that is, $a(x)$ is to be specified in a particular application]. The frailty is meant to quantify uncertainty associated with the hazard rate, which by the above specification of the conditional hazard rate acts in a multiplicative manner.

The conditional survival function of $X|\Lambda$ is therefore

$$S_{X|\Lambda}(x|\lambda) = e^{-\int_0^x h_{X|\Lambda}(t|\lambda) dt} = e^{-\lambda A(x)},$$

where $A(x) = \int_0^x a(t) dt$. In order to specify the mixture distribution (that is, the marginal distribution of X), we define the moment generating function of the frailty random variable Λ to be $M_\Lambda(t) = E(e^{t\Lambda})$. Then, the marginal survival function is

$$S_X(x) = E[e^{-\Lambda A(x)}] = M_\Lambda[-A(x)], \quad (4.5)$$

and obviously $F_X(x) = 1 - S_X(x)$.

The type of mixture to be used determines the choice of $a(x)$ and hence $A(x)$. The most important subclass of the frailty models is the class of exponential mixtures with $a(x) = 1$ and $A(x) = x$, so that $S_{X|\Lambda}(x|\lambda) = e^{-\lambda x}$, $x \geq 0$. Other useful mixtures include Weibull mixtures with $a(x) = \gamma x^{\gamma-1}$ and $A(x) = x^\gamma$.

Evaluation of the frailty distribution requires an expression for the moment generating function $M_\Lambda(t)$ of Λ . The most common choice is gamma frailty, but other choices such as inverse Gaussian frailty are also used.

Example 4.32 *Let Λ have a gamma distribution and let $X|\Lambda$ have a Weibull distribution with conditional survival function $S_{X|\Lambda}(x|\lambda) = e^{-\lambda x^\gamma}$. Determine the unconditional or marginal distribution of X .*

\square In this case it follows from Example 3.15 that the gamma moment generating function is $M_\Lambda(t) = (1 - \theta t)^{-\alpha}$, and from (4.5) it follows that X has survival function

$$S_X(x) = M_\Lambda(-x^\gamma) = (1 + \theta x^\gamma)^{-\alpha}.$$

This is a Burr distribution (see Appendix A) with the usual parameter θ replaced by $\theta^{-1/\gamma}$. Note that when $\gamma = 1$ this is an exponential mixture which is a Pareto distribution, considered previously in Example 4.29. \square

As mentioned earlier, mixing tends to create heavy-tailed distributions, and in particular a mixture of distributions which all have decreasing hazard rates also has a decreasing hazard rate. In Exercise 4.32 the reader is asked to prove this fact for frailty models. For further details on frailty models, see the book by Hougaard [63].

4.4.7 Splicing

Another method for creating a new distribution is by splicing. This approach is similar to mixing in that it might be believed that two or more separate processes are responsible for generating the losses. With mixing, the various processes operate on subsets of the population. Once the subset is identified, a simple loss model suffices. For splicing, the processes differ with regard to the loss amount. That is, one model governs the behavior of losses in some interval of possible losses while other models cover the other intervals. The following definition makes this precise.

Definition 4.33 A k -component spliced distribution has a density function that can be expressed as follows:

$$f_X(x) = \begin{cases} a_1 f_1(x), & c_0 < x < c_1, \\ a_2 f_2(x), & c_1 < x < c_2, \\ \vdots & \vdots \\ a_k f_k(x), & c_{k-1} < x < c_k. \end{cases}$$

For $j = 1, \dots, k$, each $a_j > 0$ and each $f_j(x)$ must be a legitimate density function with all probability on the interval (c_{j-1}, c_j) . Also, $a_1 + \dots + a_k = 1$.

Example 4.34 Demonstrate that Model 5 on Page 21 is a two-component spliced model.

The density function is

$$f(x) = \begin{cases} 0.01, & 0 \leq x < 50, \\ 0.02, & 50 \leq x < 75 \end{cases}$$

and the spliced model is created by letting $f_1(x) = 0.02$, $0 \leq x < 50$, which is a uniform distribution on the interval from 0 to 50, and $f_2(x) = 0.04$, $50 \leq x < 75$, which is a uniform distribution on the interval from 50 to 75. The coefficients are then $a_1 = 0.5$ and $a_2 = 0.5$. \square

It was not necessary to use density functions and coefficients, but this is one way to ensure that the result is a legitimate density function. When using parametric models, the motivation for splicing is that the tail behavior may be inconsistent with the behavior for small losses. For example, experience (based on knowledge beyond that available in the current, perhaps small, data set) may indicate that the tail follows the Pareto distribution, but there is a positive mode more in keeping with the lognormal or inverse Gaussian distributions. A second instance is when there is a large amount of data below some value but a limited amount of information elsewhere. We may want to use the empirical distribution (or a smoothed version of it) up to a certain point and a parametric model beyond that value. The definition given above is appropriate when the break points c_0, \dots, c_k are known in advance.

Another way to construct a spliced model is to use standard distributions over the range from c_0 to c_k . Let $g_j(x)$ be the j th such density function. Then, in Definition 4.33 replace $f_j(x)$ with $g_j(x)/[G(c_j) - G(c_{j-1})]$. This formulation makes it easier to have the break points become parameters that can be estimated.

Neither approach to splicing ensures that the resulting density function will be continuous (that is, the components will meet at the break points). Such a restriction could be added to the specification.

Example 4.35 Create a two-component spliced model using an exponential distribution from 0 to c and a Pareto distribution (using γ in place of θ) from c to ∞ .

The basic format is

$$f_X(x) = \begin{cases} \frac{\theta^{-1} e^{-x/\theta}}{1 - e^{-c/\theta}}, & 0 < x < c, \\ \frac{\alpha \gamma^\alpha (x + \gamma)^{-\alpha-1}}{\gamma^\alpha (c + \gamma)^{-\alpha}}, & c < x < \infty. \end{cases}$$

However, we must force the density function to integrate to 1. All that is needed is to let $a_1 = v$ and $a_2 = 1 - v$. The spliced density function becomes

$$f_X(x) = \begin{cases} v \frac{\theta^{-1} e^{-x/\theta}}{1 - e^{-c/\theta}}, & 0 < x < c, \\ (1 - v) \frac{\alpha (c + \gamma)^\alpha}{(x + \gamma)^{\alpha+1}}, & c < x < \infty \end{cases}, \quad \theta, \alpha, \gamma, c > 0, 0 < v < 1.$$

Figure 4.4 illustrates this density function using the values $c = 100$, $v = 0.6$, $\theta = 100$, $\gamma = 200$, and $\alpha = 4$. It is clear that this density is not continuous. \square

4.4.8 Exercises

4.18 Let X have cdf $F_X(x) = 1 - (1 + x)^{-\alpha}$, $x, \alpha > 0$. Determine the pdf and cdf of $Y = \theta X$.

4.19 (*) One hundred observed claims in 1995 were arranged as follows: 42 were between 0 and 300, 3 were between 300 and 350, 5 were between 350 and 400, 5 were between 400 and 450, 0 were between 450 and 500, 5 were between 500 and 600, and the remaining 40 were above 600. For the next three years, all claims are inflated by 10% per year. Based on the empirical distribution from 1995, determine a range for the probability that a claim exceeds 500 in 1998 (there is not enough information to determine the probability exactly).

4.20 Let X have the Pareto distribution. Determine the cdf of the transformed, inverse, and inverse transformed distributions. Check Appendix A to determine if any of these distributions have special names.

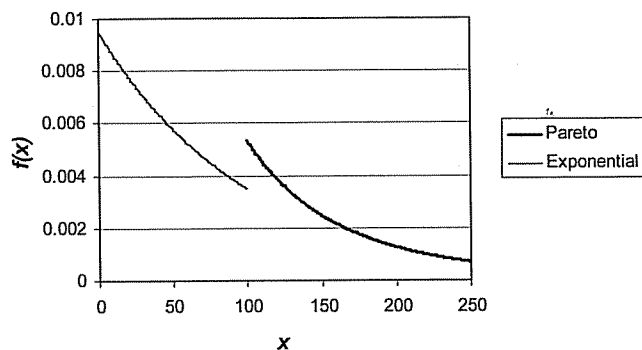


Fig. 4.4 Two-component spliced density.

4.21 Let X have the loglogistic distribution. Demonstrate that the inverse distribution also has the loglogistic distribution. Therefore there is no need to identify a separate inverse loglogistic distribution.

4.22 Let Y have the lognormal distribution with parameters μ and σ . Let $Z = \theta Y$. Show that Z also has the lognormal distribution and therefore the addition of a third parameter has not created a new distribution.

4.23 (*) Let X have a Pareto distribution with parameters α and θ . Let $Y = \ln(1 + X/\theta)$. Determine the name of the distribution of Y and its parameters.

4.24 In [132], Venter noted that if X has the transformed gamma distribution and its scale parameter θ has an inverse transformed gamma distribution (where the parameter τ is the same in both distributions) the resulting mixture has the transformed beta distribution. Demonstrate that this is true.

4.25 (*) Let N have a Poisson distribution with mean Λ . Let Λ have a gamma distribution with mean 1 and variance 2. Determine the unconditional probability that $N = 1$.

4.26 (*) Given a value of $\Theta = \theta$, the random variable X has an exponential distribution with hazard rate function $h(x) = \theta$, a constant. The random variable Θ has a uniform distribution on the interval $(1, 11)$. Determine $S_X(0.5)$ for the unconditional distribution.

4.27 (*) Let N have a Poisson distribution with mean Λ . Let Λ have a uniform distribution on the interval $(0, 5)$. Determine the unconditional probability that $N \geq 2$.

4.28 Determine the probability density function and the hazard rate of the frailty distribution.

4.29 Suppose that $X|\Lambda$ has the Weibull survival function $S_{X|\Lambda}(x|\lambda) = e^{-\lambda x^\gamma}$, $x \geq 0$, and Λ has an exponential distribution. Demonstrate that the unconditional distribution of X is loglogistic.

4.30 Consider the exponential-inverse Gaussian frailty model with $a(x) = \theta/(2\sqrt{1+\theta x})$, where $\theta > 0$.

- Verify that the conditional hazard rate $h_{X|\Lambda}(x|\lambda)$ of $X|\Lambda$ is indeed a valid hazard rate.
- Determine the conditional survival function $S_{X|\Lambda}(x|\lambda)$.
- If Λ has a gamma distribution with parameters $\theta = 1$ and α replaced by 2α , determine the marginal or unconditional survival function of X .
- Use (c) to argue that a given frailty model may arise from more than one combination of conditional distributions of $X|\Lambda$ and frailty distributions of Λ .

4.31 Suppose that X has survival function $S_X(x) = 1 - F_X(x)$ given by (4.5). Show that $S_1(x) = F_X(x)/[E(\Lambda)A(x)]$ is again a survival function of the form given by (4.5), and identify the distribution of Λ associated with $S_1(x)$.

4.32 Fix $s \geq 0$, and define an "Esscher-transformed" frailty random variable Λ_s with probability density function (or discrete probability mass function in the discrete case) $f_{\Lambda_s}(\lambda) = e^{-s\lambda} f_\Lambda(\lambda)/M_\Lambda(-s)$, $\lambda \geq 0$.

- Show that Λ_s has moment generating function

$$M_{\Lambda_s}(t) = E(e^{t\Lambda_s}) = \frac{M_\Lambda(t-s)}{M_\Lambda(-s)}.$$

- Define the cumulant generating function of Λ to be

$$c_\Lambda(t) = \ln[M_\Lambda(t)],$$

and use (a) to prove that

$$c'_\Lambda(-s) = E(\Lambda_s) \text{ and } c''_\Lambda(-s) = \text{Var}(\Lambda_s).$$

- For the frailty model with survival function given by (4.5), prove that the associated hazard rate may be expressed as $h_X(x) = a(x)c'_\Lambda[-A(x)]$, where c_Λ is defined in (b).
- Use (c) to show that

$$h'_X(x) = a'(x)c'_\Lambda[-A(x)] - [a(x)]^2 c''_\Lambda[-A(x)].$$

- (e) Prove using (d) that, if the conditional hazard rate $h_{X|\Lambda}(x|\lambda)$ is nonincreasing in x , then $h_X(x)$ is also nonincreasing in x .

4.33 Write the density function for a two-component spliced model in which the density function is proportional to a uniform density over the interval from 0 to 1,000 and is proportional to an exponential density function from 1,000 to ∞ . Ensure that the resulting density function is continuous.

4.34 Let X have pdf $f(x) = \exp(-|x/\theta|)/2\theta$ for $-\infty < x < \infty$. Let $Y = e^X$. Determine the pdf and cdf of Y .

4.35 (*) Losses in 1993 follow the density function $f(x) = 3x^{-4}$, $x \geq 1$, where x is the loss in millions of dollars. Inflation of 10% impacts all claims uniformly from 1993 to 1994. Determine the cdf of losses for 1994 and use it to determine the probability that a 1994 loss exceeds 2,200,000.

4.36 Consider the inverse Gaussian random variable X with pdf (from Appendix A)

$$f(x) = \sqrt{\frac{\theta}{2\pi x^3}} \exp\left[-\frac{\theta}{2x} \left(\frac{x-\mu}{\mu}\right)^2\right], \quad x > 0,$$

where $\theta > 0$ and $\mu > 0$ are parameters.

- (a) Derive the pdf of the reciprocal inverse Gaussian random variable $1/X$.
 (b) Prove that the “joint” moment generating function of X and $1/X$ is given by

$$\begin{aligned} M(t_1, t_2) &= E(e^{t_1 X + t_2 X^{-1}}) \\ &= \sqrt{\frac{\theta}{\theta - 2t_2}} \exp\left(\frac{\theta - \sqrt{(\theta - 2\mu^2 t_1)(\theta - 2t_2)}}{\mu}\right), \end{aligned}$$

where $t_1 < \theta/(2\mu^2)$ and $t_2 < \theta/2$.

- (c) Use (b) to show that the moment generating function of X is

$$M_X(t) = E(e^{tX}) = \exp\left[\frac{\theta}{\mu} \left(1 - \sqrt{1 - \frac{2\mu^2}{\theta}t}\right)\right], \quad t < \frac{\theta}{2\mu^2},$$

and that this agrees with the reparameterized result in Exercise 3.24.

- (d) Use (b) to show that the reciprocal inverse Gaussian random variable $1/X$ has moment generating function

$$\begin{aligned} M_{1/X}(t) &= E(e^{tX^{-1}}) \\ &= \sqrt{\frac{\theta}{\theta - 2t}} \exp\left[\frac{\theta}{\mu} \left(1 - \sqrt{1 - \frac{2}{\theta}t}\right)\right], \quad t < \frac{\theta}{2}. \end{aligned}$$

Hence prove that $1/X$ has the same distribution as $Z_1 + Z_2$, where Z_1 has a gamma distribution, Z_2 has an inverse Gaussian distribution, and Z_1 is independent of Z_2 . Also, identify the gamma and inverse Gaussian parameters in this representation.

- (e) Use (b) to show that

$$Z = \frac{1}{X} \left(\frac{X - \mu}{\mu}\right)^2$$

has a gamma distribution with parameters $\alpha = \frac{1}{2}$ and the usual parameter θ (in Appendix A) replaced by $2/\theta$.

4.5 SELECTED DISTRIBUTIONS AND THEIR RELATIONSHIPS

4.5.1 Introduction

There are many ways to organize distributions into groups. Families such as Pearson (12 types), Burr (12 types), Stoppa (5 types), and Dagum (11 types) are discussed in Chapter 2 of [73]. The same distribution can appear in more than one system, indicating that there are many relations among the distributions beyond those presented here. The systems presented in the next subsection are particularly useful for actuarial modeling because all the members have support on the positive real line and all tend to be skewed to the right. For a comprehensive set of continuous distributions, the two volumes by Johnson, Kotz, and Balakrishnan [67] and [68] are a valuable reference. In addition, there are entire books devoted to single distributions (such as Arnold [5] for the Pareto distribution).

4.5.2 Two parametric families

As noted when defining parametric families, many of the distributions presented in this section and in Appendix A are special cases of others. For example, a Weibull distribution with $\tau = 1$ and θ arbitrary is an exponential distribution. Through this process, many of our distributions can be organized into groupings, as illustrated in Figures 4.5 and 4.6. The transformed beta family includes two special cases of a different nature. The paralogistic and inverse paralogistic distributions are created by setting the two nonscale parameters of the Burr and inverse Burr distributions equal to each other rather than to a specified value.

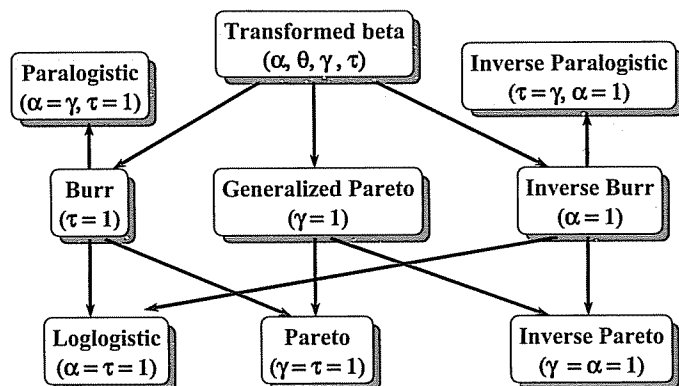


Fig. 4.5 Transformed beta family.

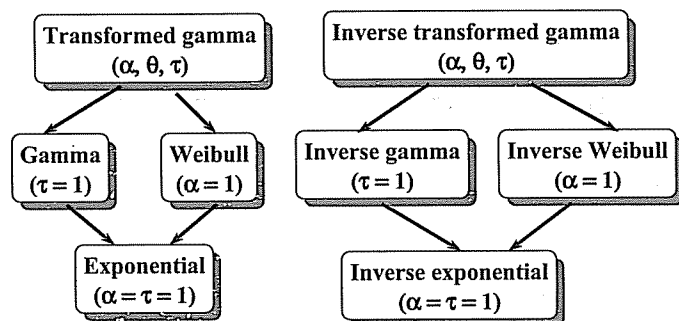


Fig. 4.6 Transformed/inverse transformed gamma family.

4.5.3 Limiting distributions

The classification in the preceding section involved distributions that are special cases of other distributions. Another way to relate distributions is to see what happens as parameters go to their limiting values of zero or infinity.

Example 4.36 Show that the transformed gamma distribution is a limiting case of the transformed beta distribution as $\theta \rightarrow \infty$, $\alpha \rightarrow \infty$, and $\theta/\alpha^{1/\gamma} \rightarrow \xi$, a constant.

The demonstration relies on two facts concerning limits:

$$\lim_{\alpha \rightarrow \infty} \frac{e^{-\alpha} \alpha^{\alpha-1/2} (2\pi)^{1/2}}{\Gamma(\alpha)} = 1 \quad (4.6)$$

and

$$\lim_{a \rightarrow \infty} \left(1 + \frac{x}{a}\right)^{a+b} = e^x. \quad (4.7)$$

The limit in (4.6) is known as Stirling's formula and provides an approximation for the gamma function. The limit in (4.7) is a standard result found in most calculus texts.

To ensure that the ratio $\theta/\alpha^{1/\gamma}$ goes to a constant, it is sufficient to force it to be constant as α and θ become larger and larger. This can be accomplished by substituting $\xi\alpha^{1/\gamma}$ for θ in the transformed beta pdf and then letting $\alpha \rightarrow \infty$. The first steps, which also include using Stirling's formula to replace two of the gamma function terms, are

$$\begin{aligned} f(x) &= \frac{\Gamma(\alpha + \tau) \gamma x^{\gamma\tau-1}}{\Gamma(\alpha) \Gamma(\tau) \theta^{\gamma\tau} (1 + x^\gamma \theta^{-\gamma})^{\alpha+\tau}} \\ &= \frac{e^{-\alpha-\tau} (\alpha + \tau)^{\alpha+\tau-1/2} (2\pi)^{1/2} \gamma x^{\gamma\tau-1}}{e^{-\alpha} \alpha^{\alpha-1/2} (2\pi)^{1/2} \Gamma(\tau) (\xi \alpha^{1/\gamma})^{\gamma\tau} (1 + x^\gamma \xi^{-\gamma} \alpha^{-1})^{\alpha+\tau}} \\ &= \frac{e^{-\tau} [(\alpha + \tau)/\alpha]^{\alpha+\tau-1/2} \gamma x^{\gamma\tau-1}}{\Gamma(\tau) \xi^{\gamma\tau} [1 + (x/\xi)^\gamma / \alpha]^{\alpha+\tau}}. \end{aligned}$$

The two limits

$$\lim_{\alpha \rightarrow \infty} \left(1 + \frac{\tau}{\alpha}\right)^{\alpha+\tau-1/2} = e^\tau, \quad \lim_{\alpha \rightarrow \infty} \left[1 + \frac{(x/\xi)^\gamma}{\alpha}\right]^{\alpha+\tau} = e^{(x/\xi)^\gamma}$$

can be substituted to yield

$$\lim_{\alpha \rightarrow \infty} f(x) = \frac{\gamma x^{\gamma\tau-1} e^{-(x/\xi)^\gamma}}{\Gamma(\tau) \xi^{\gamma\tau}}$$

which is the pdf of the transformed gamma distribution. \square

With a similar argument, the inverse transformed gamma distribution is obtained by letting τ go to infinity instead of α (see Exercise 4.39).

Because the Burr distribution is a transformed beta distribution with $\tau = 1$, its limiting case is the transformed gamma with $\tau = 1$ (using the parameterization in the previous example), which is the Weibull distribution. Similarly, the inverse Burr has the inverse Weibull as a limiting case. Finally, letting $\tau = \gamma = 1$ shows that the limiting case for the Pareto distribution is the exponential (and similarly for their inverse distributions).

As a final illustration of a limiting case, consider the transformed gamma distribution as parameterized above. Let $\gamma^{-1} \sqrt{\xi^\gamma} \rightarrow \sigma$ and $\gamma^{-1} (\xi^{\gamma\tau} - 1) \rightarrow \mu$. If this is done by letting $\tau \rightarrow \infty$ (so both γ and ξ must go to zero), the limiting distribution will be lognormal.

In Figure 4.7 some of the limiting and special case relationships are shown. Other interesting facts about the various distributions are also given.⁴

⁴Thanks to Dave Clark of American Re-Insurance Company for creating this picture.

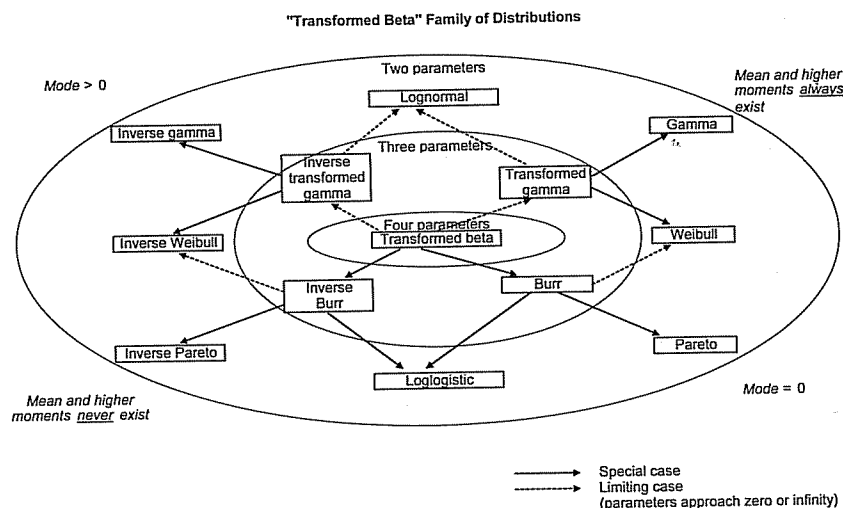


Fig. 4.7 Distributional relationships and characteristics.

4.5.4 Exercises

4.37 For a Pareto distribution, let both α and θ go to infinity with the ratio α/θ held constant. Show that the result is an exponential distribution.

4.38 Determine the limiting distribution of the generalized Pareto distribution as α and θ both go to infinity.

4.39 Show that as $\tau \rightarrow \infty$ in the transformed beta distribution the result is the inverse transformed gamma distribution.

4.6 DISCRETE DISTRIBUTIONS

4.6.1 Introduction

The purpose of this section is to introduce a large class of counting distributions. Counting distributions are discrete distributions with probabilities only on the nonnegative integers; that is, probabilities are defined only at the points $0, 1, 2, 3, 4, \dots$. In an insurance context, counting distributions describe the number of events such as losses to the insured or claims to the insurance company. With an understanding of both the number of claims and the size of claims, one can have a deeper understanding of a variety of issues surrounding insurance than if one has only information about total losses. The description of total losses in terms of numbers and amounts separately also allows one to address issues of modification of an insurance contract. Another reason for

separating numbers and amounts of claims is that models for the number of claims are fairly easy to obtain and experience has shown that the commonly used distributions really do model the propensity to generate losses.

We now formalize some of the notation that will be used for models for discrete phenomena. The **probability function** (pf) p_k denotes the probability that exactly k events (such as claims or losses) occur. Let N be a random variable representing the number of such events. Then

$$p_k = \Pr(N = k), \quad k = 0, 1, 2, \dots$$

As a reminder, the probability generating function (pgf) of a discrete random variable N with pf p_k is

$$P(z) = P_N(z) = E(z^N) = \sum_{k=0}^{\infty} p_k z^k. \quad (4.8)$$

As is true with the moment generating function, the pgf can be used to generate moments. In particular, $P'(1) = E(N)$ and $P''(1) = E[N(N-1)]$ (see Exercise 4.42). To see that the pgf really does generate probabilities,

$$\begin{aligned} P^{(m)}(z) &= E\left(\frac{d^m}{dz^m} z^N\right) = E[N(N-1)\cdots(N-m+1)z^{N-m}] \\ &= \sum_{k=m}^{\infty} k(k-1)\cdots(k-m+1)z^{k-m}p_k \\ P^{(m)}(0) &= m!p_m \text{ or } p_m = \frac{P^{(m)}(0)}{m!}. \end{aligned}$$

4.6.2 The Poisson distribution

The pf for the Poisson distribution is

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

The probability generating function from Example 3.16 is

$$P(z) = e^{\lambda(z-1)}, \quad \lambda > 0.$$

The mean and variance can be computed from the probability generating function as follows:

$$\begin{aligned} E(N) &= P'(1) = \lambda \\ E[N(N-1)] &= P''(1) = \lambda^2 \\ \text{Var}(N) &= E[N(N-1)] + E(N) - [E(N)]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

For the Poisson distribution the variance is equal to the mean. The Poisson distribution can arise from a Poisson process (to be discussed in Chapter 8). The Poisson distribution and Poisson processes are also discussed in many textbooks in statistics and actuarial science, including Panjer and Willmot [106] and Ross [116].

The Poisson distribution has at least two additional useful properties. The first is given in the following theorem.

Theorem 4.37 Let N_1, \dots, N_n be independent Poisson variables with parameters $\lambda_1, \dots, \lambda_n$. Then $N = N_1 + \dots + N_n$ has a Poisson distribution with parameter $\lambda_1 + \dots + \lambda_n$.

Proof: The pgf of the sum of independent random variables is the product of the individual pgfs. For the sum of Poisson random variables we have

$$\begin{aligned} P_N(z) &= \prod_{j=1}^n P_{N_j}(z) = \prod_{j=1}^n \exp[\lambda_j(z-1)] \\ &= \exp \left[\sum_{j=1}^n \lambda_j(z-1) \right] \\ &= e^{\lambda(z-1)}, \end{aligned}$$

where $\lambda = \lambda_1 + \dots + \lambda_n$. Just as is true with moment generating functions, the pgf is unique and therefore N must have a Poisson distribution with parameter λ . \square

The second property is particularly useful in modeling insurance risks. Suppose that the number of claims in a fixed time period, such as one year, follows a Poisson distribution. Further suppose that the claims can be classified into m distinct types. For example, claims could be classified by size, such as those below a fixed limit and those above the limit. It turns out that, if one is interested in studying the number of claims above the limit, that distribution is also Poisson but with a new Poisson parameter.

This is also useful when considering removing or adding a part of an insurance coverage. Suppose that the number of claims for a complicated medical benefit coverage follows a Poisson distribution. Consider the “types” of claims to be the different medical procedures or medical benefits under the plan. If one of the benefits is removed from the plan, again it turns out that the distribution of the number of claims under the revised plan will still have a Poisson distribution but with a new parameter.

In each of the cases mentioned in the previous paragraph, the number of claims of the different types will not only be Poisson distributed but also be independent of each other; that is, the distributions of the number of claims above the limit and the number below the limit will be independent. This

is a somewhat surprising result. For example, suppose we currently sell a policy with a deductible of 50 and experience has indicated that a Poisson distribution with a certain parameter is a valid model for the number of payments. Further suppose we are also comfortable with the assumption that the number of losses in a period also has the Poisson distribution but we do not know the parameter. Without additional information, it is impossible to infer the value of the Poisson parameter should the deductible be lowered or removed entirely. We now formalize these ideas in the following theorem.

Theorem 4.38 Suppose that the number of events N is a Poisson random variable with mean λ . Further suppose that each event can be classified into one of m types with probabilities p_1, \dots, p_m independent of all other events. Then the number of events N_1, \dots, N_m corresponding to event types 1, \dots , m respectively, are mutually independent Poisson random variables with means $\lambda p_1, \dots, \lambda p_m$, respectively.

Proof: For fixed $N = n$, the conditional joint distribution of (N_1, \dots, N_m) is multinomial with parameters (n, p_1, \dots, p_m) . Also, for fixed $N = n$, the conditional marginal distribution of N_j is binomial with parameters (n, p_j) .

The joint pf of (N_1, \dots, N_m) is given by

$$\begin{aligned} \Pr(N_1 = n_1, \dots, N_m = n_m) &= \Pr(N_1 = n_1, \dots, N_m = n_m | N = n) \\ &\quad \times \Pr(N = n) \\ &= \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} \dots p_m^{n_m} \frac{e^{-\lambda} \lambda^n}{n!} \\ &= \prod_{j=1}^m e^{-\lambda p_j} \frac{(\lambda p_j)^{n_j}}{n_j!}, \end{aligned}$$

where $n = n_1 + n_2 + \dots + n_m$. Similarly, the marginal pf of N_j is determined below.

$$\begin{aligned} \Pr(N_j = n_j) &= \sum_{n=n_j}^{\infty} \Pr(N_j = n_j | N = n) \Pr(N = n) \\ &= \sum_{n=n_j}^{\infty} \binom{n}{n_j} p_j^{n_j} (1-p_j)^{n-n_j} \frac{e^{-\lambda} \lambda^n}{n!} \\ &= e^{-\lambda} \frac{(\lambda p_j)^{n_j}}{n_j!} \sum_{n=n_j}^{\infty} \frac{[\lambda(1-p_j)]^{n-n_j}}{(n-n_j)!} \\ &= e^{-\lambda} \frac{(\lambda p_j)^{n_j}}{n_j!} e^{\lambda(1-p_j)} \\ &= e^{-\lambda p_j} \frac{(\lambda p_j)^{n_j}}{n_j!}. \end{aligned}$$

Hence the joint pf is the product of the marginal pfs, establishing mutual independence. \square

Example 4.39 In a study of medical insurance the expected number of claims per individual policy is 2.3 and the number of claims is Poisson distributed. You are considering removing one medical procedure from the coverage under this policy. Based on historical studies, this procedure accounts for approximately 10% of the claims. Determine the new frequency distribution.

From Theorem 4.38, we know that the distribution of the number of claims expected under the revised insurance policy after removing the procedure from coverage is Poisson with mean $0.9(2.3) = 2.07$. In carrying out studies of the distribution of total claims, and hence the appropriate premium under the new policy, one also needs to study the change in the amounts of losses, the severity distribution, because the distribution of amounts of losses for the procedure which was removed may be different from the distribution of amounts when all procedures are covered. \square

4.6.3 The negative binomial distribution

The negative binomial distribution has been used extensively as an alternative to the Poisson distribution. Like the Poisson distribution, it has positive probabilities on the nonnegative integers. Because it has two parameters, it has more flexibility in shape than the Poisson.

Definition 4.40 The probability function of the *negative binomial distribution* is given by

$$\Pr(N = k) = p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta} \right)^r \left(\frac{\beta}{1+\beta} \right)^k, \\ k = 0, 1, 2, \dots, \quad r > 0, \beta > 0. \quad (4.9)$$

The binomial coefficient is to be evaluated as

$$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}.$$

While k must be an integer, x may be any real number. When $x > k-1$, it can also be written as

$$\binom{x}{k} = \frac{\Gamma(x+1)}{\Gamma(k+1)\Gamma(x-k+1)},$$

which may be useful because $\ln \Gamma(x)$ is available in many spreadsheets, programming languages, and mathematics packages.

It is not difficult to show that the probability generating function for the negative binomial distribution is

$$P(z) = [1 - \beta(z-1)]^{-r}.$$

From this it follows that the mean and variance of the negative binomial distribution are

$$E(N) = r\beta \quad \text{and} \quad \text{Var}(N) = r\beta(1+\beta).$$

Because β is positive, the variance of the negative binomial distribution exceeds the mean. This is in contrast to the Poisson distribution for which the variance is equal to the mean. This suggests that for a particular set of data, if the observed variance is larger than the observed mean, the negative binomial might be a better candidate than the Poisson distribution as a model to be used.

The negative binomial distribution is a generalization of the Poisson in at least two different ways, namely as a mixed Poisson distribution with a gamma mixing distribution (demonstrated later in this subsection) and as a compound Poisson distribution with a logarithmic secondary distribution (see Section 4.6.7). Another view of the Poisson distribution is presented in Chapter 8. There, among other assumptions, the rate at which claims occur is assumed constant over time. If the rate is linearly increasing with regard to the number of past claims, then the number of claims in any period will have the negative binomial distribution. See *Insurance Risk Models* [106, Theorem 3.6.1] for this derivation of the negative binomial distribution.

The **geometric distribution** is the special case of the negative binomial distribution when $r = 1$. The geometric distribution is, in some senses, the discrete analogue of the continuous exponential distribution. Both the geometric and exponential distributions have an exponentially decaying probability function and hence the memoryless property. The memoryless property can be interpreted in various contexts as follows. If the exponential distribution is a distribution of lifetimes, then the expected future lifetime is constant for any age. If the exponential distribution describes the size of insurance claims, then the memoryless property can be interpreted as follows: *Given that a claim exceeds a certain level d , the expected amount of the claim in excess of d is constant and so does not depend on d .* That is, if a deductible of d is imposed, the expected payment per claim will be unchanged, but of course the expected number of payments will decrease. If the geometric distribution describes the number of claims, then the memoryless property can be interpreted as follows: *Given that there are at least m claims, the probability distribution of the number of claims in excess of m does not depend on m .* Among continuous distributions, the exponential distribution is used to distinguish between *subexponential* distributions with heavy (or fat) tails and distributions with light (or thin) tails. Similarly for frequency distributions, distributions that decay in the tail slower than the geometric distribution are often considered to have long tails, whereas distributions that decay more rapidly than the geometric have short tails. The negative binomial distribution has a long tail (decays more slowly than the geometric distribution) when $r < 1$ and a lighter tail than the geometric distribution when $r > 1$.

As noted earlier, one way to create the negative binomial distribution is as a mixture of Poissons. Suppose that we know that a risk has a Poisson number of claims distribution when the risk parameter λ is known. Now treat λ as being the outcome of a random variable Λ . We will denote the pf of Λ by $u(\lambda)$, where Λ may be continuous or discrete, and denote the cdf by $U(\lambda)$. The idea that λ is the outcome of a random variable can be justified in several ways. First, we can think of the population of risks as being heterogeneous with respect to the risk parameter Λ . In practice this makes sense. Consider a block of insurance policies with the same premium, such as a group of automobile drivers in the same rating category. Such categories are usually broad ranges such as 0–7,500 miles per year, garaged in a rural area, commuting less than 50 miles per week, and so on. We know that not all drivers in the same rating category are the same even though they may “appear” to be the same from the point of view of the insurer and are charged the same premium. The parameter λ measures the expected number of accidents. If λ varies across the population of drivers, then we can think of the insured individual as a sample value drawn from the population of possible drivers. This means implicitly that λ is unknown to the insurer but follows some distribution, in this case $u(\lambda)$, over the population of drivers. The true value of λ is unobservable. All we observe are the number of accidents coming from the driver. There is now an additional degree of uncertainty, that is, uncertainty about the parameter.

This is the same mixing process that was discussed with regard to continuous distributions in Section 4.4.5. In some contexts this is referred to as *parameter uncertainty*. In the Bayesian context, the distribution of Λ is called a *prior distribution* and the parameters of its distribution are sometimes called *hyperparameters*. The role of the distribution $u(\cdot)$ is very important in credibility theory, the subject of Chapter 16. When the parameter λ is unknown, the probability that exactly k claims will arise can be written as the expected value of the same probability but conditional on $\Lambda = \lambda$ where the expectation is taken with respect to the distribution of Λ . From the law of total probability, we can write

$$\begin{aligned} p_k &= \Pr(N = k) \\ &= E[\Pr(N = k | \Lambda)] \\ &= \int_0^\infty \Pr(N = k | \Lambda = \lambda) u(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} u(\lambda) d\lambda. \end{aligned}$$

Now suppose Λ has a gamma distribution. Then

$$p_k = \int_0^\infty \frac{e^{-\lambda} \lambda^k}{k!} \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\theta}}}{\theta^\alpha \Gamma(\alpha)} d\lambda = \frac{1}{\Gamma(\alpha)} \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-\lambda(1+\frac{1}{\theta})} \lambda^{k+\alpha-1} d\lambda.$$

From the definition of the gamma distribution in Appendix A, this expression can be evaluated as

$$\begin{aligned} p_k &= \frac{\Gamma(k+\alpha)}{k! \Gamma(\alpha)} \frac{\theta^k}{(1+\theta)^{k+\alpha}} \\ &= \binom{k+\alpha-1}{k} \left(\frac{\theta}{1+\theta} \right)^k \left(\frac{1}{1+\theta} \right)^\alpha. \end{aligned}$$

This formula is of the same form as (4.9), demonstrating that the mixed Poisson, with a gamma mixing distribution, is the same as a negative binomial distribution.

It is worth noting that the Poisson distribution is a limiting case of the negative binomial distribution. To see this, let r go to infinity and β go to zero while keeping their product constant. Let $\lambda = r\beta$ be that constant. Substituting $\beta = \lambda/r$ in the pgf leads to (using L'Hôpital's rule in lines 3 and 5)

$$\begin{aligned} \lim_{r \rightarrow \infty} \left[1 - \frac{\lambda(z-1)}{r} \right]^{-r} &= \exp \left\{ \lim_{r \rightarrow \infty} -r \ln \left[1 - \frac{\lambda(z-1)}{r} \right] \right\} \\ &= \exp \left\{ - \lim_{r \rightarrow \infty} \frac{\ln[1 - \lambda(z-1)/r]}{r^{-1}} \right\} \\ &= \exp \left\{ \lim_{r \rightarrow \infty} \frac{[1 - \lambda(z-1)/r]^{-1} \lambda(z-1)/r^2}{r^{-2}} \right\} \\ &= \exp \left[\lim_{r \rightarrow \infty} \frac{r \lambda(z-1)}{r - \lambda(z-1)} \right] \\ &= \exp \left\{ \lim_{r \rightarrow \infty} [\lambda(z-1)] \right\} \\ &= \exp[\lambda(z-1)] \end{aligned}$$

which is the pgf of the Poisson distribution.

4.6.4 The binomial distribution

The binomial distribution is another counting distribution that arises naturally in claim number modeling. It possesses some properties different from the Poisson and the negative binomial that make it particularly useful. First, its variance is smaller than its mean. This makes it useful for data sets in which the observed sample variance is less than the sample mean. This contrasts with the negative binomial, where the variance exceeds the mean, and it also contrasts with the Poisson distribution, where the variance is equal to the mean.

Second, it describes a physical situation in which m risks are each subject to claim or loss. We can formalize this as follows. Consider m independent and identical risks each with probability q of making a claim. This might apply to

a life insurance situation in which all the individuals under consideration are in the same mortality class; that is, they may all be male smokers at age 35 and duration 5 of an insurance policy. In that case, q is the probability that a person with those attributes will die in the next year. Then the number of claims for a single person follows a Bernoulli distribution, a distribution with probability $1 - q$ at 0 and probability q at 1. The probability generating function of the number of claims per individual is then given by

$$P(z) = (1 - q)z^0 + qz^1 = 1 + q(z - 1).$$

Now, if there are m such independent individuals, then the probability generating functions can be multiplied together to give the probability generating function of the total number of claims arising from the group of m individuals. That probability generating function is

$$P(z) = [1 + q(z - 1)]^m, \quad 0 < q < 1.$$

Then from this it is easy to show that the probability of exactly k claims from the group is

$$p_k = \Pr(N = k) = \binom{m}{k} q^k (1 - q)^{m-k}, \quad k = 0, 1, \dots, m,$$

the pf for a binomial distribution with parameters m and q . From this Bernoulli trial framework, it is clear that at most m events (claims) can occur. Hence, the distribution only has positive probabilities on the nonnegative integers up to and including m .

Consequently, an additional attribute of the binomial distribution that is sometimes useful is that it has finite support; that is, the range of values for which there exist positive probabilities has finite length. This may be useful, for instance, in modeling the number of individuals injured in an automobile accident or the number of family members covered under a health insurance policy. In each case it is reasonable to have an upper limit on the range of possible values. It is useful also in connection with situations where it is believed that it is unreasonable to assign positive probabilities beyond some point. For example, if one is modeling the number of accidents per automobile during a one-year period, it is probably physically impossible for there to be more than some number, say 12, of claims during the year given the time it would take to repair the automobile between accidents. If a model with probabilities that extend beyond 12 were used, those probabilities should be very small so that they have little impact on any decisions that are made. The mean and variance of the binomial distribution are given by

$$E(N) = mq, \quad \text{Var}(N) = mq(1 - q).$$

Table 4.1 Members of the $(a, b, 0)$ class

Distribution	a	b	p_0
Poisson	0	λ	$e^{-\lambda}$
Binomial	$-\frac{q}{1 - q}$	$(m + 1)\frac{q}{1 - q}$	$(1 - q)^m$
Negative binomial	$\frac{\beta}{1 + \beta}$	$(r - 1)\frac{\beta}{1 + \beta}$	$(1 + \beta)^{-r}$
Geometric	$\frac{\beta}{1 + \beta}$	0	$(1 + \beta)^{-1}$

4.6.5 The $(a, b, 0)$ class

The following definition characterizes the members of this class of distributions.

Definition 4.41 Let p_k be the pf of a discrete random variable. It is a member of the $(a, b, 0)$ class of distributions, provided that there exists constants a and b such that

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 1, 2, 3, \dots$$

This recursion describes the relative size of successive probabilities in the counting distribution. The probability at zero, p_0 , can be obtained from the recursive formula because the probabilities must add up to 1. This provides a boundary condition. The $(a, b, 0)$ class of distributions is a two-parameter class, the two parameters being a and b . By substituting in the probability function for each of the Poisson, binomial, and negative binomial distributions on the left-hand side of the recursion, it can be seen that each of these three distributions satisfies the recursion and that the values of a and b are as given in Table 4.1. In addition the table gives the value of p_0 , the starting value for the recursion. Also in the table is the geometric distribution, the one-parameter special case ($r = 1$) of the negative binomial distribution.

It can be shown (see Panjer and Willmot [106, Chapter 6]) that these are the only possible distributions satisfying this recursive formula.

The recursive formula can be rewritten as

$$k \frac{p_k}{p_{k-1}} = ak + b, \quad k = 1, 2, 3, \dots$$

The expression on the left-hand side is a linear function in k . Note from Table 4.1 that the slope a of the straight line is 0 for the Poisson distribution, is negative for the binomial distribution, and is positive for the negative binomial distribution, including the geometric. This suggests a graphical way of indicating which of the three distributions might be selected for fitting. First,

Table 4.2 Accident profile

Number of accidents, k	Number of policies, n_k	$k \frac{n_k}{n_{k-1}}$
0	7,840	
1	1,317	0.17
2	239	0.36
3	42	0.53
4	14	1.33
5	4	1.43
6	4	6.00
7	1	1.75
8+	0	
Total	9,461	

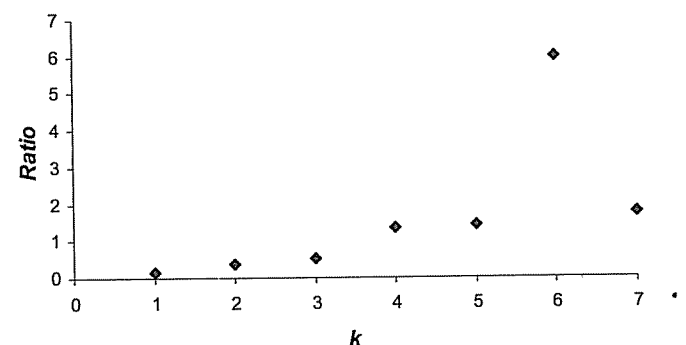
one can plot

$$k \frac{\hat{p}_k}{\hat{p}_{k-1}} = k \frac{n_k}{n_{k-1}}$$

against k . The observed values should form approximately a straight line if one of these models is to be selected, and the value of the slope should be an indication of which of the models should be selected. Note that this cannot be done if any of the n_k are 0. Hence this procedure is less useful for a small number of observations.

Example 4.42 Consider the accident data in Table 4.2, which is taken from Thyron [128]. For the 9,461 automobile insurance policies studied, the number of accidents under the policy is recorded in the table. Also recorded in the table is the observed value of the quantity that should be linear.

Figure 4.8 plots the value of the quantity of interest against k , the number of accidents. It can be seen from the graph that the quantity of interest looks approximately linear except for the point at $k = 6$. The reliability of the quantities as k increases diminishes because the number of observations becomes small and the variability of the results grows. This illustrates the weakness of this ad hoc procedure. Visually, all the points appear to have equal value. However, the points on the left are more reliable than the points on the right. From the graph, it can be seen that the slope is positive and the data appear approximately linear. This suggests the negative binomial distribution is an appropriate model. Whether or not the slope is significantly different from 0 is also not easily judged from the graph. By rescaling the vertical axis of the graph, the slope can be made to look steeper and hence the slope could be made to appear to be significantly different from 0. Graphically, it is difficult to distinguish between the Poisson and the negative binomial

Fig. 4.8 Plot of the ratio kn_k/n_{k-1} against k .

distribution because the Poisson requires a slope of 0. However, we can say that the binomial distribution is probably not a good choice since there is no evidence of a negative slope. In this case it is advisable to fit both the Poisson and negative binomial distributions and conduct a more formal test to choose between them.

It is also possible to compare the appropriateness of the distributions by looking at the relationship of the variance to the mean. For this data set, the mean number of claims per policy is 0.2144. The variance is 0.2889. Because the variance exceeds the mean, the negative binomial should be considered as an alternative to the Poisson. Again this is a qualitative comment because we have, at this point, no formal way of determining whether the variance is sufficiently larger than the mean to warrant use of the negative binomial. In order to do some formal analysis, Table 4.3 gives the results of maximum likelihood estimation (to be discussed in Chapter 12) of the parameters of the Poisson and negative binomial distributions and the negative loglikelihood in each case. In Chapter 13 formal selection methods are presented. They would indicate that the negative binomial is superior to the Poisson as a model for this data set. However, those methods also indicate that the negative binomial is not a particularly good model, and thus some of the distributions yet to be introduced should be considered.

In subsequent subsections we will expand the class of the distributions beyond the three discussed in this section by constructing more general models related to the Poisson, binomial, and negative binomial distributions.

4.6.6 Truncation and modification at zero

At times, the distributions discussed previously do not adequately describe the characteristics of some data sets encountered in practice. This may be

Table 4.3 Poisson-negative binomial comparison

Distribution	Parameter estimates	-Loglikelihood
Poisson	$\hat{\lambda} = 0.2143537$	5,490.78
Negative binomial	$\hat{\beta} = 0.3055594$ $\hat{r} = 0.7015122$	5,348.04

because the tail of the negative binomial is not heavy enough or because the distributions in the $(a, b, 0)$ class cannot capture the shape of the data set in some other part of the distribution.

In this section, we address the problem of a poor fit at the left-hand end of the distribution, in particular the probability at zero.

For insurance count data, the probability at zero is the probability that no claims occur during the period under study. For applications in insurance where the probability of occurrence of a loss is low, the probability at zero has the largest value. Thus, it is important to pay special attention to the fit at this point.

There are also situations that naturally occur which generate unusually large probabilities at zero. Consider the case of group dental insurance. If, in a family, both husband and wife have coverage with their employer-sponsored plans and both group insurance contracts provide coverage for all family members, the claims will be made to the insurer of the plan which provides the better benefits, and no claims may be made under the other contract. Then, in conducting studies for a specific insurer, one may find a higher than expected number of individuals who made no claim.

Similarly, it is possible to have situations in which there is less than the expected number, or even zero, occurrences at zero. For example, if one is counting the number of claims from accidents resulting in a claim, the minimum observed value is 1.

An adjustment of the probability at zero is easily handled for the Poisson, binomial, and negative binomial distributions.

Definition 4.43 Let p_k be the pf of a discrete random variable. It is a member of the $(a, b, 1)$ class of distributions provided that there exists constants a and b such that

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 2, 3, 4, \dots$$

Note that the only difference from the $(a, b, 0)$ class is that the recursion begins at p_1 rather than p_0 . This forces the distribution from $k = 1$ to $k = \infty$ to have the same shape as the $(a, b, 0)$ class in the sense that the probabilities

are the same up to a constant of proportionality because $\sum_{k=1}^{\infty} p_k$ can be set to any number in the interval $(0, 1]$. The remaining probability is at $k = 0$.

We will distinguish between the situations in which $p_0 = 0$ and those where $p_0 > 0$. The first subclass is called the **truncated** (more specifically, **zero-truncated**) distributions. The members are the zero-truncated Poisson, zero-truncated binomial, and zero-truncated negative binomial (and its special case, the zero-truncated geometric) distributions.

The second subclass will be referred to as the **zero-modified** distributions because the probability is modified from that for the $(a, b, 0)$ class. These distributions can be viewed as a mixture of an $(a, b, 0)$ distribution and a degenerate distribution with all the probability at zero. Alternatively, they can be called **truncated with zeros** distributions since the distribution can be viewed as a mixture of a truncated distribution and a degenerate distribution with all the probability at zero. We now show this more formally. Note that all zero-truncated distributions can be considered as zero-modified distributions, with the particular modification being to set $p_0 = 0$.

With three types of distributions, notation can become confusing. When writing about discrete distributions in general, we will continue to let $p_k = \Pr(N = k)$. When referring to a zero-truncated distribution, we will use p_k^T , and when referring to a zero-modified distribution, we will use p_k^M . Once again, it is possible for a zero-modified distribution to be a zero-truncated distribution.

Let $P(z) = \sum_{k=0}^{\infty} p_k z^k$ denote the pgf of a member of the $(a, b, 0)$ class. Let $P^M(z) = \sum_{k=0}^{\infty} p_k^M z^k$ denote the pgf of the corresponding member of the $(a, b, 1)$ class; that is,

$$p_k^M = c p_k, \quad k = 1, 2, 3, \dots,$$

and p_0^M is an arbitrary number. Then

$$\begin{aligned} P^M(z) &= p_0^M + \sum_{k=1}^{\infty} p_k^M z^k \\ &= p_0^M + c \sum_{k=1}^{\infty} p_k z^k \\ &= p_0^M + c[P(z) - p_0]. \end{aligned}$$

Because $P^M(1) = P(1) = 1$,

$$1 = p_0^M + c(1 - p_0),$$

resulting in

$$c = \frac{1 - p_0^M}{1 - p_0} \text{ or } p_0^M = 1 - c(1 - p_0).$$

This relationship is necessary to ensure that the p_k^M sum to 1. We then have

$$\begin{aligned} P^M(z) &= p_0^M + \frac{1-p_0^M}{1-p_0} [P(z) - p_0] \\ &= \left(1 - \frac{1-p_0^M}{1-p_0}\right) 1 + \frac{1-p_0^M}{1-p_0} P(z). \end{aligned} \quad (4.10)$$

This is a weighted average of the pgfs of the degenerate distribution and the corresponding $(a, b, 0)$ member. Furthermore,

$$p_k^M = \frac{1-p_0^M}{1-p_0} p_k, \quad k = 1, 2, \dots \quad (4.11)$$

Let $P^T(z)$ denote the pgf of the zero-truncated distribution corresponding to an $(a, b, 0)$ pgf $P(z)$. Then, by setting $p_0^M = 0$ in (4.10) and (4.11),

$$P^T(z) = \frac{P(z) - p_0}{1 - p_0}$$

and

$$p_k^T = \frac{p_k}{1 - p_0}, \quad k = 1, 2, \dots \quad (4.12)$$

Then from (4.11)

$$p_k^M = (1 - p_0^M) p_k^T, \quad k = 1, 2, \dots, \quad (4.13)$$

and

$$P^M(z) = p_0^M (1) + (1 - p_0^M) P^T(z). \quad (4.14)$$

Then the zero-modified distribution is also the weighted average of a degenerate distribution and the zero-truncated member of the $(a, b, 0)$ class. The following example illustrates these relationships.

Example 4.44 Consider a negative binomial random variable with parameters $\beta = 0.5$ and $r = 2.5$. Determine the first four probabilities for this random variable. Then determine the corresponding probabilities for the zero-truncated and zero-modified (with $p_0^M = 0.6$) versions.

From Table 4.4 on Page 89 we have, for the negative binomial distribution,

$$\begin{aligned} p_0 &= (1 + 0.5)^{-2.5} = 0.362887, \\ a &= \frac{0.5}{1.5} = \frac{1}{3}, \\ b &= \frac{(2.5 - 1)(0.5)}{1.5} = \frac{1}{2}. \end{aligned}$$

The first three recursions are

$$\begin{aligned} p_1 &= 0.362887 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{1}\right) = 0.302406, \\ p_2 &= 0.302406 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{2}\right) = 0.176404, \\ p_3 &= 0.176404 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{3}\right) = 0.088202. \end{aligned}$$

For the zero-truncated random variable, $p_0^T = 0$ by definition. The recursions start with [from (4.12)] $p_1^T = 0.302406/(1 - 0.362887) = 0.474651$. Then

$$\begin{aligned} p_2^T &= 0.474651 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{2}\right) = 0.276880, \\ p_3^T &= 0.276880 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{3}\right) = 0.138440. \end{aligned}$$

If the original values were all available, then the zero-truncated probabilities could have all been obtained by multiplying the original values by $1/(1 - 0.362887) = 1.569580$.

For the zero-modified random variable, $p_0^M = 0.6$ arbitrarily. From (4.11), $p_1^M = (1 - 0.6)(0.302406)/(1 - 0.362887) = 0.189860$. Then

$$\begin{aligned} p_2^M &= 0.189860 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{2}\right) = 0.110752, \\ p_3^M &= 0.110752 \left(\frac{1}{3} + \frac{1}{2} \frac{1}{3}\right) = 0.055376. \end{aligned}$$

In this case, each original negative binomial probability has been multiplied by $(1 - 0.6)/(1 - 0.362887) = 0.627832$. Also note that, for $j \geq 1$, $p_j^M = 0.4p_j^T$. \square

Although we have only discussed the zero-modified distributions of the $(a, b, 0)$ class, the $(a, b, 1)$ class admits additional distributions. The (a, b) parameter space can be expanded to admit an extension of the negative binomial distribution to include cases where $-1 < r < 0$. For the $(a, b, 0)$ class, $r > 0$ is required. By adding the additional region to the sample space, the "extended" truncated negative binomial (ETNB) distribution has parameter restrictions $\beta > 0$, $r > -1$, $r \neq 0$.

To show that the recursive equation

$$p_k = p_{k-1} \left(a + \frac{b}{k}\right), \quad k = 2, 3, \dots, \quad (4.15)$$

with $p_0 = 0$ defines a proper distribution, it is sufficient to show that for any value of p_1 , the successive values of p_k obtained recursively are each positive and that $\sum_{k=1}^{\infty} p_k < \infty$. For the ETNB, this must be done for the parameter space

$$\begin{aligned} a &= \frac{\beta}{1 + \beta}, \quad \beta > 0, \\ b &= (r - 1) \frac{\beta}{1 + \beta}, \quad r > -1, r \neq 0 \end{aligned}$$

(see Exercise 4.44).

When $r \rightarrow 0$, the limiting case of the ETNB is the logarithmic distribution with

$$p_k^T = \frac{[\beta/(1 + \beta)]^k}{k \ln(1 + \beta)}, \quad k = 1, 2, 3, \dots \quad (4.16)$$

(see Exercise 4.45). The pgf of the logarithmic distribution is

$$P^T(z) = 1 - \frac{\ln[1 - \beta(z-1)]}{\ln(1+\beta)} \quad (4.17)$$

(see Exercise 4.46). The zero-modified logarithmic distribution is created by assigning an arbitrary probability at zero and reducing the remaining probabilities.

It is also interesting that the special extreme case with $-1 < r < 0$ and $\beta \rightarrow \infty$ is a proper distribution, and is sometimes called the Sibuya distribution. It has pgf $P(z) = 1 - (1-z)^{-r}$ and no moments exist (see Exercise 4.47). Distributions with no moments are not particularly interesting for modeling claim numbers (unless the right tail is subsequently modified) because then an infinite number of claims are expected. This might be difficult to price!

Example 4.45 Determine the probabilities for an ETNB distribution with $r = -0.5$ and $\beta = 1$. Do this both for the truncated version and for the modified version with $p_0^M = 0.6$ set arbitrarily.

We have $a = 1/(1+1) = 0.5$ and $b = (-0.5-1)(1)/(1+1) = -0.75$. From Appendix B we also have $p_1^T = -0.5(1)/[(1+1)^{0.5} - (1+1)] = 0.853553$. Subsequent values are

$$p_2^T = \left(0.5 - \frac{0.75}{2}\right)(0.853553) = 0.106694,$$

$$p_3^T = \left(0.5 - \frac{0.75}{3}\right)(0.106694) = 0.026674.$$

For the modified probabilities, the truncated probabilities need to be multiplied by 0.4 to produce $p_1^M = 0.341421$, $p_2^M = 0.042678$, and $p_3^M = 0.010670$.

A reasonable question is to ask if there is a "natural" member of the ETNB distribution, that is, one for which the recursion would begin with p_1 rather than p_2 . For that to be the case, the natural value of p_0 would have to satisfy $p_1 = (0.5 - 0.75/1)p_0 = -0.25p_0$. This would force one of the two probabilities to be negative and so there is no acceptable solution. It is easy to show that this occurs for any $r < 0$. \square

There are no other members of the $(a, b, 1)$ class beyond the ones discussed above. A summary is given in Table 4.4.

4.6.7 Compound frequency models

A larger class of distributions can be created by the processes of compounding any two discrete distributions. The term *compounding* reflects the idea that the pgf of the new distribution $P(z)$ is written as

$$P(z) = P_N[P_M(z)], \quad (4.18)$$

Table 4.4 Members of the $(a, b, 1)$ class

Distribution ^a	p_0	a	b	Parameter space
Poisson	$e^{-\lambda}$	0	λ	$\lambda > 0$
ZT Poisson	0	0	λ	$\lambda > 0$
ZM Poisson	Arbitrary	0	λ	$\lambda > 0$
Binomial	$(1-q)^m$	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$0 < q < 1$
ZT binomial	0	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$0 < q < 1$
ZM binomial	Arbitrary	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$0 < q < 1$
Negative binomial	$(1+\beta)^{-r}$	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$r > 0, \beta > 0$
ETNB	0	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$r > -1,^b \beta > 0$
ZM ETNB	Arbitrary	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$r > -1,^b \beta > 0$
Geometric	$(1+\beta)^{-1}$	$\frac{\beta}{1+\beta}$	0	$\beta > 0$
ZT geometric	0	$\frac{\beta}{1+\beta}$	0	$\beta > 0$
ZM geometric	Arbitrary	$\frac{\beta}{1+\beta}$	0	$\beta > 0$
Logarithmic	0	$\frac{\beta}{1+\beta}$	$-\frac{\beta}{1+\beta}$	$\beta > 0$
ZM logarithmic	Arbitrary	$\frac{\beta}{1+\beta}$	$-\frac{\beta}{1+\beta}$	$\beta > 0$

^aZT = zero truncated, ZM = zero modified.

^bExcluding $r = 0$, which is the logarithmic distribution.

where $P_N(z)$ and $P_M(z)$ are called the *primary* and *secondary* distributions, respectively.

The compound distributions arise naturally as follows. Let N be a counting random variable with pgf $P_N(z)$. Let M_1, M_2, \dots be identically and independently distributed random variables with pgf $P_M(z)$. Assuming that the M_j s do not depend on N , the pgf of the random sum $S = M_1 + M_2 + \dots + M_N$ (where $N = 0$ implies that $S = 0$) is $P_S(z) = P_N[P_M(z)]$. This is shown as

$$P_S(z) = \sum_{k=0}^{\infty} \Pr(S=k)z^k = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \Pr(S=k|N=n) \Pr(N=n)z^k$$

$$\begin{aligned}
&= \sum_{n=0}^{\infty} \Pr(N=n) \sum_{k=0}^{\infty} \Pr(M_1 + \cdots + M_n = k | N=n) z^k \\
&= \sum_{n=0}^{\infty} \Pr(N=n) [P_M(z)]^n \\
&= P_N[P_M(z)].
\end{aligned}$$

In insurance contexts, this distribution can arise naturally. If N represents the number of accidents arising in a portfolio of risks and $\{M_k; k=1, 2, \dots, N\}$ represents the number of claims (injuries, number of cars, etc.) from the accidents, then S represents the total number of claims from the portfolio. This kind of interpretation is not necessary to justify the use of a compound distribution. If a compound distribution fits data well, that may be enough justification itself. Also, there are other motivations for these distributions, as presented in Section 4.6.12.

✕ **Example 4.46** *Demonstrate that any zero-modified distribution is a compound distribution.*

Consider a primary Bernoulli distribution. It has pgf $P_N(z) = 1 - q + qz$. Then consider an arbitrary secondary distribution with pgf $P_M(z)$. Then, from (4.18) we obtain

$$P_S(z) = P_N[P_M(z)] = 1 - q + qP_M(z).$$

From (4.10) this is the pgf of a ZM distribution with

$$q = \frac{1 - p_0^M}{1 - p_0}.$$

That is, the ZM distribution has assigned arbitrary probability p_0^M at zero, while p_0 is the probability assigned at zero by the secondary distribution. \square

Example 4.47 *Consider the case where both M and N have the Poisson distribution. Determine the pgf of this distribution.*

This distribution is called the Poisson-Poisson or Neyman Type A distribution. Let $P_N(z) = e^{\lambda_1(z-1)}$ and $P_M(z) = e^{\lambda_2(z-1)}$. Then

$$P_S(z) = e^{\lambda_1[e^{\lambda_2(z-1)} - 1]}.$$

When λ_2 is a lot larger than λ_1 —for example, $\lambda_1 = 0.1$ and $\lambda_2 = 10$ —the resulting distribution will have two local modes. \square

The probability of exactly k claims can be written as

$$\begin{aligned}
\Pr(S=k) &= \sum_{n=0}^{\infty} \Pr(S=k | N=n) \Pr(N=n) \\
&= \sum_{n=0}^{\infty} \Pr(M_1 + \cdots + M_N = k | N=n) \Pr(N=n) \\
&= \sum_{n=0}^{\infty} \Pr(M_1 + \cdots + M_n = k) \Pr(N=n). \quad (4.19)
\end{aligned}$$

Letting $g_n = \Pr(S=n)$, $p_n = \Pr(N=n)$, and $f_n = \Pr(M=n)$, this is rewritten as

$$g_k = \sum_{n=0}^{\infty} p_n f_k^{*n} \quad (4.20)$$

where f_k^{*n} , $k=0, 1, \dots$, is the “ n -fold convolution” of the function f_k , $k=0, 1, \dots$, that is, the probability that the sum of n random variables which are each independent and identically distributed (i.i.d.) with probability function f_k will take on value k .

When $P_N(z)$ is chosen to be a member of the $(a, b, 0)$ class,

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k=1, 2, \dots, \quad (4.21)$$

and a simple recursive formula can be used. This formula avoids the use of convolutions and thus reduces the computations considerably.

Theorem 4.48 *If the primary distribution is a member of the $(a, b, 0)$ class, the recursive formula is*

$$g_k = \frac{1}{1 - af_0} \sum_{j=1}^k \left(a + \frac{bj}{k}\right) f_j g_{k-j}, \quad k=1, 2, 3, \dots \quad (4.22)$$

Proof: From (4.21),

$$np_n = a(n-1)p_{n-1} + (a+b)p_{n-1}.$$

Multiplying each side by $[P_M(z)]^{n-1} P'_M(z)$ and summing over n yields

$$\begin{aligned}
\sum_{n=1}^{\infty} np_n [P_M(z)]^{n-1} P'_M(z) &= a \sum_{n=1}^{\infty} (n-1) p_{n-1} [P_M(z)]^{n-1} P'_M(z) \\
&\quad + (a+b) \sum_{n=1}^{\infty} p_{n-1} [P_M(z)]^{n-1} P'_M(z).
\end{aligned}$$

Because $P_S(z) = \sum_{n=0}^{\infty} p_n [P_M(z)]^n$, the previous equation is

$$P'_S(z) = a \sum_{n=0}^{\infty} n p_n [P_M(z)]^n P'_M(z) + (a+b) \sum_{n=0}^{\infty} p_n [P_M(z)]^n P'_M(z).$$

Therefore

$$P'_S(z) = a P'_S(z) P_M(z) + (a+b) P_S(z) P'_M(z).$$

Each side can be expanded in powers of z . The coefficients of z^{k-1} in such an expansion must be the same on both sides of the equation. Hence, for $k = 1, 2, \dots$ we have

$$\begin{aligned} k g_k &= a \sum_{j=0}^k (k-j) f_j g_{k-j} + (a+b) \sum_{j=0}^k j f_j g_{k-j} \\ &= a k f_0 g_k + a \sum_{j=1}^k (k-j) f_j g_{k-j} + (a+b) \sum_{j=1}^k j f_j g_{k-j} \\ &= a k f_0 g_k + a k \sum_{j=1}^k f_j g_{k-j} + b \sum_{j=1}^k j f_j g_{k-j}. \end{aligned}$$

Therefore,

$$g_k = a f_0 g_k + \sum_{j=1}^k \left(a + \frac{b j}{k} \right) f_j g_{k-j}.$$

Rearrangement yields (4.22). \square

In order to use (4.22), the starting value g_0 is required and is given in Theorem 4.51 below. If the primary distribution is a member of the $(a, b, 1)$ class, the proof must be modified to reflect the fact that the recursion for the primary distribution begins at $k = 2$. The result is the following.

Theorem 4.49 *If the primary distribution is a member of the $(a, b, 1)$ class, the recursive formula is*

$$g_k = \frac{[p_1 - (a+b)p_0]f_k + \sum_{j=1}^k (a + b j/k) f_j g_{k-j}}{1 - a f_0}, \quad k = 1, 2, 3, \dots \quad (4.23)$$

Proof: It is similar to the proof of Theorem 4.48 and is left to the reader. \square

Example 4.50 *Develop the recursive formula for the case where the primary distribution is Poisson.*

In this case $a = 0$ and $b = \lambda$, yielding the recursive form

$$g_k = \frac{\lambda}{k} \sum_{j=1}^k j f_j g_{k-j}.$$

The starting value is, from (4.18),

$$\begin{aligned} g_0 &= \Pr(S = 0) = P(0) \\ &= P_N[P_M(0)] = P_N(f_0) \\ &= e^{-\lambda(1-f_0)}. \end{aligned} \quad (4.24)$$

Distributions of this type are called compound Poisson distributions. When the secondary distribution is specified, the compound distribution is called Poisson-X, where X is the name of the secondary distribution. \square

The method used to obtain g_0 applies to any compound distribution.

Theorem 4.51 *For any compound distribution, $g_0 = P_N(f_0)$, where $P_N(z)$ is the pgf of the primary distribution and f_0 is the probability that the secondary distribution takes on the value zero.*

Proof: See the second line of (4.24). \square

$\chi \checkmark$ Note that the secondary distribution is not required to be in any special form. However, to keep the number of distributions manageable, secondary distributions will be selected from the $(a, b, 0)$ or the $(a, b, 1)$ class.

Example 4.52 *Calculate the probabilities for the Poisson-ETNB distribution where $\lambda = 3$ for the Poisson distribution and the ETNB distribution has $r = -0.5$ and $\beta = 1$.*

From Example 4.45 the secondary probabilities are $f_0 = 0$, $f_1 = 0.853553$, $f_2 = 0.106694$, and $f_3 = 0.026674$. From (4.24), $g_0 = \exp[-3(1-0)] = 0.049787$. For the Poisson primary distribution, $a = 0$ and $b = 3$. The recursive formula in (4.22) becomes

$$g_k = \frac{\sum_{j=1}^k (3j/k) f_j g_{k-j}}{1 - 0(0)} = \sum_{j=1}^k \frac{3j}{k} f_j g_{k-j}.$$

Then,

$$\begin{aligned} g_1 &= \frac{3(1)}{1} 0.853553(0.049787) = 0.127488, \\ g_2 &= \frac{3(1)}{2} 0.853553(0.127488) + \frac{3(2)}{2} 0.106694(0.049787) = 0.179163, \\ g_3 &= \frac{3(1)}{3} 0.853553(0.179163) + \frac{3(2)}{3} 0.106694(0.127488) \\ &\quad + \frac{3(3)}{3} 0.026674(0.049787) = 0.184114. \end{aligned} \quad \square$$

Example 4.53 Demonstrate that the Poisson-logarithmic distribution is a negative binomial distribution.

The negative binomial distribution has pgf

$$P(z) = [1 - \beta(z-1)]^{-r}.$$

Suppose $P_N(z)$ is Poisson(λ) and $P_M(z)$ is logarithmic(β); then

$$\begin{aligned} P_N[P_M(z)] &= \exp\{\lambda[P_M(z) - 1]\} \\ &= \exp\left\{\lambda\left[1 - \frac{\ln[1 - \beta(z-1)]}{\ln(1+\beta)} - 1\right]\right\} \\ &= \exp\left\{\frac{-\lambda}{\ln(1+\beta)} \ln[1 - \beta(z-1)]\right\} \\ &= [1 - \beta(z-1)]^{-\lambda/\ln(1+\beta)} \\ &= [1 - \beta(z-1)]^{-r}, \end{aligned}$$

where $r = \lambda/\ln(1+\beta)$. This shows that the negative binomial distribution can be written as a compound Poisson distribution with a logarithmic secondary distribution. \square

The above example shows that the "Poisson-logarithmic" distribution does not create a new distribution beyond the $(a, b, 0)$ and $(a, b, 1)$ classes. As a result, this combination of distributions is not useful to us. Another combination which does not create a new distribution beyond the $(a, b, 1)$ class is the compound geometric distribution where both the primary and secondary distributions are geometric. The resulting distribution is a zero-modified geometric distribution, as shown in Exercise 4.51. The following theorem shows that certain other combinations are also of no use in expanding the class of distributions through compounding. Suppose $P_S(z) = P_N[P_M(z); \theta]$ as before. Now, $P_M(z)$ can always be written as

$$P_M(z) = f_0 + (1 - f_0)P_M^*(z) \quad (4.25)$$

where $P_M^*(z)$ is the pgf of the conditional distribution over the positive range (in other words, the zero-truncated version).

Theorem 4.54 Suppose the pgf $P_N(z; \theta)$ satisfies

$$P_N(z; \theta) = B[\theta(z-1)]$$

for some parameter θ and some function $B(z)$ which is independent of θ . That is, the parameter θ and the argument z only appear in the pgf as $\theta(z-1)$. There may be other parameters as well, and they may appear anywhere in the pgf. Then $P_S(z) = P_N[P_M(z); \theta]$ can be rewritten as

$$P_S(z) = P_N[P_M^T(z); \theta(1 - f_0)].$$

Proof:

$$\begin{aligned} P_S(z) &= P_N[P_M(z); \theta] \\ &= P_N[f_0 + (1 - f_0)P_M^T(z); \theta] \\ &= B\{\theta[f_0 + (1 - f_0)P_M^T(z) - 1]\} \\ &= B\{\theta(1 - f_0)[P_M^T(z) - 1]\} \\ &= P_N[P_M^T(z); \theta(1 - f_0)]. \end{aligned} \quad \square$$

This shows that adding, deleting, or modifying the probability at zero in the secondary distribution does not add a new distribution because it is equivalent to modifying the parameter θ of the primary distribution. This means that, for example, a Poisson primary distribution with a Poisson, zero-truncated Poisson, or zero-modified Poisson secondary distribution will still lead to a Neyman Type A (Poisson-Poisson) distribution.

Example 4.55 Determine the probabilities for a Poisson-zero-modified ETNB distribution where the parameters are $\lambda = 7.5$, $p_0^M = 0.6$, $r = -0.5$, and $\beta = 1$.

From Example 4.45 the secondary probabilities are $f_0 = 0.6$, $f_1 = 0.341421$, $f_2 = 0.042678$, and $f_3 = 0.010670$. From (4.24), $g_0 = \exp[-7.5(1 - 0.6)] = 0.049787$. For the Poisson primary distribution, $a = 0$ and $b = 7.5$. The recursive formula in (4.22) becomes

$$g_k = \frac{\sum_{j=1}^k (7.5j/k) f_j g_{k-j}}{1 - 0(0.6)} = \sum_{j=1}^k \frac{7.5j}{k} f_j g_{k-j}.$$

Then,

$$\begin{aligned} g_1 &= \frac{7.5(1)}{1} 0.341421(0.049787) = 0.127487, \\ g_2 &= \frac{7.5(1)}{2} 0.341421(0.127487) + \frac{7.5(2)}{2} 0.042678(0.049787) = 0.179161, \\ g_3 &= \frac{7.5(1)}{3} 0.341421(0.179161) + \frac{7.5(2)}{3} 0.042678(0.127487) \\ &\quad + \frac{7.5(3)}{3} 0.010670(0.049787) = 0.184112. \end{aligned}$$

Except for slight rounding differences, these probabilities are the same as those obtained in Example 4.52. \square

4.6.8 Further properties of the compound Poisson class

Of central importance within the class of compound frequency models is the class of compound Poisson frequency distributions. Physical motivation for

this model arises from the fact that the Poisson distribution is often a good model to describe the number of claim-causing accidents, and the number of claims from an accident is often itself a random variable. This physical motivation is discussed in the previous subsection. However, there are numerous convenient mathematical properties enjoyed by the compound Poisson class. In particular, those involving recursive evaluation of the probabilities were also discussed in the previous subsection. In addition, there is a close connection between the compound Poisson distributions and the mixed Poisson frequency distributions which is discussed in more detail in Section 4.6.10. Here we consider some other properties of these distributions. The compound Poisson pgf may be expressed as

$$P(z) = \exp\{\lambda[Q(z) - 1]\}, \quad (4.26)$$

where $Q(z)$ is the pgf of the secondary distribution.

Example 4.56 Obtain the pgf for the Poisson-ETNB distribution and show that it looks like the pgf of a Poisson-negative binomial distribution.

The ETNB distribution has pgf

$$Q(z) = \frac{[1 - \beta(z-1)]^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}}$$

for $\beta > 0$, $r > -1$, and $r \neq 0$. Then the Poisson-ETNB distribution has as the logarithm of its pgf

$$\begin{aligned} \ln P(z) &= \lambda \left\{ \frac{[1 - \beta(z-1)]^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}} - 1 \right\} \\ &= \lambda \left\{ \frac{[1 - \beta(z-1)]^{-r} - 1}{1 - (1 + \beta)^{-r}} \right\} \\ &= \mu \{ [1 - \beta(z-1)]^{-r} - 1 \}, \end{aligned}$$

where $\mu = \lambda/[1 - (1 + \beta)^{-r}]$. This defines a compound Poisson distribution with primary mean μ and secondary pgf $[1 - \beta(z-1)]^{-r}$, which is the pgf of a negative binomial random variable, as long as r and hence μ are positive. This illustrates that the probability at zero in the secondary distribution has no impact on the compound Poisson form. Also, the above calculation demonstrates that the Poisson-ETNB pgf $P(z)$, with $\ln P(z) = \mu\{[1 - \beta(z-1)]^{-r} - 1\}$, has parameter space $\{\beta > 0, r > -1, \mu r > 0\}$, a useful observation with respect to estimation and analysis of the parameters. \square

We can compare the skewness (third moment) of these distributions to develop an appreciation of the amount by which the skewness, and hence the tails of these distributions, can vary even when the mean and variance

are fixed. From (4.26) (see Exercise 4.53) and Definition 3.4, the mean and second and third central moments of the compound Poisson distribution are

$$\begin{aligned} \mu'_1 &= \mu = \lambda m'_1, \\ \mu'_2 &= \sigma^2 = \lambda m'_2, \\ \mu'_3 &= \lambda m'_3, \end{aligned} \quad (4.27)$$

where m'_j is the j th raw moment of the secondary distribution. The coefficient of skewness is

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{m'_3}{\lambda^{1/2} (m'_2)^{3/2}}.$$

For the Poisson-binomial distribution, with a bit of algebra (see Exercise 4.54) we obtain

$$\begin{aligned} \mu &= \lambda m q, \\ \sigma^2 &= \mu[1 + (m-1)q], \\ \mu_3 &= 3\sigma^2 - 2\mu + \frac{m-2}{m-1} \frac{(\sigma^2 - \mu)^2}{\mu}. \end{aligned} \quad (4.28)$$

Carrying out similar exercises for the negative binomial, Polya-Aeppli, Neyman Type A, and Poisson-ETNB distributions yields

$$\begin{aligned} \text{Negative binomial:} \quad \mu_3 &= 3\sigma^2 - 2\mu + 2 \frac{(\sigma^2 - \mu)^2}{\mu} \\ \text{Polya-Aeppli:} \quad \mu_3 &= 3\sigma^2 - 2\mu + \frac{3}{2} \frac{(\sigma^2 - \mu)^2}{\mu} \\ \text{Neyman Type A:} \quad \mu_3 &= 3\sigma^2 - 2\mu + \frac{(\sigma^2 - \mu)^2}{\mu} \\ \text{Poisson-ETNB:} \quad \mu_3 &= 3\sigma^2 - 2\mu + \frac{r+2}{r+1} \frac{(\sigma^2 - \mu)^2}{\mu} \end{aligned}$$

For the Poisson-ETNB distribution, the range of r is $-1 < r < \infty$, $r \neq 0$. Note that as $r \rightarrow 0$ the secondary distribution is logarithmic, resulting in the negative binomial distribution.

Note that for fixed mean and variance the third moment only changes through the coefficient in the last term for each of the five distributions. For the Poisson distribution, $\mu_3 = \lambda = 3\sigma^2 - 2\mu$, and so the third term for each expression for μ_3 represents the change from the Poisson distribution. For the Poisson-binomial distribution, if $m = 1$, the distribution is Poisson because it is equivalent to a Poisson-zero-truncated binomial as truncation at zero leaves only probability at 1. Another view is that from (4.28) we have

$$\begin{aligned} \mu_3 &= 3\sigma^2 - 2\mu + \frac{m-2}{m-1} \frac{(m-1)^2 q^4 \lambda^2 m^2}{\lambda m q} \\ &= 3\sigma^2 - 2\mu + (m-2)(m-1)q^3 \lambda m, \end{aligned}$$

which reduces to the Poisson value for μ_3 when $m = 1$. Hence, it is necessary that $m \geq 2$ for non-Poisson distributions to be created. Then the coefficient satisfies

$$0 \leq \frac{m-2}{m-1} < 1.$$

For the Poisson-ETNB, because $r > -1$, the coefficient satisfies

$$1 < \frac{r+2}{r+1} < \infty,$$

noting that when $r = 0$ this refers to the negative binomial distribution. For the Neyman Type A distribution, the coefficient is exactly 1. Hence, these three distributions provide any desired degree of skewness greater than that of the Poisson distribution. Note that the Polya-Aeppli and the negative binomial distributions are special and limiting cases of the Poisson-ETNB with $r = 1$ and $r \rightarrow 0$, respectively.

Example 4.57 The data in Table 4.5 are taken from Hossack et al. [62] and give the distribution of the number of claims on automobile insurance policies in Australia. Determine an appropriate frequency model based on the skewness results of this section.

The mean, variance, and third central moment are 0.1254614, 0.1299599, and 0.1401737, respectively. For these numbers,

$$\frac{\mu_3 - 3\sigma^2 + 2\mu}{(\sigma^2 - \mu)^2/\mu} = 7.543865.$$

From among the Poisson-binomial, negative binomial, Polya-Aeppli, Neyman Type A, and Poisson-ETNB distributions, only the latter is appropriate. For this distribution, an estimate of r can be obtained from

$$7.543865 = \frac{r+2}{r+1}$$

resulting in $r = -0.8471851$. In Example 13.14 a more formal estimation and selection procedure will be applied, but the conclusion will be the same. \square

A very useful property of the compound Poisson class of probability distributions is the fact that it is closed under convolution. We have the following theorem.

Theorem 4.58 Suppose that S_i has a compound Poisson distribution with Poisson parameter λ_i and secondary distribution $\{q_n(i); n = 0, 1, 2, \dots\}$ for $i = 1, 2, 3, \dots, k$. Suppose also that S_1, S_2, \dots, S_k are independent random variables. Then $S = S_1 + S_2 + \dots + S_k$ also has a compound Poisson distribution with Poisson parameter $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_k$ and secondary distribution $\{q_n; n = 0, 1, 2, \dots\}$, where $q_n = [\lambda_1 q_n(1) + \lambda_2 q_n(2) + \dots + \lambda_k q_n(k)]/\lambda$.

Table 4.5 Hossack et al. data

No. of claims	Observed frequency
0	565,664
1	68,714
2	5,177
3	365
4	24
5	6
6+	0

Proof: Let $Q_i(z) = \sum_{n=0}^{\infty} q_n(i)z^n$ for $i = 1, 2, \dots, k$. Then S_i has pgf $P_{S_i}(z) = E(z^{S_i}) = \exp\{\lambda_i[Q_i(z) - 1]\}$. Because the S_i s are independent, S has pgf

$$\begin{aligned} P_S(z) &= \prod_{i=1}^k P_{S_i}(z) \\ &= \prod_{i=1}^k \exp\{\lambda_i[Q_i(z) - 1]\} \\ &= \exp\left[\sum_{i=1}^k \lambda_i Q_i(z) - \sum_{i=1}^k \lambda_i\right] \\ &= \exp\{\lambda[Q(z) - 1]\} \end{aligned}$$

where $\lambda = \sum_{i=1}^k \lambda_i$ and $Q(z) = \sum_{i=1}^k \lambda_i Q_i(z)/\lambda$. The result follows by the uniqueness of the generating function. \square

One main advantage of this result is computational. If we are interested in the sum of independent compound Poisson random variables, then we do not need to compute the distribution of each compound Poisson random variable separately (i.e., recursively using Example 4.50) because Theorem 4.58 implies that a single application of the compound Poisson recursive formula in Example 4.50 will suffice. The following example illustrates this idea.

Example 4.59 Suppose that $k = 2$ and S_1 has a compound Poisson distribution with $\lambda_1 = 2$ and secondary distribution $q_1(1) = 0.2, q_2(1) = 0.7$, and $q_3(1) = 0.1$. Also, S_2 (independent of S_1) has a compound Poisson distribution with $\lambda_2 = 3$ and secondary distribution $q_2(2) = 0.25, q_3(2) = 0.6$, and $q_4(2) = 0.15$. Determine the distribution of $S = S_1 + S_2$.

We have $\lambda = \lambda_1 + \lambda_2 = 2 + 3 = 5$. Then

$$\begin{aligned} q_1 &= 0.4(0.2) + 0.6(0) = 0.08, \\ q_2 &= 0.4(0.7) + 0.6(0.25) = 0.43, \\ q_3 &= 0.4(0.1) + 0.6(0.6) = 0.40, \\ q_4 &= 0.4(0) + 0.6(0.15) = 0.09. \end{aligned}$$

Thus, S has a compound Poisson distribution with Poisson parameter $\lambda = 5$ and secondary distribution $q_1 = 0.08, q_2 = 0.43, q_3 = 0.40$, and $q_4 = 0.09$. Numerical values of the distribution of S may be obtained using the recursive formula

$$\Pr(S = x) = \frac{5}{x} \sum_{n=1}^x n q_n \Pr(S = x - n), \quad x = 1, 2, \dots,$$

beginning with $\Pr(S = 0) = e^{-5}$. \square

In various situations the convolution of negative binomial distributions is of interest. The following example indicates how this distribution may be evaluated.

Example 4.60 (Convolutions of negative binomial distributions). Suppose that N_i has a negative binomial distribution with parameters r_i and β_i for $i = 1, 2, \dots, k$ and that N_1, N_2, \dots, N_k are independent. Determine the distribution of $N = N_1 + N_2 + \dots + N_k$.

The pgf of N_i is $P_{N_i}(z) = [1 - \beta_i(z - 1)]^{-r_i}$ and that of N is $P_N(z) = \prod_{i=1}^k P_{N_i}(z) = \prod_{i=1}^k [1 - \beta_i(z - 1)]^{-r_i}$. If $\beta_i = \beta$ for $i = 1, 2, \dots, k$, then $P_N(z) = [1 - \beta(z - 1)]^{-(r_1 + r_2 + \dots + r_k)}$, and N has a negative binomial distribution with parameters $r = r_1 + r_2 + \dots + r_k$ and β .

If not all the β_i s are identical, however, we may proceed as follows. From Example 4.53,

$$P_{N_i}(z) = [1 - \beta_i(z - 1)]^{-r_i} = e^{\lambda_i [Q_i(z) - 1]}$$

where $\lambda_i = r_i \ln(1 + \beta_i)$ and

$$Q_i(z) = 1 - \frac{\ln[1 - \beta_i(z - 1)]}{\ln(1 + \beta_i)} = \sum_{n=1}^{\infty} q_n(i) z^n$$

with

$$q_n(i) = \frac{[\beta_i / (1 + \beta_i)]^n}{n \ln(1 + \beta_i)}; \quad n = 1, 2, \dots$$

But Theorem 4.58 implies that $N = N_1 + N_2 + \dots + N_k$ has a compound Poisson distribution with Poisson parameter

$$\lambda = \sum_{i=1}^k r_i \ln(1 + \beta_i)$$

and secondary distribution

$$\begin{aligned} q_n &= \sum_{i=1}^k \frac{\lambda_i}{\lambda} q_n(i) \\ &= \frac{\sum_{i=1}^k r_i [\beta_i / (1 + \beta_i)]^n}{n \sum_{i=1}^k r_i \ln(1 + \beta_i)}, \quad n = 1, 2, 3, \dots \end{aligned}$$

The distribution of N may be computed recursively using the formula

$$\Pr(N = n) = \frac{\lambda}{n} \sum_{k=1}^n k q_k \Pr(N = n - k), \quad n = 1, 2, \dots,$$

beginning with $\Pr(N = 0) = e^{-\lambda} = \prod_{i=1}^k (1 + \beta_i)^{-r_i}$ and with λ and q_n as given above. \square

It is not hard to see that Theorem 4.58 is a generalization of Theorem 4.37, which may be recovered with $q_1(i) = 1$ for $i = 1, 2, \dots, k$. Similarly, the decomposition result of Theorem 4.38 may also be extended to compound Poisson random variables, where the decomposition is on the basis of the region of support of the secondary distribution. See Panjer and Willmot [106], Sec. 6.4 or Karlin and Taylor [72], Sec. 16.9 for further details.

4.6.9 Mixed frequency models

Many compound distributions can arise in a way that is very different from compounding. In this section, we examine mixture distributions by treating one or more parameters as being "random" in some sense. This section expands on the ideas discussed in Section 4.6.3 in connection with the gamma mixture of the Poisson distribution being negative binomial.

We assume that the parameter is distributed over the population under consideration (the *collective*) and that the sampling scheme that generates our data has two stages. First, a value of the parameter is selected. Then, given that parameter value, an observation is generated using that parameter value.

In automobile insurance, for example, classification schemes attempt to put individuals into (relatively) homogeneous groups for the purpose of pricing. Variables used to develop the classification scheme might include age, experience, a history of violations, accident history, and other variables. Because there will always be some residual variation in accident risk within each class, mixed distributions provide a framework for modeling this heterogeneity.

Let $P(z|\theta)$ denote the pgf of the number of events (e.g., claims) if the risk parameter is known to be θ . The parameter, θ , might be the Poisson mean, for example, in which case the measurement of risk is the expected number of events in a fixed time period.

Let $U(\theta) = \Pr(\Theta \leq \theta)$ be the cdf of Θ , where Θ is the risk parameter, which is viewed as a random variable. Then $U(\theta)$ represents the probability that, when a value of Θ is selected (e.g., a driver is included in our sample), the value of the risk parameter does not exceed θ . Let $u(\theta)$ be the pf or pdf of Θ . Then

$$P(z) = \int P(z|\theta) u(\theta) d\theta \quad \text{or} \quad P(z) = \sum P(z|\theta_j) u(\theta_j) \quad (4.29)$$

is the unconditional pgf of the number of events (where the formula selected depends on whether Θ is discrete or continuous⁵). The corresponding probabilities are denoted by

$$p_k = \int p_k(\theta)u(\theta)d\theta \text{ or } p_k = \sum p_k(\theta_j)u(\theta_j). \quad (4.30)$$

The **mixing distribution** denoted by $U(\theta)$ may be of the discrete or continuous type or even a combination of discrete and continuous types. **Discrete mixtures** are mixtures of distributions when the mixing function is of the discrete type. Similarly for **continuous mixtures**. This phenomenon was introduced for continuous mixtures of severity distributions in Section 4.4.5 and for finite discrete mixtures in Section 4.2.3.

It should be noted that the mixing distribution is unobservable because the data are drawn from the mixed distribution.

Example 4.61 Demonstrate that the zero-modified distributions may be created by using a two-point mixture.

Suppose

$$P(z) = p \cdot 1 + (1 - p)P_2(z).$$

This is a (discrete) two-point mixture of a degenerate distribution that places all probability at zero and a distribution with pgf $P_2(z)$. From (4.25) this is also a compound Bernoulli distribution. \square

Many mixed models can be constructed beginning with a simple distribution. Two examples are given here.

Example 4.62 Determine the pf for a mixed binomial with a beta mixing distribution. This distribution is called binomial-beta, negative hypergeometric, or Polya-Eggenberger.

The beta distribution has pdf

$$u(q) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1}(1-q)^{b-1}, \quad a > 0, b > 0.$$

⁵We could have written the more general $P(z) = \int P(z|\theta) dU(\theta)$, which would include situations where Θ has a distribution that is partly continuous and partly discrete.

Then the mixed distribution has probabilities

$$\begin{aligned} p_k &= \int_0^1 \binom{m}{k} q^k (1-q)^{m-k} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1} (1-q)^{b-1} dq \\ &= \frac{\Gamma(a+b)\Gamma(m+1)\Gamma(a+k)\Gamma(b+m-k)}{\Gamma(a)\Gamma(b)\Gamma(k+1)\Gamma(m-k+1)\Gamma(a+b+m)} \\ &= \frac{\binom{-a}{k} \binom{-b}{m-k}}{\binom{-a-b}{m}}, \quad k = 0, 1, 2, \dots \end{aligned}$$

\square

Example 4.63 Determine the pf for a mixed negative binomial distribution with mixing on the parameter $p = (1 + \beta)^{-1}$. Let p have a beta distribution. The mixed distribution is called the generalized Waring.

Arguing as in Example 4.62 we have

$$\begin{aligned} p_k &= \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a+r-1} (1-p)^{b+k-1} dp \\ &= \frac{\Gamma(r+k)}{\Gamma(r)\Gamma(k+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+r)\Gamma(b+k)}{\Gamma(a+r+b+k)}, \quad k = 0, 1, 2, \dots \end{aligned}$$

When $b = 1$, this distribution is called the Waring distribution. When $r = b = 1$, it is termed the Yule distribution. \square

4.6.10 Poisson mixtures

If we let $p_k(\theta)$ in (4.30) have the Poisson distribution, this leads to a class of distributions with useful properties. A simple example of a Poisson mixture is the two-point mixture.

Example 4.64 Suppose drivers can be classified as "good drivers" and "bad drivers," each group with its own Poisson distribution. Determine the pf for this model and fit it to the data from Example 12.56. This model and its application to the data set are from Tröbliger [130].

From (4.30) the pf is

$$p_k = p \frac{e^{-\lambda_1} \lambda_1^k}{k!} + (1-p) \frac{e^{-\lambda_2} \lambda_2^k}{k!}.$$

The maximum likelihood estimates⁶ were calculated by Tröbliger to be $\hat{p} = 0.94$, $\hat{\lambda}_1 = 0.11$ and $\hat{\lambda}_2 = 0.70$. This means that about 94% of drivers

⁶Maximum likelihood estimation is discussed in Section 12.2.

were “good” with a risk of $\lambda_1 = 0.11$ expected accidents per year and 6% were “bad” with a risk of $\lambda_2 = 0.70$ expected accidents per year. Note that it is not possible to return to the data set and identify which were the bad drivers. \square

This example illustrates two important points about finite mixtures. First, the model is probably oversimplified in the sense that risks (e.g., drivers) probably exhibit a continuum of risk levels rather than just two. The second point is that finite mixture models have a lot of parameters to be estimated. The simple two-point Poisson mixture above has three parameters. Increasing the number of distributions in the mixture to r will then involve $r - 1$ mixing parameters in addition to the total number of parameters in the r component distributions. As a result of this, continuous mixtures are frequently preferred.

The class of mixed Poisson distributions has some interesting properties that will be developed here.

Let $P(z)$ be the pgf of a mixed Poisson distribution with arbitrary mixing distribution $U(\theta)$. Then (with formulas given for the continuous case), by introducing a scale parameter λ , we have

$$\begin{aligned} P(z) &= \int e^{\lambda\theta(z-1)} u(\theta) d\theta = \int [e^{\lambda(z-1)}]^\theta u(\theta) d\theta \\ &= E \left\{ [e^{\lambda(z-1)}]^\theta \right\} = M_\Theta [\lambda(z-1)], \end{aligned} \quad (4.31)$$

where $M_\Theta(z)$ is the mgf of the mixing distribution.

Therefore, $P'(z) = \lambda M'_\Theta[\lambda(z-1)]$ and with $z = 1$ we obtain $E(N) = \lambda E(\Theta)$, where N has the mixed Poisson distribution. Also, $P''(z) = \lambda^2 M''_\Theta[\lambda(z-1)]$, implying that $E[N(N-1)] = \lambda^2 E(\Theta^2)$ and therefore

$$\begin{aligned} \text{Var}(N) &= E[N(N-1)] + E(N) - [E(N)]^2 \\ &= \lambda^2 E(\Theta^2) + E(N) - \lambda^2 [E(\Theta)]^2 \\ &= \lambda^2 \text{Var}(\Theta) + E(N) \\ &> E(N) \end{aligned}$$

and thus for mixed Poisson distributions the variance is always greater than the mean.

Douglas [29] proves that for any mixed Poisson distribution the mixing distribution is unique. This means that two different mixing distributions cannot lead to the same mixed Poisson distribution. This allows us to identify the mixing distribution in some cases.

There is also an important connection between mixed Poisson distributions and compound Poisson distributions.

Definition 4.65 A distribution is said to be *infinitely divisible* if for all values of $n = 1, 2, 3, \dots$ its characteristic function $\varphi(z)$ can be written as

$$\varphi(z) = [\varphi_n(z)]^n,$$

where $\varphi_n(z)$ is the characteristic function of some random variable.

In other words, taking the $(1/n)$ th power of the characteristic function still results in a characteristic function. The characteristic function is defined as follows.

Definition 4.66 The *characteristic function* of a random variable X is

$$\varphi_X(z) = E(e^{izX}) = E(\cos zX + i \sin zX),$$

where $i = \sqrt{-1}$.

In Definition 4.65, “characteristic function” could have been replaced by “moment generating function” or “probability generating function,” or some other transform. That is, if the definition is satisfied for one of these transforms, it will be satisfied for all others which exist for the particular random variable. We choose the characteristic function because it exists for all distributions while the moment generating function does not exist for some distributions with heavy tails. Because many earlier results involved probability generating functions, it is useful to note the relationship between it and the characteristic function.

Theorem 4.67 If the probability generating function exists for a random variable X , then $P_X(z) = \varphi(-i \ln z)$ and $\varphi_X(z) = P(e^{iz})$.

Proof:

$$P_X(z) = E(z^X) = E(e^{X \ln z}) = E[e^{-i(i \ln z)X}] = \varphi_X(-i \ln z)$$

and

$$\varphi_X(z) = E(e^{izX}) = E[(e^{iz})^X] = P_X(e^{iz}). \quad \square$$

The following distributions, among others, are infinitely divisible: normal, gamma, Poisson, negative binomial. The binomial distribution is not infinitely divisible because the exponent m in its pgf must take on integer values. Dividing m by $n = 1, 2, 3, \dots$ will result in nonintegral values. In fact, no distributions with a finite range of support (the range over which positive probabilities exist) can be infinitely divisible. Now to the important result.

Theorem 4.68 Suppose $P(z)$ is a mixed Poisson pgf with an infinitely divisible mixing distribution. Then $P(z)$ is also a compound Poisson pgf and may be expressed as

$$P(z) = e^{\lambda[P_2(z)-1]},$$

where $P_2(z)$ is a pgf. If one adopts the convention that $P_2(0) = 0$, then $P_2(z)$ is unique.

A proof can be found in Feller [35], Ch. 12. If one chooses any infinitely divisible mixing distribution, the corresponding mixed Poisson distribution can be equivalently described as a compound Poisson distribution. For some distributions, this is a distinct advantage when carrying out numerical work because the recursive formula (4.22) can be used in evaluating the probabilities once the secondary distribution is identified. For most cases, this identification is easily carried out. A second advantage is that, because the same distribution can be motivated in two different ways, a specific explanation is not required in order to use it. Conversely, the fact that one of these models fits well does not imply that it is the result of mixing or compounding. For example, the fact that claims follow a negative binomial distribution does not imply that individuals have the Poisson distribution and the Poisson parameter has a gamma distribution.

Example 4.69 Use the above results and (4.31) to demonstrate that a gamma mixture of Poisson variables is negative binomial.

If the mixing distribution is gamma, it has the following moment generating function (as derived in Example 3.15 and where β plays the role of $1/\theta$):

$$M_{\Theta}(t) = \left(\frac{\beta}{\beta - t} \right)^{\alpha}, \quad \beta > 0, \alpha > 0, t < \beta.$$

It is clearly infinitely divisible because $[M_{\Theta}(t)]^{1/n}$ is the mgf of a gamma distribution with parameters α/n and β . Then the pgf of the mixed Poisson distribution is

$$P(z) = \left[\frac{\beta}{\beta - \lambda(z-1)} \right]^{\alpha} = \left[1 - \frac{\lambda}{\beta}(z-1) \right]^{-\alpha},$$

which is the form of the pgf of the negative binomial distribution where the negative binomial parameter r is equal to α and the parameter β is equal to λ/β . \square

It was shown in Example 4.53 that a compound Poisson distribution with a logarithmic secondary distribution is a negative binomial distribution. Therefore the theorem holds true for this case. It is not difficult to see that, if $u(\theta)$ is the pf for any discrete random variable with pgf $P_{\Theta}(z)$, then the pgf of the mixed Poisson distribution is $P_{\Theta}[e^{\lambda(z-1)}]$, a compound distribution with a Poisson secondary distribution.

Example 4.70 Demonstrate that the Neyman Type A distribution can be obtained by mixing.

If in (4.31) the mixing distribution has pgf

$$P_{\Theta}(z) = e^{\mu(z-1)},$$

then the mixed Poisson distribution has pgf

$$P(z) = \exp\{\mu[e^{\lambda(z-1)} - 1]\},$$

the pgf of a compound Poisson with a Poisson secondary distribution, that is, the Neyman Type A distribution. \square

A further interesting result obtained by Holgate [60] is that, if a mixing distribution is absolutely continuous and unimodal, then the resulting mixed Poisson distribution is also unimodal. Multimodality can occur when discrete mixing functions are used. For example, the Neyman Type A distribution can have more than one mode. The reader should try this calculation for various combinations of the two parameters.

Most continuous distributions in this book involve a scale parameter. This means that scale changes to distributions do not cause a change in the form of the distribution, only in the value of its scale parameter. For the mixed Poisson distribution, with pgf (4.31), any change in λ is equivalent to a change in the scale parameter of the mixing distribution. Hence, it may be convenient to simply set $\lambda = 1$ where a mixing distribution with a scale parameter is used.

Example 4.71 Show that a mixed Poisson with an inverse Gaussian mixing distribution is the same as a Poisson-ETNB distribution with $r = -0.5$.

The inverse Gaussian distribution is described in Appendix A. It has pdf

$$f(x) = \left(\frac{\theta}{2\pi x^3} \right)^{1/2} \exp \left[-\frac{\theta}{2x} \left(\frac{x-\mu}{\mu} \right)^2 \right], \quad x > 0,$$

which is conveniently rewritten as

$$f(x) = \frac{\mu}{(2\pi\beta x^3)^{1/2}} \exp \left[-\frac{(x-\mu)^2}{2\beta x} \right], \quad x > 0,$$

where $\beta = \mu^2/\theta$. The mgf of this distribution is (see Exercise 3.24)

$$M(t) = \exp \left\{ -\frac{\mu}{\beta} [(1 - 2\beta t)^{1/2} - 1] \right\}.$$

Hence, the inverse Gaussian distribution is infinitely divisible ($[M(t)]^{1/n}$ is the mgf of an inverse Gaussian distribution with μ replaced by μ/n). From (4.31) with $\lambda = 1$, the pgf of the mixed distribution is

$$P(z) = \exp \left(-\frac{\mu}{\beta} \{ [1 + 2\beta(1-z)]^{1/2} - 1 \} \right).$$

By setting

$$\lambda = \frac{\mu}{\beta} [(1 + 2\beta)^{1/2} - 1]$$

Table 4.6 Pairs of compound and mixed Poisson distributions

Name	Compound secondary distribution	Mixing distribution
Negative binomial	logarithmic	gamma
Neyman-A	Poisson	Poisson
Poisson-inverse Gaussian	ETNB ($r = -0.5$)	inverse Gaussian

and

$$P_2(z) = \frac{[1 - 2\beta(z-1)]^{1/2} - (1 + 2\beta)^{1/2}}{1 - (1 + 2\beta)^{1/2}},$$

we see that

$$P(z) = \exp\{\lambda[P_2(z) - 1]\},$$

where $P_2(z)$ is the pgf of the extended truncated negative binomial distribution with $r = -\frac{1}{2}$.

Hence, the Poisson-inverse Gaussian distribution is a compound Poisson distribution with an ETNB ($r = -\frac{1}{2}$) secondary distribution. \square

The relationships between mixed and compound Poisson distributions are given in Table 4.6.

In this chapter, we focused on distributions that are easily handled computationally. Although many other discrete distributions are available, we believe that those discussed form a sufficiently rich class for most problems.

4.6.11 Effect of exposure on frequency

Assume that the current portfolio consists of n entities, each of which could produce claims. Let N_j be the number of claims produced by the j th entity. Then $N = N_1 + \cdots + N_n$. If we assume that the N_j are independent and identically distributed, then

$$P_N(z) = [P_{N_1}(z)]^n.$$

Now suppose the portfolio is expected to expand to n^* entities with frequency N^* . Then

$$P_{N^*}(z) = [P_{N_1}(z)]^{n^*} = [P_N(z)]^{n^*/n}.$$

Thus, if N is infinitely divisible, the distribution of N^* will have the same form as that of N , but with modified parameters.

Example 4.72 *It has been determined from past studies that the number of workers compensation claims for a group of 300 employees in a certain occupation class has the negative binomial distribution with $\beta = 0.3$ and $r = 10$. Determine the frequency distribution for a group of 500 such individuals.*

The pgf of N^* is

$$\begin{aligned} P_{N^*}(z) &= [P_N(z)]^{500/300} = \{[1 - 0.3(z-1)]^{-10}\}^{500/300} \\ &= [1 - 0.3(z-1)]^{-16.67}, \end{aligned}$$

which is negative binomial with $\beta = 0.3$ and $r = 16.67$. \square

For the $(a, b, 0)$ class, all members except the binomial have this property. For the $(a, b, 1)$ class, none of the members do. For compound distributions, it is the primary distribution that must be infinitely divisible. In particular, compound Poisson and compound negative binomial (including the geometric) distributions will be preserved under an increase in exposure. Earlier, some reasons were given to support the use of zero-modified distributions. If exposure adjustments are anticipated, it may be better to choose a compound model, even if the fit is not quite as good. It should be noted that compound models have the ability to place large amounts of probability at zero.

4.6.12 An inventory of discrete distributions

We have introduced the simple $(a, b, 0)$ class, generalized to the $(a, b, 1)$ class, and then used compounding and mixing to create a larger class of distributions. Calculation of the probabilities of these distributions can be carried out by using simple recursive procedures. In this section we note that there are relationships among the various distributions similar to those of Section 4.5.2. The specific relationships are given in Table 4.7.

It is clear from earlier developments that members of the $(a, b, 0)$ class are special cases of members of the $(a, b, 1)$ class and that zero-truncated distributions are special cases of zero-modified distributions. The limiting cases are best discovered through the probability generating function, as was done on page 79 where the Poisson distribution is shown to be a limiting case of the negative binomial distribution.

We have not listed compound distributions where the primary distribution is one of the two parameter models such as the negative binomial or Poisson-inverse Gaussian. This was done because these distributions are often themselves compound Poisson distributions and, as such, are generalizations of distributions already presented. This collection forms a particularly rich set of distributions in terms of shape. However, many other distributions are also possible. Many others are discussed in Johnson, Kotz, and Kemp [69], Douglas [29], and Panjer and Willmot [106].

4.6.13 Exercises

4.40 For each of the data sets in Exercises 12.96 and 12.98 on page 400 calculate values similar to those in Table 4.2. For each, determine the most appropriate model from the $(a, b, 0)$ class.

Table 4.7 Relationships among discrete distributions

Distribution	Is a special case of	Is a limiting case of
Poisson	ZM Poisson	negative binomial Poisson-binomial Poisson-inv. Gaussian Polya-Aeppli ^a Neyman-A ^b
ZT Poisson	ZM Poisson	ZT negative binomial
ZM Poisson		ZM negative binomial
geometric	negative binomial, ZM geometric	geometric-Poisson
ZT geometric	ZT negative binomial	
ZM geometric	ZM negative binomial	
logarithmic		ZT negative binomial
ZM logarithmic		ZM negative binomial
binomial	ZM binomial	
negative binomial	ZM negative binomial, Poisson-ETNB	
Poisson-inverse Gaussian	Poisson-ETNB	
Polya-Aeppli	Poisson-ETNB	
Neyman-A		Poisson-ETNB

^aAlso called Poisson-geometric.^bAlso called Poisson-Poisson.

4.41 Calculate $\Pr(N = 0)$, $\Pr(N = 1)$, and $\Pr(N = 2)$ for each of the following distributions.

- Poisson($\lambda = 4$)
- Geometric($\beta = 4$)
- Negative binomial($r = 2, \beta = 2$)
- Binomial($m = 8, q = 0.5$)
- Logarithmic($\beta = 4$)
- ETNB($r = -0.5, \beta = 4$)
- Poisson-inverse Gaussian($\lambda = 2, \beta = 4$)
- Zero-modified geometric($p_0^M = 0.5, \beta = 4$)
- Poisson-Poisson(Neyman Type A)($\lambda_{\text{primary}} = 4, \lambda_{\text{secondary}} = 1$)
- Poisson-ETNB($\lambda = 4, r = 2, \beta = 0.5$)
- Poisson-zero-modified geometric distribution($\lambda = 8, p_0^M = 0.5, r = 2, \beta = 0.5$)

4.42 The **moment generating function** (mgf) for discrete variables is defined as

$$M_N(z) = E(e^{zN}) = \sum_{k=0}^{\infty} p_k e^{zk}.$$

Demonstrate that $P_N(z) = M_N(\ln z)$. Use the fact that $E(N^k) = M_N^{(k)}(0)$ to show that $P'(1) = E(N)$ and $P''(1) = E[N(N-1)]$.

4.43 Use your knowledge of the permissible ranges for the parameters of the Poisson, negative binomial, and binomial to determine all possible values of a and b for these members of the $(a, b, 0)$ class. Because these are the only members of the class, all other pairs must not lead to a legitimate probability distribution (nonnegative values that sum to 1). Show that the pair $a = -1$ and $b = 1.5$ (which is not on the list of possible values) does not lead to a legitimate distribution.

4.44 Show that for the negative binomial distribution with any $\beta > 0$ and $r > -1$, but $r \neq 0$, the successive values of p_k given by (4.15) are, for any p_1 , positive and $\sum_{k=1}^{\infty} p_k < \infty$.

4.45 Show that when, in the zero-truncated negative binomial distribution, $r \rightarrow 0$ the pf is as given in (4.16).

4.46 Show that the pgf of the logarithmic distribution is as given in (4.17).

4.47 Show that for the Sibuya distribution, which is the ETNB distribution with $-1 < r < 0$ and $\beta \rightarrow \infty$, the mean does not exist (that is, the sum which defines the mean does not converge). Because this random variable takes on nonnegative values, this also shows that no other positive moments exist.

4.48 A frequency model that has not been mentioned to this point is the **zeta distribution**. It is a zero-truncated distribution with $p_k^T = k^{-(\rho+1)}/\zeta(\rho+1)$, $k = 1, 2, \dots, \rho > 0$. The denominator is the zeta function, which must be evaluated numerically as $\zeta(\rho+1) = \sum_{k=1}^{\infty} k^{-(\rho+1)}$. The zero-modified zeta distribution can be formed in the usual way. More information can be found in Luong and Doray [88]. Verify that the zeta distribution is not a member of the $(a, b, 1)$ class.

4.49 Do all the members of the $(a, b, 0)$ class satisfy the condition of Theorem 4.54? For those that do, identify the parameter (or function of its parameters) that plays the role of θ in the theorem.

4.50 For $i = 1, \dots, n$ let S_i have independent compound Poisson frequency distributions with Poisson parameter λ_i and a secondary distribution with pgf $P_2(z)$. Note that all n of the variables have the same secondary distribution. Determine the distribution of $S = S_1 + \dots + S_n$.

4.51 Show that the following three distributions are identical: (1) geometric-geometric, (2) Bernoulli-geometric, (3) zero-modified geometric. That is, for any one of the distributions with arbitrary parameters, show that there is a member of the other two distribution types that has the same pf or pgf.

4.52 Show that the binomial-geometric and negative binomial-geometric (with negative binomial parameter r a positive integer) distributions are identical.

4.53 Show that, for any pgf, $P^{(k)}(1) = E[N(N-1)\cdots(N-k+1)]$ provided the expectation exists. Here $P^{(k)}(z)$ indicates the k th derivative. Use this result to confirm the three moments as given in (4.27).

4.54 Verify the three moments as given in (4.28).

4.55 Show that the negative binomial-Poisson compound distribution is the same as a mixed Poisson distribution with a negative binomial mixing distribution.

4.56 For $i = 1, \dots, n$ let N_i have a mixed Poisson distribution with parameter λ . Let the mixing distribution for N_i have pgf $P_i(z)$. Show that $N = N_1 + \cdots + N_n$ has a mixed Poisson distribution and determine the pgf of the mixing distribution.

4.57 Let N have a Poisson distribution with (given that $\Theta = \theta$) parameter $\lambda\theta$. Let the distribution of the random variable Θ have a scale parameter. Show that the mixed distribution does not depend on the value of λ .

4.58 Let N have a Poisson distribution with (given that $\Theta = \theta$) parameter θ . Let the distribution of the random variable Θ have pdf $u(\theta) = \alpha^2(\alpha + 1)^{-1}(\theta + 1)e^{-\alpha\theta}$, $\theta > 0$. Determine the pf of the mixed distribution. Also, show that the mixed distribution is also a compound distribution.

4.59 For the discrete counting random variable N with probabilities $p_n = \Pr(N = n)$; $n = 0, 1, 2, \dots$, let $a_n = \Pr(N > n) = \sum_{k=n+1}^{\infty} p_k$; $n = 0, 1, 2, \dots$.

- Demonstrate that $E(N) = \sum_{n=0}^{\infty} a_n$.
- Demonstrate that $A(z) = \sum_{n=0}^{\infty} a_n z^n$ and $P(z) = \sum_{n=0}^{\infty} p_n z^n$ are related by $A(z) = [1 - P(z)] / (1 - z)$. What happens as $z \rightarrow 1$?
- Suppose that N has the negative binomial distribution

$$p_n = \binom{n+r-1}{n} \left(\frac{1}{1+\beta} \right)^r \left(\frac{\beta}{1+\beta} \right)^n, \quad n = 0, 1, 2, \dots,$$

where r is a positive integer. Prove that

$$a_n = \beta \sum_{k=1}^r \binom{n+k-1}{n} \left(\frac{1}{1+\beta} \right)^k \left(\frac{\beta}{1+\beta} \right)^n, \quad n = 0, 1, 2, \dots$$

(d) Suppose that N has the Sibuya distribution with pgf $P(z) = 1 - (1 - z)^{-r}$, $-1 < r < 0$. Prove that

$$p_n = \frac{(-r)\Gamma(n+r)}{n!\Gamma(1+r)}, \quad n = 1, 2, 3, \dots,$$

and that

$$a_n = \binom{n+r}{n}, \quad n = 0, 1, 2, \dots$$

(e) Suppose that N has the mixed Poisson distribution with

$$p_n = \int_0^{\infty} \frac{(\lambda\theta)^n e^{-\lambda\theta}}{n!} dU(\theta), \quad n = 0, 1, 2, \dots,$$

where $U(\theta)$ is a cumulative distribution function. Prove that

$$a_n = \lambda \int_0^{\infty} \frac{(\lambda\theta)^n e^{-\lambda\theta}}{n!} [1 - U(\theta)] d\theta, \quad n = 0, 1, 2, \dots$$

4.60 Consider the mixed Poisson distribution

$$p_n = \Pr(N = n) = \int_0^1 \frac{(\lambda\theta)^n e^{-\lambda\theta}}{n!} U'(\theta) d\theta, \quad n = 0, 1, \dots,$$

where $U(\theta) = 1 - (1 - \theta)^k$, $0 < \theta < 1$, $k = 1, 2, \dots$.

(a) Prove that

$$p_n = k e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+n} (m+k-1)!}{m! (m+k+n)!}, \quad n = 0, 1, \dots$$

(b) Using Exercise 4.59 prove that

$$\Pr(N > n) = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+n+1} (m+k)!}{m! (m+k+n+1)!}.$$

(c) When $k = 1$, prove that

$$p_n = \frac{1 - \sum_{m=0}^n \lambda^m e^{-\lambda} / m!}{\lambda}, \quad n = 0, 1, 2, \dots$$

4.61 Consider the mixed Poisson distribution

$$p_n = \int_0^{\infty} \frac{(\lambda\theta)^n e^{-\lambda\theta}}{n!} u(\theta) d\theta, \quad n = 0, 1, \dots,$$

where the pdf $u(\theta)$ is that of the positive stable distribution (see, for example, Feller [36], pp. 448, 583) given by

$$u(\theta) = \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} (-1)^{k-1} \theta^{-k\alpha-1} \sin(k\alpha\pi), \quad \theta > 0,$$

where $0 < \alpha < 1$. The Laplace transform is $\int_0^{\infty} e^{-s\theta} u(\theta) d\theta = \exp(-s^\alpha)$, $s \geq 0$. Prove that $\{p_n; n = 0, 1, \dots\}$ is a compound Poisson distribution with Sibuya secondary distribution (this mixed Poisson distribution is sometimes called a discrete stable distribution).

4.62 Consider a mixed Poisson distribution with a reciprocal inverse Gaussian distribution as the mixing distribution.

- Use Exercise 4.36 to show that this distribution is the convolution of a negative binomial distribution and a Poisson-ETNB distribution with $r = -\frac{1}{2}$ (i.e., a Poisson-inverse Gaussian distribution).
- Show that the mixed Poisson distribution in (a) is a compound Poisson distribution and identify the secondary distribution.

5

Frequency and severity with coverage modifications

5.1 INTRODUCTION

We have seen a variety of examples that involve functions of random variables. In this chapter we relate those functions to insurance applications. Throughout this chapter we assume that all random variables have support on all or a subset of the nonnegative real numbers. At times in this chapter and later in the text we will need to distinguish between a random variable that measures the payment per loss (so zero is a possibility, taking place when there is a loss without a payment) and a variable that measures the payment per payment (the random variable is not defined when there is no payment). For notation, a per-loss variable will be denoted Y^L and a per-payment variable will be noted Y^P . When the distinction is not material (for example, setting a maximum payment does not create a difference), the superscript will be left off.

5.2 DEDUCTIBLES

Insurance policies are often sold with a per-loss deductible of d . When the loss, x , is at or below d , the insurance pays nothing. When the loss is above

d , the insurance pays $x - d$. In the language of Chapter 3 such a deductible can be defined as follows.

Definition 5.1 An ordinary deductible modifies a random variable into either the excess loss or left censored and shifted variable (see Definition 3.6). The difference depends on whether the result of applying the deductible is to be per payment or per loss, respectively.

This concept has already been introduced along with formulas for determining its moments. The per-payment variable is

$$Y^P = \begin{cases} \text{undefined,} & X \leq d, \\ X - d, & X > d, \end{cases}$$

while the per-loss variable is

$$Y^L = \begin{cases} 0, & X \leq d, \\ X - d, & X > d. \end{cases}$$

Note that the per-payment variable $Y^P = Y^L | Y^L > 0$. That is, the per-payment variable is the per-loss variable conditioned on the loss being positive. For the excess loss/per-payment variable, the density function is

$$f_{Y^P}(y) = \frac{f_X(y+d)}{S_X(d)}, \quad y > 0, \quad (5.1)$$

noting that for a discrete distribution the density function need only be replaced by the probability function. Other key functions are

$$\begin{aligned} S_{Y^P}(y) &= \frac{S_X(y+d)}{S_X(d)}, \\ F_{Y^P}(y) &= \frac{F_X(y+d) - F_X(d)}{1 - F_X(d)}, \\ h_{Y^P}(y) &= \frac{f_X(y+d)}{S_X(y+d)} = h_X(y+d). \end{aligned}$$

Note that as a per-payment variable the excess loss variable places no probability at 0.

The left censored and shifted variable has discrete probability at zero of $F_X(d)$ representing the probability that a payment of zero is made because the loss did not exceed d . Above zero, the density function is

$$f_{Y^L}(y) = f_X(y+d), \quad y > 0, \quad (5.2)$$

while the other key functions are¹ (for $y \geq 0$)

$$\begin{aligned} S_{Y^L}(y) &= S_X(y+d), \\ F_{Y^L}(y) &= F_X(y+d). \end{aligned}$$

It is important to recognize that when counting claims on a per payment, changing the deductible will change the frequency with which payments are made (while the frequency of losses will be unchanged). The nature of these changes will be discussed in Section 5.6.

Example 5.2 Determine similar quantities for a Pareto distribution with $\alpha = 3$ and $\theta = 2000$ for an ordinary deductible of 500.

Using the above formulas, for the excess loss variable,

$$\begin{aligned} f_{Y^P}(y) &= \frac{3(2,000)^3(2,000+y+500)^{-4}}{(2,000)^3(2,000+500)^{-3}} = \frac{3(2,500)^3}{(2,500+y)^4}, \\ S_{Y^P}(y) &= \left(\frac{2,500}{2,500+y} \right)^3, \\ F_{Y^P}(y) &= 1 - \left(\frac{2,500}{2,500+y} \right)^3, \\ h_{Y^P}(y) &= \frac{3}{2,500+y}. \end{aligned}$$

Note that this is a Pareto distribution with $\alpha = 3$ and $\theta = 2,500$. For the left censored and shifted variable,

$$\begin{aligned} f_{Y^L}(y) &= \begin{cases} 0.488, & y = 0, \\ \frac{3(2,000)^3}{(2,500+y)^4}, & y > 0, \end{cases} & S_{Y^L}(y) &= \begin{cases} 0.512, & y = 0, \\ \frac{(2,000)^3}{(2,500+y)^3}, & y > 0, \end{cases} \\ F_{Y^L}(y) &= \begin{cases} 0.488, & y = 0, \\ 1 - \frac{(2,000)^3}{(2,500+y)^3}, & y > 0, \end{cases} & h_{Y^L}(y) &= \begin{cases} \text{undefined,} & y = 0, \\ \frac{3}{2,500+y}, & y > 0. \end{cases} \end{aligned}$$

Figure 5.1 contains a plot of the densities. The modified densities are created as follows. For the excess loss variable, take the portion of the original density from 500 and above. Then shift it to start at zero and multiply it by a constant so that the area under it is still 1. The left censored and shifted variable also takes the original density function above 500 and shifts it to the origin, but then leaves it alone. The remaining probability is concentrated at zero, rather than spread out. \square

¹The hazard rate function is not presented because it is not defined at zero, making it of limited value. Note that for the excess loss variable the hazard rate function is simply shifted.

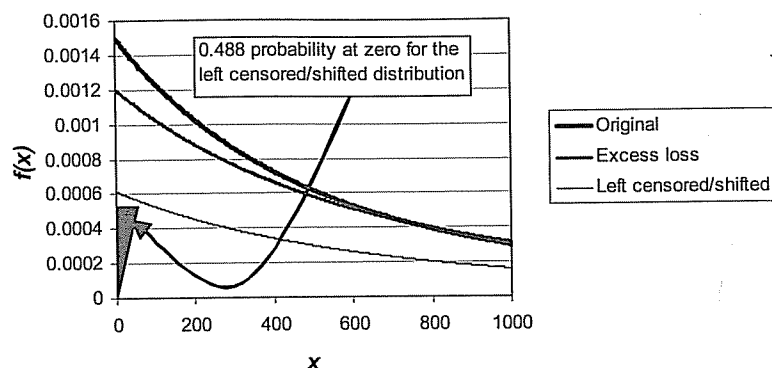


Fig. 5.1 Densities for Example 5.2.

An alternative to the ordinary deductible is the franchise deductible. This deductible differs from the ordinary deductible in that, when the loss exceeds the deductible, the loss is paid in full. One example is in disability insurance where, for example, if a disability lasts seven or fewer days, no benefits are paid. However, if the disability lasts more than seven days, daily benefits are paid retroactively to the onset of the disability.

Definition 5.3 A franchise deductible modifies the ordinary deductible by adding the deductible when there is a positive amount paid.

The terms *left censored and shifted* and *excess loss* are not used here. Because this modification is unique to insurance applications, we will use per-payment and per-loss terminology. The per-loss variable is

$$Y^L = \begin{cases} 0, & X \leq d, \\ X, & X > d, \end{cases}$$

while the per-payment variable is

$$Y^P = \begin{cases} \text{undefined}, & X \leq d, \\ X, & X > d. \end{cases}$$

Note that, as usual, the per-payment variable is a conditional random variable. The quantities derived above for the ordinary deductible are now

$$f_{Y^L}(y) = \begin{cases} F_X(d), & y = 0, \\ f_X(y), & y > d, \end{cases} \quad S_{Y^L}(y) = \begin{cases} S_X(d), & 0 \leq y \leq d, \\ S_X(y), & y > d, \end{cases}$$

$$F_{Y^L}(y) = \begin{cases} F_X(d), & 0 \leq y \leq d, \\ F_X(y), & y > d, \end{cases} \quad h_{Y^L}(y) = \begin{cases} 0, & 0 < y < d, \\ h_X(y), & y > d, \end{cases}$$

for the per loss variable and

$$f_{Y^P}(y) = \frac{f_X(y)}{S_X(d)}, \quad y > d, \quad S_{Y^P}(y) = \begin{cases} 1, & 0 \leq y \leq d, \\ \frac{S_X(y)}{S_X(d)}, & y > d, \end{cases}$$

$$F_{Y^P}(y) = \begin{cases} 0, & 0 \leq y \leq d, \\ \frac{F_X(y) - F_X(d)}{1 - F_X(d)}, & y > d, \end{cases}$$

$$h_{Y^P}(y) = \begin{cases} 0, & 0 < y < d, \\ h_X(y), & y > d \end{cases}$$

for the per-payment variable.

Example 5.4 Repeat Example 5.2 for a franchise deductible.

Using the above formulas for the per-payment variable, for $y > 500$,

$$f_{Y^P}(y) = \frac{3(2,000)^3(2,000 + y)^{-4}}{(2,000)^3(2,000 + 500)^{-3}} = \frac{3(2,500)^3}{(2,000 + y)^4},$$

$$S_{Y^P}(y) = \left(\frac{2,500}{2,000 + y} \right)^3,$$

$$F_{Y^P}(y) = 1 - \left(\frac{2,500}{2,000 + y} \right)^3,$$

$$h_{Y^P}(y) = \frac{3}{2,000 + y}.$$

For the per-loss variable,

$$f_{Y^L}(y) = \begin{cases} 0.488, & y = 0, \\ \frac{3(2,000)^3}{(2,000 + y)^4}, & y > 500, \end{cases}$$

$$S_{Y^L}(y) = \begin{cases} 0.512, & 0 \leq y \leq 500, \\ \frac{(2,000)^3}{(2,000 + y)^3}, & y > 500, \end{cases}$$

$$F_{Y^L}(y) = \begin{cases} 0.488, & 0 \leq y \leq 500, \\ 1 - \frac{(2,000)^3}{(2,000 + y)^3}, & y > 500, \end{cases}$$

$$h_{Y^L}(y) = \begin{cases} 0, & 0 < y < 500, \\ \frac{3}{2,000 + y}, & y > 500. \end{cases}$$

□

Expected costs for the two types of deductible may also be calculated.

Theorem 5.5 For an ordinary deductible, the expected cost per loss is $E(X) - E(X \wedge d)$ and the expected cost per payment is $[E(X) - E(X \wedge d)]/[1 - F(d)]$. For a franchise deductible the expected cost per loss is $E(X) - E(X \wedge d) + d[1 - F(d)]$ and the expected cost per payment is $[E(X) - E(X \wedge d)]/[1 - F(d)] + d$.

Proof: For the per-loss expectation with an ordinary deductible, we have, from (3.7) and (3.10) that the expectation is $E(X) - E(X \wedge d)$. From (5.1) and (5.2) we see that, to change to a per-payment basis, division by $1 - F(d)$ is required. The adjustments for the franchise deductible come from the fact that when there is a payment it will exceed that for the ordinary deductible by d . \square

Example 5.6 Determine the four expectations for the Pareto distribution from Examples 5.2 and 5.4 using a deductible of 500.

Expectations could be derived directly from the density functions obtained in Examples 5.2 and 5.4. Using Theorem 5.5 and recognizing that we have a Pareto distribution, we can also look up the required values (the formulas are in Appendix A). That is,

$$F(500) = 1 - \left(\frac{2,000}{2,000 + 500} \right)^3 = 0.488,$$

$$E(X \wedge 500) = \frac{2,000}{2} \left[1 - \left(\frac{2,000}{2,000 + 500} \right)^2 \right] = 360.$$

With $E(X) = 1,000$ we have, for the ordinary deductible, the expected cost per loss is $1,000 - 360 = 640$ while the expected cost per payment is $640/0.512 = 1,250$. For the franchise deductible the expectations are $640 + 500(1 - 0.488) = 896$ and $1,250 + 500 = 1,750$. \square

5.2.1 Exercises

5.1 Perform the calculations in Example 5.2 for the following distribution (which is Model 4 on page 15) using an ordinary deductible of 5,000.

$$F_4(x) = \begin{cases} 0, & x < 0, \\ 1 - 0.3e^{-0.00001x}, & x \geq 0. \end{cases}$$

5.2 Repeat Exercise 5.1 for a franchise deductible.

5.3 Repeat Example 5.6 for the model in Exercise 5.1 and a 5,000 deductible.

5.4 (*) Risk 1 has a Pareto distribution with parameters $\alpha > 2$ and θ . Risk 2 has a Pareto distribution with parameters 0.8α and θ . Each risk is covered by a separate policy, each with an ordinary deductible of k . Determine the expected cost per loss for risk 1. Determine the limit as k goes to infinity of

Table 5.1 Data for Exercise 5.5

x	$F(x)$	$E(X \wedge x)$
10,000	0.60	6,000
15,000	0.70	7,700
22,500	0.80	9,500
32,500	0.90	11,000
∞	1.00	20,000

the ratio of the expected cost per loss for risk 2 to the expected cost per loss for risk 1.

5.5 (*) Losses (prior to any deductibles being applied) have a distribution as reflected in Table 5.1. There is a per-loss ordinary deductible of 10,000. The deductible is then raised so that half the number of losses exceed the new deductible as exceeded the old deductible. Determine the percentage change in the expected cost per payment when the deductible is raised.

5.3 THE LOSS ELIMINATION RATIO AND THE EFFECT OF INFLATION FOR ORDINARY DEDUCTIBLES

A ratio that can be meaningful in evaluating the impact of a deductible is the loss elimination ratio.

Definition 5.7 The loss elimination ratio is the ratio of the decrease in the expected payment with an ordinary deductible to the expected payment without the deductible.

While many types of coverage modifications can decrease the expected payment, the term loss elimination ratio is reserved for the effect of changing the deductible. Without the deductible, the expected payment is $E(X)$. With the deductible, the expected payment (from Theorem 5.5) is $E(X) - E(X \wedge d)$. Therefore the loss elimination ratio is

$$\frac{E(X) - [E(X) - E(X \wedge d)]}{E(X)} = \frac{E(X \wedge d)}{E(X)}$$

provided $E(X)$ exists.

Example 5.8 Determine the loss elimination ratio for the Pareto distribution with $\alpha = 3$ and $\theta = 2,000$ with an ordinary deductible of 500.

From Example 5.6 we have a loss elimination ratio of $360/1,000 = 0.36$. Thus 36% of losses can be eliminated by introducing an ordinary deductible of 500. \square

Inflation increases costs, but it turns out that when there is a deductible the effect of inflation is magnified. First, some events that formerly produced losses below the deductible will now lead to payments. Second, the relative effect of inflation is magnified because the deductible is subtracted after inflation. For example, suppose an event formerly produced a loss of 600. With a 500 deductible the payment is 100. Inflation at 10% will increase the loss to 660 and the payment to 160, a 60% increase in the cost to the insurer.

Theorem 5.9 For an ordinary deductible of d after uniform inflation of $1+r$, the expected cost per loss is

$$(1+r)\{E(X) - E[X \wedge d/(1+r)]\}.$$

If $F[d/(1+r)] < 1$, then the expected cost per payment is obtained by dividing by $1 - F[d/(1+r)]$.

Proof: After inflation, losses are given by the random variable $Y = (1+r)X$. From Theorem 4.19, $f_Y(y) = f_X[y/(1+r)]/(1+r)$ and $F_Y(y) = F_X[y/(1+r)]$. Using (3.8),

$$\begin{aligned} E(Y \wedge d) &= \int_0^d y f_Y(y) dy + d[1 - F_Y(d)] \\ &= \int_0^d \frac{y f_X[y/(1+r)]}{1+r} dy + d \left[1 - F_X \left(\frac{d}{1+r} \right) \right] \\ &= \int_0^{d/(1+r)} (1+r) x f_X(x) dx + d \left[1 - F_X \left(\frac{d}{1+r} \right) \right] \\ &= (1+r) \left\{ \int_0^{d/(1+r)} x f_X(x) dx + \frac{d}{1+r} \left[1 - F_X \left(\frac{d}{1+r} \right) \right] \right\} \\ &= (1+r) E \left(X \wedge \frac{d}{1+r} \right) \end{aligned}$$

where the third line follows from the substitution $x = y/(1+r)$. Noting that $E(Y) = (1+r)E(X)$ completes the first statement of the theorem and the per-payment result follows from the relationship between the distribution functions of Y and X . \square

Example 5.10 Determine the effect of inflation at 10% on an ordinary deductible of 500 applied to a Pareto distribution with $\alpha = 3$ and $\theta = 2,000$.

From Example 5.6 the expected costs are 640 and 1,250 per loss and per payment respectively. With 10% inflation we need

$$\begin{aligned} E \left(X \wedge \frac{500}{1.1} \right) &= E(X \wedge 454.55) \\ &= \frac{2,000}{2} \left[1 - \left(\frac{2,000}{2,000 + 454.55} \right)^2 \right] = 336.08. \end{aligned}$$

The expected cost per loss after inflation is $1.1(1,000 - 336.08) = 730.32$, an increase of 14.11%. On a per-payment basis we need

$$\begin{aligned} F_Y(500) &= F_X(454.55) \\ &= 1 - \left(\frac{2,000}{2,000 + 454.55} \right)^3 \\ &= 0.459. \end{aligned}$$

The expected cost per payment is $730.32/(1 - 0.459) = 1,350$, an increase of 8%. \square

5.3.1 Exercises

5.6 Determine the loss elimination ratio for the distribution given below with an ordinary deductible of 5,000. This is the same model used in Exercise 5.1.

$$F_4(x) = \begin{cases} 0, & x < 0, \\ 1 - 0.3e^{-0.00001x}, & x \geq 0. \end{cases}$$

5.7 Determine the effect of inflation at 10% on an ordinary deductible of 5,000 applied to the distribution in Exercise 5.6.

5.8 (*) Losses have a lognormal distribution with $\mu = 7$ and $\sigma = 2$. There is a deductible of 2,000 and 10 losses are expected each year. Determine the loss elimination ratio. If there is uniform inflation of 20% but the deductible remains at 2,000, how many payments will be expected?

5.9 (*) Losses have a Pareto distribution with $\alpha = 2$ and $\theta = k$. There is an ordinary deductible of $2k$. Determine the loss elimination ratio before and after 100% inflation.

5.10 (*) Losses have an exponential distribution with a mean of 1,000. There is a deductible of 500. Determine the amount by which the deductible would have to be raised to double the loss elimination ratio.

5.11 (*) The values in Table 5.2 are available for a random variable X . There is a deductible of 15,000 per loss and no policy limit. Determine the expected cost per payment using X and then assuming 50% inflation (with the deductible remaining at 15,000).

5.12 (*) Losses have a lognormal distribution with $\mu = 6.9078$ and $\sigma = 1.5174$. Determine the ratio of the loss elimination ratio at 10,000 to the loss elimination ratio at 1,000. Then determine the percentage increase in the number of losses that exceed 1,000 if all losses are increased by 10%.

5.13 (*) Losses have a mean of 2,000. With a deductible of 1,000 the loss elimination ratio is 0.3. The probability of a loss being greater than 1,000 is 0.4. Determine the average size of a loss that is less than or equal to 1,000.

Table 5.2 Data for Exercise 5.11

x	$F(x)$	$E(X \wedge x)$
10,000	0.60	6,000
15,000	0.70	7,700
22,500	0.80	9,500
∞	1.00	20,000

5.4 POLICY LIMITS

The opposite of a deductible is a policy limit. The typical policy limit arises in a contract where for losses below u the insurance pays the full loss but for losses above u the insurance pays only u . The effect of the limit is to produce a right censored random variable. It will have a mixed distribution with distribution and density function given by (where Y is the random variable after the limit has been imposed)

$$F_Y(y) = \begin{cases} F_X(y), & y < u, \\ 1, & y \geq u \end{cases}$$

and

$$f_Y(y) = \begin{cases} f_X(y), & y < u, \\ 1 - F_X(u), & y = u. \end{cases}$$

The effect of inflation can be calculated as follows.

Theorem 5.11 For a policy limit of u , after uniform inflation of $1 + r$, the expected cost is $(1 + r)E[X \wedge u/(1 + r)]$.

Proof: The expected cost is $E(Y \wedge u)$. The proof of Theorem 5.9 shows that this equals the expression given in this theorem. \square

For policy limits the concept of per payment and per loss is not relevant. All losses that produced payments prior to imposing the limit will produce payments after the limit is imposed.

Example 5.12 Impose a limit of 3,000 on a Pareto distribution with $\alpha = 3$ and $\theta = 2,000$. Determine the expected cost per loss with the limit as well as the proportional reduction in expected cost. Repeat these calculations after 10% uniform inflation is applied.

For this Pareto distribution, the expected cost is

$$E(X \wedge 3,000) = \frac{2,000}{2} \left[1 - \left(\frac{2,000}{2,000 + 3,000} \right)^2 \right] = 840$$

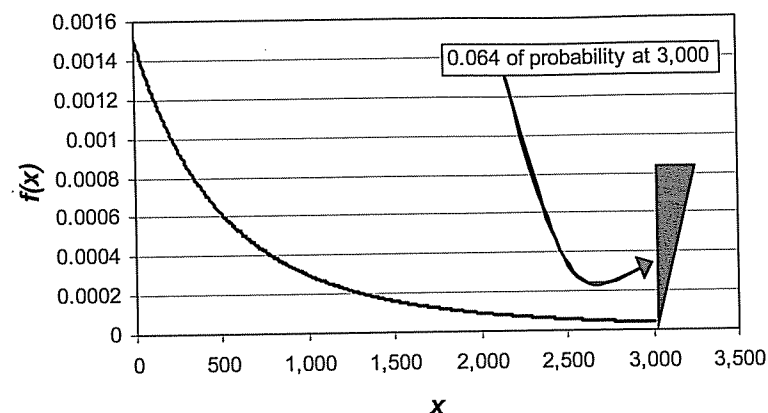


Fig. 5.2 Density function for Example 5.12.

and the proportional reduction is $(1,000 - 840)/1,000 = 0.16$. After inflation the expected cost is

$$1.1E(X \wedge 3,000/1.1) = 1.1 \frac{2,000}{2} \left[1 - \left(\frac{2,000}{2,000 + 3,000/1.1} \right)^2 \right] = 903.11$$

for a proportional reduction of $(1,100 - 903.11)/1,100 = 0.179$. Also note that after inflation the expected cost has increased 7.51%, less than the general inflation rate. The effect is the opposite of the deductible—inflation is tempered, not exacerbated.

Figure 5.2 shows the density function for the right censored random variable. From 0 to 3,000 it matches the original Pareto distribution. The probability of exceeding 3,000, $\Pr(X > 3,000) = (2,000/5,000)^3 = 0.064$ is concentrated at 3,000. \square

A policy limit and an ordinary deductible go together in the sense that, whichever applies to the insurance company's payments, the other applies to the policyholder's payments. For example, when the policy has a deductible of 500, the cost per loss to the policyholder is a random variable that is right censored at 500. When the policy has a limit of 3,000, the policyholder's payments are a variable that is left truncated and shifted (as in an ordinary deductible). The opposite of the franchise deductible is a coverage that right truncates any losses (see Exercise 3.12). This coverage is rarely, if ever, sold. (Would you buy a policy that pays you nothing if your loss exceeds u ?)

5.4.1 Exercises

5.14 Determine the effect of 10% inflation on a policy limit of 150,000 on the following distribution. This is the same distribution used in Exercises 5.1 and 5.6.

$$F_4(x) = \begin{cases} 0, & x < 0 \\ 1 - 0.3e^{-0.00001x}, & x \geq 0. \end{cases}$$

5.15 (*) Let X has a Pareto distribution with $\alpha = 2$ and $\theta = 100$. Determine the range of the mean residual life function $e(d)$ as d ranges over all positive numbers. Then let $Y = 1.1X$. Determine the range of the ratio $e_Y(d)/e_X(d)$ as d ranges over all positive numbers. Finally, let Z be X right censored at 500 (that is, a limit of 500 is applied to X). Determine the range of $e_Z(d)$ as d ranges over the interval 0 to 500.

5.5 COINSURANCE, DEDUCTIBLES, AND LIMITS

The final common coverage modification is coinsurance. In this case the insurance company pays a proportion, α , of the loss and the policyholder pays the remaining fraction. If coinsurance is the only modification, this changes the loss variable X to the payment variable, $Y = \alpha X$. The effect of multiplication has already been covered. When all four items covered in this chapter are present (ordinary deductible, limit, coinsurance, and inflation), we create the following per-loss random variable:

$$Y^L = \begin{cases} 0, & X < \frac{d}{1+r}, \\ \alpha[(1+r)X - d], & \frac{d}{1+r} \leq X < \frac{u}{1+r}, \\ \alpha(u - d), & X \geq \frac{u}{1+r}. \end{cases}$$

For this definition, the quantities are applied in a particular order. In particular, the coinsurance is applied last. For the contract illustrated above, the policy limit is $\alpha(u - d)$, the maximum amount payable. In this definition, u is the loss above which no additional benefits are paid and will be called the **maximum covered loss**. For the per-payment variable, Y^P is undefined for $X < d/(1+r)$.

Previous results can be combined to produce the following theorem, given without proof.

Theorem 5.13 For the per-loss variable,

$$E(Y^L) = \alpha(1+r) \left[E\left(X \wedge \frac{u}{1+r}\right) - E\left(X \wedge \frac{d}{1+r}\right) \right].$$

The expected value of the per-payment variable is obtained as

$$E(Y^P) = \frac{E(Y^L)}{1 - F_X\left(\frac{d}{1+r}\right)}.$$

Higher moments are more difficult. The next theorem gives the formula for the second moment. The variance can then be obtained by subtracting the square of the mean.

Theorem 5.14 For the per-loss variable

$$E[(Y^L)^2] = \alpha^2(1+r)^2 \{E[(X \wedge u^*)^2] - E[(X \wedge d^*)^2] - 2d^*E(X \wedge u^*) + 2d^*E(X \wedge d^*)\},$$

where $u^* = u/(1+r)$ and $d^* = d/(1+r)$. For the second moment of the per-payment variable, divide this expression by $1 - F_X(d^*)$.

Proof: From the definition of Y^L ,

$$\begin{aligned} E[(Y^L)^2] &= \int_{d^*}^{u^*} \alpha^2[(1+r)x - d]^2 f(x) dx + \int_{u^*}^{\infty} \alpha^2(u - d)^2 f(x) dx \\ \frac{E[(Y^L)^2]}{\alpha^2} &= (1+r)^2 \left[\int_0^{u^*} x^2 f(x) dx - \int_0^{d^*} x^2 f(x) dx \right] \\ &\quad - 2(1+r)d \left[\int_0^{u^*} x f(x) dx - \int_0^{d^*} x f(x) dx \right] \\ &\quad + d^2[F(u^*) - F(d^*)] + (u - d)^2[1 - F(u^*)] \\ &= (1+r)^2 \{E[(X \wedge u^*)^2] - u^{*2}[1 - F(u^*)] \\ &\quad - E[(X \wedge d^*)^2] + d^{*2}[1 - F(d^*)]\} \\ &\quad - 2(1+r)^2 d^* \{E(X \wedge u^*) - u^*[1 - F(u^*)] \\ &\quad - E(X \wedge d^*) + d^*[1 - F(d^*)]\} \\ &\quad + (1+r)^2 d^{*2} [F(u^*) - F(d^*)] \\ &\quad + (1+r)^2 (u^* - d^*)^2 [1 - F(u^*)] \\ \frac{E[(Y^L)^2]}{[\alpha(1+r)]^2} &= E[(X \wedge u^*)^2] - E[(X \wedge d^*)^2] - 2d^*[E(X \wedge u^*) - E(X \wedge d^*)]. \end{aligned}$$

□

Example 5.15 Determine the mean and standard deviation per loss for a Pareto distribution with $\alpha = 3$ and $\theta = 2,000$ with a deductible of 500 and a policy limit of 2,500. Note that the maximum covered loss is $u = 3,000$.

From earlier examples, $E(X \wedge 500) = 360$ and $E(X \wedge 3,000) = 840$. The second limited moment is

$$\begin{aligned} E[(X \wedge u)^2] &= \int_0^u x^2 \frac{3(2,000)^3}{(x+2,000)^4} dx + u^2 \left(\frac{2,000}{u+2,000} \right)^3 \\ &= 3(2,000)^3 \int_{2,000}^{u+2,000} (y-2,000)^2 y^{-4} dy + u^2 \left(\frac{2,000}{u+2,000} \right)^3 \\ &= 3(2,000)^3 \left(-y^{-1} + 2,000y^{-2} - \frac{2,000^2}{3}y^{-3} \right) \Big|_{2,000}^{u+2,000} \\ &\quad + u^2 \left(\frac{2,000}{u+2,000} \right)^3 \\ &= 3(2,000)^3 \left[-\frac{1}{u+2,000} + \frac{2,000}{(u+2,000)^2} - \frac{2,000^2}{3(u+2,000)^3} \right] \\ &\quad + 3(2,000)^3 \left[\frac{1}{2,000} - \frac{2,000}{2,000^2} + \frac{2,000^2}{3(2,000)^3} \right] \\ &\quad + u^2 \left(\frac{2,000}{u+2,000} \right)^3 \\ &= (2,000)^2 - \left(\frac{2,000}{u+2,000} \right)^3 (2u+2,000)(u+2,000). \end{aligned}$$

Then, $E[(X \wedge 500)^2] = 160,000$ and $E[(X \wedge 3,000)^2] = 1,440,000$ and so

$$E(Y) = 840 - 360 = 480$$

$$E(Y^2) = 1,440,000 - 160,000 - 2(500)(840) + 2(500)(360) = 800,000$$

for a variance of $800,000 - 480^2 = 569,600$ and a standard deviation of 754.72. \square

5.5.1 Exercises

5.16 (*) You are given that $e(0) = 25$, $S(x) = 1 - x/w$, $0 \leq x \leq w$, and Y is the excess loss variable for $d = 10$. Determine the variance of Y .

5.17 (*) The loss ratio (R) is defined as total losses (L) divided by earned premiums (P). An agent will receive a bonus (B) if the loss ratio on his business is less than 0.7. The bonus is given as $B = P(0.7 - R)/3$ if this quantity is positive, otherwise it is zero. Let $P = 500,000$ and L have a Pareto distribution with parameters $\alpha = 3$ and $\theta = 600,000$. Determine the expected value of the bonus.

5.18 (*) Losses this year have a distribution such that $E(X \wedge d) = -0.025d^2 + 1.475d - 2.25$ for $d = 10, 11, 12, \dots, 26$. Next year, losses will be uniformly

higher by 10%. An insurance policy reimburses 100% of losses subject to a deductible of 11 up to a maximum reimbursement of 11. Determine the ratio of next year's reimbursements to this year's reimbursements.

5.19 (*) Losses have an exponential distribution with a mean of 1,000. An insurance company will pay the amount of each claim in excess of a deductible of 100. Determine the variance of the amount paid by the insurance company for one claim, including the possibility that the amount paid is zero.

5.20 (*) Total claims for a health plan have a Pareto distribution with $\alpha = 2$ and $\theta = 500$. The health plan implements an incentive to physicians that will pay a bonus of 50% of the amount by which total claims are less than 500; otherwise no bonus is paid. It is anticipated that with the incentive plan the claim distribution will change to become Pareto with $\alpha = 2$ and $\theta = K$. With the new distribution it turns out that expected claims plus the expected bonus is equal to expected claims prior to the bonus system. Determine the value of K .

5.21 (*) In year a, total expected losses are 10,000,000. Individual losses in year a have a Pareto distribution with $\alpha = 2$ and $\theta = 2,000$. A reinsurer pays the excess of each individual loss over 3,000. For this, the reinsurer is paid a premium equal to 110% of expected covered losses. In year b, losses will experience 5% inflation over year a, but the frequency of losses will not change. Determine the ratio of the premium in year b to the premium in year a.

5.22 (*) Losses have a uniform distribution from 0 to 50,000. There is a per-loss deductible of 5,000 and a policy limit of 20,000 (meaning that the maximum covered loss is 25,000). Determine the expected payment given that a payment has been made.

5.23 (*) Losses have a lognormal distribution with $\mu = 10$ and $\sigma = 1$. For losses below 50,000, no payment is made. For losses between 50,000 and 100,000, the full amount of the loss is paid. For losses in excess of 100,000 the limit of 100,000 is paid. Determine the expected cost per loss.

5.6 THE IMPACT OF DEDUCTIBLES ON CLAIM FREQUENCY

An important component in analyzing the effect of policy modifications pertains to the change in the frequency distribution of payments when the deductible (ordinary or franchise) is imposed or changed. When a deductible is imposed or increased, there will be fewer payments per period, while if a deductible is lowered, there will be more payments.

We can quantify this process, providing it can be assumed that the imposition of coverage modifications does not affect the process that produces

losses or the type of individual who will purchase insurance. For example, those who buy a 250 deductible on an automobile property damage coverage may (correctly) view themselves as less likely to be involved in an accident than those who buy full coverage. Similarly, an employer may find that the rate of permanent disability declines when reduced benefits are provided to employees in the first few years of employment.

To begin, suppose X_j , the severity, represents the ground-up loss on the j th such loss and there are no coverage modifications. Let N^L denote the number of losses. Now consider a coverage modification such that v is the probability that a loss will result in a payment. For example, if there is a deductible of d , $v = \Pr(X > d)$. Next, define the indicator random variable I_j by $I_j = 1$ if the j th loss results in a payment and $I_j = 0$ otherwise. Then I_j has a Bernoulli distribution with parameter v and the pgf of I_j is $P_{I_j}(z) = 1 - v + vz$. Then $N^P = I_1 + \dots + I_{N^L}$ represents the number of payments. If I_1, I_2, \dots are mutually independent and are also independent of N^L , then N^P has a compound distribution with N^L as the primary distribution and a Bernoulli secondary distribution. Thus

$$P_{N^P}(z) = P_{N^L}[P_{I_j}(z)] = P_{N^L}[1 + v(z - 1)].$$

In the important special case in which the distribution of N^L depends on a parameter θ such that

$$P_{N^L}(z) = P_{N^L}(z; \theta) = B[\theta(z - 1)],$$

where $B(z)$ is functionally independent of θ (as in Theorem 4.54), then

$$\begin{aligned} P_{N^P}(z) &= B[\theta(1 - v + vz - 1)] \\ &= B[v\theta(z - 1)] \\ &= P_{N^L}(z; v\theta). \end{aligned}$$

This implies that N^L and N^P are both from the same parametric family and only the parameter θ need be changed.

Example 5.16 Demonstrate that the above result applies to the negative binomial distribution and illustrate the effect when a deductible of 250 is imposed on a negative binomial distribution with $r = 2$ and $\beta = 3$. Assume that losses have a Pareto distribution with $\alpha = 3$ and $\theta = 1,000$.

The negative binomial pgf is $P_{N^L}(z) = [1 - \beta(z - 1)]^{-r}$. Here β takes on the role of θ in the result and $B(z) = (1 - z)^{-r}$. Then N^P must have a negative binomial distribution with $r^* = r$ and $\beta^* = v\beta$. For the particular situation described,

$$v = 1 - F(250) = \left(\frac{1,000}{1,000 + 250} \right)^3 = 0.512$$

and so $r^* = 2$ and $\beta^* = 3(0.512) = 1.536$. \square

This result may be generalized for zero-modified and zero-truncated distributions. Suppose N^L depends on parameters θ and α such that

$$P_{N^L}(z) = P_{N^L}(z; \theta, \alpha) = \alpha + (1 - \alpha) \frac{B[\theta(z - 1)] - B(-\theta)}{1 - B(-\theta)}. \quad (5.3)$$

Note that $\alpha = P_{N^L}(0) = \Pr(N^L = 0)$ and so is the modified probability at zero. It is also the case that, if $B[\theta(z - 1)]$ is itself a pgf, then the pgf given in (5.3) is that for the corresponding zero-modified distribution. However, it is not necessary for $B[\theta(z - 1)]$ to be a pgf in order for $P_{N^L}(z)$ as given in (5.3) to be a pgf. In particular, $B(z) = 1 + \ln(1 - z)$ yields the zero-modified (ZM) logarithmic distribution, even though there is no distribution with $B(z)$ as its pgf. Similarly, $B(z) = (1 - z)^{-r}$ for $-1 < r < 0$ yields the ETNB distribution. A few algebraic steps reveal that for (5.3)

$$P_{N^P}(z) = P_{N^L}(z; v\theta, \alpha^*),$$

where $\alpha^* = \Pr(N^P = 0) = P_{N^P}(0) = P_{N^L}(1 - v; \theta, \alpha)$. It is expected that imposing a deductible will increase the value of α because periods with no payments will become more likely. In particular, if N^L is zero truncated, N^P will be zero modified.

Example 5.17 Repeat the previous example, only now let the frequency distribution be zero-modified negative binomial with $r = 2$, $\beta = 3$, and $p_0^M = 0.4$.

The pgf is

$$P_{N^L}(z) = p_0^M + (1 - p_0^M) \frac{[1 - \beta(z - 1)]^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}}.$$

Then $\alpha = p_0^M$ and $B(z) = (1 - z)^{-r}$. We then have $r^* = r$, $\beta^* = v\beta$, and

$$\begin{aligned} \alpha^* &= p_0^{M*} = p_0^M + (1 - p_0^M) \frac{[1 + v\beta]^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}} \\ &= \frac{p_0^M - (1 + \beta)^{-r} + (1 + v\beta)^{-r} - p_0^M(1 + v\beta)^{-r}}{1 - (1 + \beta)^{-r}}. \end{aligned}$$

For the particular distribution given, the new parameters are $r^* = 2$, $\beta^* = 3(0.512) = 1.536$, and

$$p_0^{M*} = \frac{0.4 - 4^{-2} + 2.536^{-2} - 0.4(2.536)^{-2}}{1 - 4^{-2}} = 0.4595. \quad \square$$

In applications, it may be the case that we want to determine the distribution of N^L from that of N^P . For example, data may have been collected on the number of payments in the presence of a deductible and from that data

the parameters of N^P can be estimated. We may then want to know the distribution of payments if the deductible is removed. Arguing as before,

$$P_{N^L}(z) = P_{N^P}(1 - v^{-1} + zv^{-1}).$$

This implies that the formulas derived previously hold with v replaced by $1/v$. However, it is possible that the resulting pgf for N^L is not valid. In this case one of the modeling assumptions is invalid (for example, the assumption that changing the deductible does not change claim-related behavior).

Example 5.18 Suppose payments on a policy with a deductible of 250 have the zero-modified negative binomial distribution with $r^* = 2$, $\beta^* = 1.536$, and $p_0^{M*} = 0.4595$. Losses have the Pareto distribution with $\alpha = 3$ and $\theta = 1,000$. Determine the distribution of the number of payments when the deductible is removed. Repeat this calculation assuming $p_0^{M*} = 0.002$.

In this case the formulas use $v = 1/0.512 = 1.953125$ and so $r = 2$ and $\beta = 1.953125(1.536) = 3$. Also,

$$p_0^{M*} = \frac{0.4595 - 2.536^{-2} + 4^{-2} - 0.4595(4)^{-2}}{1 - 2.536^{-2}} = 0.4$$

as expected. For the second case,

$$p_0^{M*} = \frac{0.002 - 2.536^{-2} + 4^{-2} - 0.002(4)^{-2}}{1 - 2.536^{-2}} = -0.1079,$$

which is not a legitimate probability. \square

All members of the $(a, b, 0)$ and $(a, b, 1)$ classes meet the conditions of this section. Table 5.3 indicates how the parameters change when moving from N^L to N^P . If N^L has a compound distribution, then we can write $P_{N^L}(z) = P_1[P_2(z)]$ and therefore

$$P_{N^P}(z) = P_{N^L}[1 + v(z - 1)] = P_1\{P_2[1 + v(z - 1)]\}.$$

Thus N^P will also have a compound distribution with the secondary distribution modified as indicated. If the secondary distribution has an $(a, b, 0)$ distribution, then it can be modified as in Table 5.3. The following example indicates the adjustment to be made if the secondary distribution has an $(a, b, 1)$ distribution.

Example 5.19 Suppose N^L is Poisson-ETNB with $\lambda = 5$, $\beta = 0.3$, and $r = 4$. If $v = 0.5$, determine the distribution of N^P .

From the discussion above, N^P is compound Poisson with $\lambda^* = 5$, but the secondary distribution is a zero-modified negative binomial with (from Table 5.3) $\beta^* = 0.5(0.3) = 0.15$,

$$p_0^{M*} = \frac{0 - 1.3^{-4} + 1.15^{-4} - 0(1.15)^{-4}}{1 - 1.3^{-4}} = 0.34103,$$

Table 5.3 Frequency adjustments

N^L	Parameters for N^P
Poisson	$\lambda^* = v\lambda$
ZM Poisson	$p_0^{M*} = \frac{p_0^M - e^{-\lambda} + e^{-v\lambda} - p_0^M e^{-v\lambda}}{1 - e^{-\lambda}}, \lambda^* = v\lambda$
Binomial	$q^* = vq$
ZM binomial	$p_0^{M*} = \frac{p_0^M - (1 - q)^m + (1 - vq)^m - p_0^M(1 - vq)^m}{1 - (1 - q)^m}$
Negative binomial	$q^* = vq$ $\beta^* = v\beta, r^* = r$
ZM negative binomial	$p_0^{M*} = \frac{p_0^M - (1 + \beta)^{-r} + (1 + v\beta)^{-r} - p_0^M(1 + v\beta)^{-r}}{1 - (1 + \beta)^{-r}}$
ZM logarithmic	$\beta^* = v\beta, r^* = r$ $p_0^{M*} = 1 - (1 - p_0^M) \ln(1 + v\beta) / \ln(1 + \beta)$ $\beta^* = v\beta$

and $r^* = 4$. This would be sufficient, except we have acquired the habit of using the ETNB as the secondary distribution. From Theorem 4.54 a compound Poisson distribution with a zero-modified secondary distribution is equivalent to a compound Poisson distribution with a zero-truncated secondary distribution. The Poisson parameter must be changed to $(1 - p_0^{M*})\lambda^*$. Therefore, N^P has a Poisson-ETNB distribution with $\lambda^* = (1 - 0.34103)5 = 3.29485$, $\beta^* = 0.15$, and $r^* = 4$. \square

The results can be further generalized to an increase or decrease in the deductible. Let N^d be the frequency when the deductible is d and let N^{d^*} be the frequency when the deductible is d^* . Let $v = [1 - F_X(d^*)]/[1 - F_X(d)]$, and then Table 5.3 can be used to move from the parameters of N^d to the parameters of N^{d^*} . As long as $d^* > d$, we will have $v < 1$ and the formulas will lead to a legitimate distribution for N^{d^*} . This includes the special case of $d = 0$ that was used at the start of this section. If $d^* < d$, then $v > 1$ and there is no assurance that a legitimate distribution will result. This includes the special case $d^* = 0$ (removal of a deductible) covered earlier.

Finally, it should be noted that policy limits have no effect on the frequency distribution. Imposing, removing, or changing a limit will not change the number of payments made.

5.6.1 Exercises

5.24 A group life insurance policy has an accidental death rider. For ordinary deaths, the benefit is 10,000; however, for accidental deaths, the benefit is

20,000. The insureds are approximately the same age, so it is reasonable to assume they all have the same claim probabilities. Let them be 0.97 for no claim, 0.01 for an ordinary death claim, and 0.02 for an accidental death claim. A reinsurer has been asked to bid on providing an excess reinsurance that will pay 10,000 for each accidental death.

- (a) The claim process can be modeled with a frequency component that has the Bernoulli distribution (the event is claim/no claim) and a two-point severity component (the probabilities are associated with the two claim levels, given that a claim occurred). Specify the probability distributions for the frequency and severity random variables.
- (b) Suppose the reinsurer wants to retain the same frequency distribution. Determine the modified severity distribution that will reflect the reinsurer's payments.
- (c) Determine the reinsurer's frequency and severity distributions when the severity distribution is to be conditional on a reinsurance payment being made.

5.25 Individual losses have a Pareto distribution with $\alpha = 2$ and $\theta = 1,000$. With a deductible of 500 the frequency distribution for the number of payments is Poisson-inverse Gaussian with $\lambda = 3$ and $\beta = 2$. If the deductible is raised to 1,000, determine the distribution for the number of payments. Also, determine the pdf of the severity distribution (per payment) when the new deductible is in place.

5.26 Losses have a Pareto distribution with $\alpha = 2$ and $\theta = 1,000$. The frequency distribution for a deductible of 500 is zero-truncated logarithmic with $\beta = 4$. Determine a model for the number of payments when the deductible is reduced to 0.

5.27 Suppose that the number of losses N^L has the Sibuya distribution (see Exercises 4.47 and 4.59) with pgf $P_{N^L}(z) = 1 - (1 - z)^{-r}$, where $-1 < r < 0$. Demonstrate that the number of payments has a zero-modified Sibuya distribution.

6

Aggregate loss models

6.1 INTRODUCTION

The purpose of this chapter is to develop models of aggregate losses, the total amount paid on all claims occurring in a fixed time period on a defined set of insurance contracts. There are two ways to go about adding the claims in order to obtain the total for the period.

One method is to record the payments as they are made and then add them up. In that case we can represent the aggregate losses as a sum, S , of a random number, N , of individual payment amounts (X_1, X_2, \dots, X_N) . Hence,

$$S = X_1 + X_2 + \dots + X_N, \quad N = 0, 1, 2, \dots, \quad (6.1)$$

where $S = 0$ when $N = 0$.

Definition 6.1 *The collective risk model has the representation in (6.1) with the X_j s being independent and identically distributed (i.i.d.) random variables, unless otherwise specified. More formally, the independence assumptions are:*

1. *Conditional on $N = n$, the random variables X_1, X_2, \dots, X_n are i.i.d. random variables.*

2. Conditional on $N = n$, the common distribution of the random variables X_1, X_2, \dots, X_n does not depend on n .
3. The distribution of N does not depend in any way on the values of X_1, X_2, \dots .

An alternative model is given below

Definition 6.2 The *individual risk model* represents the aggregate loss as a sum, $S = X_1 + \dots + X_n$, of a fixed number, n , of insurance contracts. The loss amounts for the n contracts are (X_1, X_2, \dots, X_n) , where the X_j s are assumed to be independent but are not assumed to be identically distributed. The distribution of the X_j s usually has a probability mass at zero, corresponding to the probability of no loss or payment.

The individual risk model is used to add together the losses or payments from a fixed number of insurance contracts or sets of insurance contracts. It is used in modeling the losses of a group life or health insurance policy that covers a group of n employees. Each employee can have different coverage (life insurance benefit as a multiple of salary) and different levels of loss probabilities (different ages and health status).

In the special case where the X_j s are identically distributed, the individual risk model becomes the special case of the collective risk model, with the distribution of N being the degenerate distribution with all of the probability at $N = n$; that is, $\Pr(N = n) = 1$.

The distribution of S in (6.1) is obtained from the distribution of N and the distribution of the X_j s. Using this approach, the frequency and the severity of claims are modeled separately. The information about these distributions is used to obtain information about S . An alternative to this approach is to simply gather information about S (e.g., total losses each month for a period of months) and to use some model from Chapter 4 to model the distribution of S . Modeling the distribution of N and the distribution of the X_j s separately has some distinct advantages:

1. The expected number of claims changes as the number of insured policies changes. Growth in the volume of business needs to be accounted for in forecasting the number of claims in future years based on past years' data.
2. The effects of general economic inflation and additional claims inflation are reflected in the losses incurred by insured parties and the claims paid by insurance companies. Such effects are often masked when insurance policies have deductibles and policy limits which do not depend on inflation and aggregate results are used.
3. The impact of changing individual deductibles and policy limits is more easily studied. This is done by changing the specification of the severity distribution.

4. The impact on claims frequencies of changing deductibles is better understood.
5. Data that are heterogeneous in terms of deductibles and limits can be combined to obtain the hypothetical loss size distribution. This is useful when data from several years in which policy provisions were changing are combined.
6. Models developed for noncovered losses to insureds, claim costs to insurers, and claim costs to reinsurers can be mutually consistent. This is useful for a direct insurer which is studying the consequence of shifting losses to a reinsurer.
7. The shape of the distribution of S depends on the shapes of both distributions of N and X . The understanding of the relative shapes is useful when modifying policy details. For example, if the severity distribution has a much heavier tail than the frequency distribution, the shape of the tail of the distribution of aggregate claims or losses will be determined by the severity distribution and will be insensitive to the choice of frequency distribution.

In summary, a more accurate and flexible model can be constructed by examining frequency and severity separately.

In constructing the model (6.1) for S , if N represents the actual number of losses to the insured, then the X_j s can represent (i) the losses to the insured, (ii) the claim payments of the insurer, (iii) the claim payments of a reinsurer, or (iv) the deductibles (self-insurance) paid by the insured. In each case, the interpretation of S is different and the severity distribution can be constructed in a consistent manner.

Because the random variables N, X_1, X_2, \dots , and S provide much of the focus for this chapter and the two that follow, we want to be especially careful when referring to them. We will refer to N as the **claim count random variable** and will refer to its distribution as the **claim count distribution**. The expression **number of claims** will also be used, and, occasionally, just **claims** will be used. Another term that will commonly be used is **frequency distribution**. The X_j s are the **individual or single-loss random variables**. The modifier **individual or single** will be dropped when the reference is clear. In Chapter 5 a distinction was made between losses and payments. Strictly speaking, the X_j s are payments because they represent a real cash transaction. However, the term **loss** is more customary, and we will continue with it. Another common term for the X_j s is **severity**. Finally, S is the **aggregate loss random variable** or the **total loss random variable**.

Example 6.3 An insurer estimates that individual losses to an insured follow a distribution with pdf $f_X(x)$. The insurer pays 80% of individual losses in excess of 1,000 with a maximum payment of 100,000. It reinsures that portion

of any payments in excess of 50,000. Develop models for each of the following: (a) the total loss preinsurance to the policyholder, (b) the aggregate loss to the insurer prior to the reinsurance payment, (c) the aggregate loss to the reinsurer, (d) the aggregate loss to the insurer, after the reinsurance payment, and (e) the aggregate loss to the insured.

- (a) The aggregate losses with no insurance are $S = X_1 + X_2 + \cdots + X_N$, where the X_j s have the distribution with pdf $f_X(x)$.
- (b) The aggregate payments by the insurer (before recovery of reinsurance) are $S = Y_1 + Y_2 + \cdots + Y_N$, where

$$Y_j = \begin{cases} 0, & X_j \leq 1,000, \\ 0.80(X_j - 1,000), & 1,000 < X_j \leq 126,000, \\ 100,000, & X_j > 126,000. \end{cases}$$

- (c) The aggregate payments by the reinsurer are $S = Y_1 + Y_2 + \cdots + Y_N$, where

$$Y_j = \begin{cases} 0, & X_j \leq 63,500, \\ 0.80(X_j - 63,500), & 63,500 < X_j \leq 126,000, \\ 50,000, & X_j > 126,000. \end{cases}$$

- (d) The aggregate costs to the insurer after recovery of reinsurance payments are $S = Y_1 + Y_2 + \cdots + Y_N$, where

$$Y_j = \begin{cases} 0, & X_j \leq 1,000, \\ 0.80(X_j - 1,000), & 1,000 < X_j \leq 63,500, \\ 50,000, & X_j > 63,500. \end{cases}$$

- (e) The aggregate costs to the insured, that is, the uninsured costs, are $S = Y_1 + Y_2 + \cdots + Y_N$, where

$$Y_j = \begin{cases} X_j, & X_j \leq 1,000, \\ 800 + 0.20X_j, & 1,000 < X_j \leq 126,000, \\ X_j - 100,000, & X_j > 126,000. \end{cases} \quad \square$$

6.1.1 Exercise

6.1 Show that the costs to the insured, the insurer, and the reinsurer in Example 6.3 sum to the aggregate loss.

6.2 MODEL CHOICES

In many cases of fitting frequency or severity distributions to data, several distributions may be good candidates for models. However, some distributions may be preferable for a variety of practical reasons.

In general, it is useful for the severity distribution to be a scale distribution (see Definition 4.2) because the choice of currency (e.g., U.S. dollars or British pounds) should not affect the result. Also, scale families are easy to adjust for inflationary effects over time (this is, in effect, a change in currency; e.g., 1994 U.S. dollars to 1995 U.S. dollars). When forecasting the costs for a future year, the anticipated rate of inflation can be factored in easily by adjusting the parameters.

A similar consideration applies to frequency distributions. As a block of an insurance company's business grows, the number of claims can be expected to grow, all other things being equal. If one chooses models that have probability generating functions of the form

$$P_N(z; \alpha) = Q(z)^\alpha \quad (6.2)$$

for some parameter α , then the expected number of claims is proportional to α . Increasing the volume of business by 100% results in expected claims being proportional to $\alpha^* = (1 + r)\alpha$. This was discussed in Section 4.6.11. Because r is any value satisfying $r > -1$, the distributions satisfying (6.2) should allow α to take on any positive values. Such distributions can be shown to be infinitely divisible (see Definition 4.65).

A related consideration also suggests frequency distributions that are infinitely divisible. This relates to the concept of invariance over the time period of study. Ideally the model selected should not depend on the length of the time period used in the study of claims frequency. The expected frequency should be proportional to the length of the time period after any adjustment for growth in business. This means that a study conducted over a period of 10 years can be used to develop claims frequency distributions for periods of a month, a year, or any other period. Furthermore, the form of the distribution for a one-year period is the same as for a one-month period with a change of parameter. The parameter α corresponds to the length of a time period. For example, if $\alpha = 1.7$ in (6.2) for a one-month period, then the identical model with $\alpha = 20.4$ is an appropriate model for a one-year period.

Distributions that have a modification at zero are not of the form (6.2). However, it may still be desirable to use a zero-modified distribution if the physical situation suggests it. For example, if a certain proportion of policies never make a claim, due to duplication of coverage or other reason, it may be appropriate to use this same proportion in future periods for a policy selected at random.

6.2.1 Exercises

6.2 For pgfs satisfying (6.2), show that the mean is proportional to α .

6.3 Which of the distributions in Appendix B satisfy (6.2) for any positive value of α ?

6.3 THE COMPOUND MODEL FOR AGGREGATE CLAIMS

Let S denote aggregate losses associated with a set of N observed claims X_1, X_2, \dots, X_N satisfying the independence assumptions following (6.1). The approach in this chapter is to:

1. Develop a model for the distribution of N based on data.
2. Develop a model for the common distribution of the X_j s based on data.
3. Using these two models, carry out necessary calculations to obtain the distribution of S .

Completion of the first two steps follows the ideas developed elsewhere in this text. We now presume that these two models are developed and that we only need to carry out numerical work in obtaining solutions to problems associated with the distribution of S . These might involve pricing a stop-loss reinsurance contract, and they require analyzing the impact of changes in deductibles, coinsurance levels, and maximum payments on individual losses.

The random sum

$$S = X_1 + X_2 + \dots + X_N$$

(where N has a counting distribution) has a distribution function

$$\begin{aligned} F_S(x) &= \Pr(S \leq x) \\ &= \sum_{n=0}^{\infty} p_n \Pr(S \leq x | N = n) \\ &= \sum_{n=0}^{\infty} p_n F_X^{*n}(x), \end{aligned} \quad (6.3)$$

where $F_X(x) = \Pr(X \leq x)$ is the common distribution function of the X_j s and $p_n = \Pr(N = n)$. In (6.3), $F_X^{*n}(x)$ is the “ n -fold convolution” of the cdf of X . It can be obtained as

$$F_X^{*0}(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases}$$

and

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) dF_X(y) \text{ for } k = 1, 2, \dots \quad (6.4)$$

If X is a continuous random variable with probability zero on negative values, (6.4) reduces to

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x-y) f_X(y) dy \text{ for } k = 2, 3, \dots$$

For $k = 1$ this equation reduces to $F_X^{*1}(x) = F_X(x)$. By differentiating, the pdf is

$$f_X^{*k}(x) = \int_0^x f_X^{*(k-1)}(x-y) f_X(y) dy \text{ for } k = 2, 3, \dots$$

In the case of discrete random variables with positive probabilities at $0, 1, 2, \dots$, Equation (6.4) reduces to

$$F_X^{*k}(x) = \sum_{y=0}^x F_X^{*(k-1)}(x-y) f_X(y) \text{ for } x = 0, 1, \dots, k = 2, 3, \dots$$

The corresponding pf is

$$f_X^{*k}(x) = \sum_{y=0}^x f_X^{*(k-1)}(x-y) f_X(y) \text{ for } x = 0, 1, \dots, k = 2, 3, \dots$$

The distribution (6.3) is called a **compound distribution** and the pf for the distribution of aggregate losses is

$$f_S(x) = \sum_{n=0}^{\infty} p_n f_X^{*n}(x).$$

Arguing as in Section 4.6.7, the pgf of S is

$$\begin{aligned} P_S(z) &= E[z^S] \\ &= \sum_{n=0}^{\infty} E[z^{X_1+X_2+\dots+X_n} | N = n] \Pr(N = n) \\ &= \sum_{n=0}^{\infty} E \left[\prod_{j=1}^n z^{X_j} \right] \Pr(N = n) \\ &= \sum_{n=0}^{\infty} \Pr(N = n) [P_X(z)]^n \\ &= E[P_X(z)^N] = P_N[P_X(z)] \end{aligned} \quad (6.5)$$

due to the independence of X_1, \dots, X_n for fixed n .

A similar relationship exists for the other generating functions. It is sometimes more convenient to use the characteristic function

$$\varphi_S(z) = E(e^{izS}) = P_N[\varphi_X(z)],$$

which always exists. Panjer and Willmot [106] use the Laplace transform

$$L_S(z) = E(e^{-zS}) = P_N[L_X(z)]$$

which always exists for random variables defined on nonnegative values. With regard to the moment generating function, we have

$$M_S(z) = P_N[M_X(z)].$$

The pgf of compound distributions was discussed in Section 4.6.7 where the "secondary" distribution plays the role of the claim size distribution in this chapter.

In the case where $P_N(z) = P_1[P_2(z)]$ —that is, N is itself a compound distribution— $P_S(z) = P_1\{P_2[P_X(z)]\}$, which in itself produces no additional difficulties.

From (6.5), the moments of S can be obtained in terms of the moments of N and the X_j s. The first three moments are

$$\begin{aligned} E(S) &= \mu'_{S1} = \mu'_{N1}\mu'_{X1} = E(N)E(X), \\ \text{Var}(S) &= \mu_{S2} = \mu'_{N1}\mu_{X2} + \mu_{N2}(\mu'_{X1})^2, \\ E\{[S - E(S)]^3\} &= \mu_{S3} = \mu'_{N1}\mu_{X3} + 3\mu_{N2}\mu'_{X1}\mu_{X2} + \mu_{N3}(\mu'_{X1})^3. \end{aligned} \quad (6.6)$$

Here, the first subscript indicates the appropriate random variable, the second subscript indicates the order of the moment, and the superscript is a prime (') for raw moments (moments about the origin) and is unprimed for central moments (moments about the mean). The moments can be used on their own to provide approximations for probabilities of aggregate claims by matching the first few model and sample moments.

Example 6.4 *The observed mean (and standard deviation) of the number of claims and the individual losses over the past 10 months are 6.7 (2.3) and 179,247 (52,141), respectively. Determine the mean and variance of aggregate claims per month.*

$$\begin{aligned} E(S) &= 6.7(179,247) = 1,200,955, \\ \text{Var}(S) &= 6.7(52,141)^2 + (2.3)^2(179,247)^2 \\ &= 1.88180 \times 10^{11}. \end{aligned}$$

Hence, the mean and standard deviation of aggregate claims are 1,200,955 and 433,797, respectively. \square

Example 6.5 (Example 6.4 continued) *Using normal and lognormal distributions as approximating distributions for aggregate claims, calculate the probability that claims will exceed 140% of expected costs. That is,*

$$\Pr(S > 1.40 \times 1,200,955) = \Pr(S > 1,681,337).$$

For the normal distribution

$$\begin{aligned} \Pr(S > 1,681,337) &= \Pr\left(\frac{S - E(S)}{\sqrt{\text{Var}(S)}} > \frac{1,681,337 - 1,200,955}{433,797}\right) \\ &= \Pr(Z > 1.107) = 1 - \Phi(1.107) = 0.134. \end{aligned}$$

For the lognormal distribution, from Appendix A, the mean and second raw moment of the lognormal distribution are

$$E(S) = \exp(\mu + \frac{1}{2}\sigma^2) \quad \text{and} \quad E(S^2) = \exp(2\mu + 2\sigma^2).$$

Equating these to 1.200955×10^6 and $1.88180 \times 10^{11} + (1.200955 \times 10^6)^2 = 1.63047 \times 10^{12}$ and taking logarithms results in the following two equations in two unknowns:

$$\mu + \frac{1}{2}\sigma^2 = 13.99863, \quad 2\mu + 2\sigma^2 = 28.11989.$$

From this $\mu = 13.93731$ and $\sigma^2 = 0.1226361$. Then

$$\begin{aligned} \Pr(S > 1,681,337) &= 1 - \Phi\left[\frac{\ln 1,681,337 - 13.93731}{(0.1226361)^{0.5}}\right] \\ &= 1 - \Phi(1.135913) = 0.128. \end{aligned}$$

The normal distribution provides a good approximation when $E(N)$ is large. In particular, if N has the Poisson, binomial, or negative binomial distribution, a version of the central limit theorem indicates that, as λ , m , or r , respectively, goes to infinity, the distribution of S becomes normal. In this example, $E(N)$ is small so the distribution of S is likely to be skewed. In this case the lognormal distribution may provide a good approximation, although there is no theory to support this choice. \square

Example 6.6 (Group dental insurance) *Under a group dental insurance plan covering employees and their families, the premium for each married employee is the same regardless of the number of family members. The insurance company has compiled statistics showing that the annual cost (adjusted to current dollars) of dental care per person for the benefits provided by the plan has the distribution in Table 6.1 (given in units of 25 dollars).*

Furthermore, the distribution of the number of persons per insurance certificate (that is, per employee) receiving dental care in any year has the distribution given in Table 6.2.

Table 6.1 Loss distribution for Example 6.6

x	$f_X(x)$
1	0.150
2	0.200
3	0.250
4	0.125
5	0.075
6	0.050
7	0.050
8	0.050
9	0.025
10	0.025

Table 6.2 Frequency distribution for Example 6.6

n	p_n
0	0.05
1	0.10
2	0.15
3	0.20
4	0.25
5	0.15
6	0.06
7	0.03
8	0.01

The insurer is now in a position to calculate the distribution of the cost per year per married employee in the group. The cost per married employee is

$$f_S(x) = \sum_{n=0}^8 p_n f_X^{*n}(x).$$

Determine the pf of S up to 525. Determine the mean and standard deviation of total payments per employee.

The distribution up to amounts of 525 is given in Table 6.3. To obtain $f_S(x)$, each row of the matrix of convolutions of $f_X(x)$ is multiplied by the probabilities from the row below the table and the products are summed.

The reader may wish to verify using (6.6) that the first two moments of the distribution $f_S(x)$ are

$$E(S) = 12.58, \quad \text{Var}(S) = 58.7464.$$

Table 6.3 Aggregate probabilities for Example 6.6

x	f_X^{*0}	f_X^{*1}	f_X^{*2}	f_X^{*3}	f_X^{*4}	f_X^{*5}	f_X^{*6}	f_X^{*7}	f_X^{*8}	$f_S(x)$
0	1	0	0	0	0	0	0	0	0	.05000
1	0	.150	0	0	0	0	0	0	0	.01500
2	0	.200	.02250	0	0	0	0	0	0	.02338
3	0	.250	.06000	.00338	0	0	0	0	0	.03468
4	0	.125	.11500	.01350	.00051	0	0	0	0	.03258
5	0	.075	.13750	.03488	.00270	.00008	0	0	0	.03579
6	0	.050	.13500	.06144	.00878	.00051	.00001	0	0	.03981
7	0	.050	.10750	.08569	.01999	.00198	.00009	.00000	0	.04356
8	0	.050	.08813	.09750	.03580	.00549	.00042	.00002	.00000	.04752
9	0	.025	.07875	.09841	.05266	.01194	.00136	.00008	.00000	.04903
10	0	.025	.07063	.09338	.06682	.02138	.00345	.00031	.00002	.05190
11	0	0	.06250	.08813	.07597	.03282	.00726	.00091	.00007	.05138
12	0	0	.04500	.08370	.08068	.04450	.01305	.00218	.00022	.05119
13	0	0	.03125	.07673	.08266	.05486	.02062	.00448	.00060	.05030
14	0	0	.01750	.06689	.08278	.06314	.02930	.00808	.00138	.04818
15	0	0	.01125	.05377	.08081	.06934	.03826	.01304	.00279	.04576
16	0	0	.00750	.04125	.07584	.07361	.04677	.01919	.00505	.04281
17	0	0	.00500	.03052	.06811	.07578	.05438	.02616	.00829	.03938
18	0	0	.00313	.02267	.05854	.07552	.06080	.03352	.01254	.03575
19	0	0	.00125	.01673	.04878	.07263	.06573	.04083	.01768	.03197
20	0	0	.00063	.01186	.03977	.06747	.06882	.04775	.02351	.02832
21	0	0	0	.00800	.03187	.06079	.06982	.05389	.02977	.02479
p_n	.05	.10	.15	.20	.25	.15	.06	.03	.01	

Hence the annual cost of the dental plan has mean $12.58 \times 25 = 314.50$ dollars and standard deviation 191.6155 dollars. (Why can't the calculations be done from Table 6.3?) \square

It is common for insurance to be offered in which a deductible is applied to the aggregate losses for the period. When the losses occur to a policyholder it is called insurance coverage and when the losses occur to an insurance company it is called reinsurance coverage. The latter version is a common method for an insurance company to protect itself against an adverse year (as opposed to protecting against a single, very large claim). More formally, we present the following definition.

Definition 6.7 Insurance on the aggregate losses, subject to a deductible, is called *stop-loss insurance*. The expected cost of this insurance is called the *net stop-loss premium* and can be computed as $E[(S - d)_+]$, where d is the deductible and the notation $(\cdot)_+$ means to use the value in parentheses if it is positive but to use zero otherwise.

For any aggregate distribution,

$$E[(S - d)_+] = \int_d^\infty [1 - F_S(x)] dx.$$

If the distribution is continuous, the net stop-loss premium can be computed directly from the definition as

$$E[(S - d)_+] = \int_d^{\infty} (x - d) f_S(x) dx.$$

Similarly, for discrete random variables,

$$E[(S - d)_+] = \sum_{x>d} (x - d) f_S(x).$$

Any time there is an interval with no aggregate probability, the following result may simplify calculations.

Theorem 6.8 Suppose $\Pr(a < S < b) = 0$. Then, for $a \leq d \leq b$,

$$E[(S - d)_+] = \frac{b-d}{b-a} E[(S - a)_+] + \frac{d-a}{b-a} E[(S - b)_+].$$

That is, the net stop-loss premium can be calculated via linear interpolation.

Proof: From the assumption, $F_S(x) = F_S(a)$, $a \leq x < b$. Then,

$$\begin{aligned} E[(S - d)_+] &= \int_d^{\infty} [1 - F_S(x)] dx \\ &= \int_a^{\infty} [1 - F_S(x)] dx - \int_a^d [1 - F_S(x)] dx \\ &= E[(S - a)_+] - \int_a^d [1 - F_S(a)] dx \\ &= E[(S - a)_+] - (d - a)[1 - F_S(a)]. \end{aligned} \quad (6.7)$$

Then, by setting $d = b$ in (6.7),

$$E[(S - b)_+] = E[(S - a)_+] - (b - a)[1 - F_S(a)]$$

and therefore

$$1 - F_S(a) = \frac{E[(S - a)_+] - E[(S - b)_+]}{b - a}.$$

Substituting this in (6.7) produces the desired result. \square

Further simplification is available in the discrete case, provided S places probability at equally spaced values.

Theorem 6.9 Assume $\Pr(S = kh) = f_k \geq 0$ for some fixed $h > 0$ and $k = 0, 1, \dots$ and $\Pr(S = x) = 0$ for all other x . Then, provided $d = jh$, with j a non-negative integer

$$E[(S - d)_+] = h \sum_{m=0}^{\infty} \{1 - F_S[(m + j)h]\}.$$

Proof:

$$\begin{aligned} E[(S - d)_+] &= \sum_{x>d} (x - d) f_S(x) \\ &= \sum_{k=j}^{\infty} (kh - jh) f_k \\ &= h \sum_{k=j}^{\infty} \sum_{m=0}^{k-j-1} f_k \\ &= h \sum_{m=0}^{\infty} \sum_{k=m+j+1}^{\infty} f_k \\ &= h \sum_{m=0}^{\infty} \{1 - F_S[(m + j)h]\}. \end{aligned} \quad \square$$

In the discrete case with probability at equally spaced values, a simple recursion holds.

Corollary 6.10 Under the conditions of Theorem 6.9,

$$E\{[S - (j + 1)h]_+\} = E[(S - jh)_+] - h[1 - F_S(jh)].$$

This result is easy to use because, when $d = 0$, $E[(S - 0)_+] = E(S) = E(N)E(X)$, which can be obtained directly from the frequency and severity distributions.

Example 6.11 (Example 6.6 continued) *The insurer is examining the effect of imposing an aggregate deductible per employee. Determine the reduction in the net premium as a result of imposing deductibles of 25, 30, 50, and 100 dollars.*

From Table 6.3, the cdf at 0, 25, 50, and 75 dollars has values 0.05, 0.065, 0.08838, and 0.12306. With $E(S) = 25(12.58) = 314.5$ we have

$$\begin{aligned} E[(S - 25)_+] &= 314.5 - 25(1 - 0.05) = 290.75, \\ E[(S - 50)_+] &= 290.75 - 25(1 - 0.065) = 267.375, \\ E[(S - 75)_+] &= 267.375 - 25(1 - 0.08838) = 244.5845, \\ E[(S - 100)_+] &= 244.5845 - 25(1 - 0.12306) = 222.661. \end{aligned}$$

From Theorem 6.8, $E[(S - 30)_+] = \frac{20}{25}290.75 + \frac{5}{25}267.375 = 286.07$. When compared to the original premium of 314.5, the reductions are 23.75, 28.43, 47.125, and 91.839 for the four deductibles. \square

6.3.1 Exercises

6.4 From (6.5), show that the relationships between the moments in (6.6) hold.

6.5 When an individual is admitted to the hospital, the hospital charges have the following characteristics:

1. Charges	Mean	Standard deviation
Room	1,000	500
Other	500	300

2. The covariance between an individual's room charges and other charges is 100,000.

An insurer issues a policy that reimburses 100% for room charges and 80% for other charges. The number of hospital admissions has a Poisson distribution with parameter 4. Determine the mean and standard deviation of the insurer's payout for the policy.

6.6 Aggregate claims have been modeled by a compound negative binomial distribution with parameters $r = 15$ and $\beta = 5$. The claim amounts are uniformly distributed on the interval $(0, 10)$. Using the normal approximation, determine the premium such that the probability that claims will exceed premium is 0.05.

6.7 Automobile drivers can be divided into three homogeneous classes. The number of claims for each driver follows a Poisson distribution with parameter λ . Determine the variance of the number of claims for a randomly selected driver, using the following data.

Class	Proportion of population	λ
1	0.25	5
2	0.25	3
3	0.50	2

6.8 Assume X_1 , X_2 , and X_3 are mutually independent loss random variables with probability functions as given in Table 6.4. Determine the pf of $S = X_1 + X_2 + X_3$.

6.9 Assume X_1 , X_2 , and X_3 are mutually independent random variables with probability functions as given in Table 6.5. If $S = X_1 + X_2 + X_3$ and $f_S(5) = 0.06$, determine p .

Table 6.4 Distributions for Exercise 6.8

x	$f_1(x)$	$f_2(x)$	$f_3(x)$
0	0.90	0.50	0.25
1	0.10	0.30	0.25
2	0.00	0.20	0.25
3	0.00	0.00	0.25

Table 6.5 Distributions for Exercise 6.9

x	$f_1(x)$	$f_2(x)$	$f_3(x)$
0	p	0.6	0.25
1	$1 - p$	0.2	0.25
2	0	0.1	0.25
3	0	0.1	0.25

6.10 Consider the following information about AIDS patients.

1. The conditional distribution of an individual's medical care costs, given that the individual does not have AIDS, has mean 1,000 and variance 250,000.
2. The conditional distribution of an individual's medical care costs, given that the individual does have AIDS, has mean 70,000 and variance 1,600,000.
3. The number of individuals with AIDS in a group of m randomly selected adults has a binomial distribution with parameters m and $q = 0.01$.

An insurance company determines premiums for a group as the mean plus 10% of the standard deviation of the group's aggregate claims distribution. The premium for a group of 10 independent lives for which all individuals have been proven *not* to have AIDS is P . The premium for a group of 10 randomly selected adults is Q . Determine P/Q .

6.11 You have been asked by a city planner to analyze office cigarette smoking patterns. The planner has provided the information in Table 6.6 about the distribution of the number of cigarettes smoked during a workday.

The number of male employees in a randomly selected office of N employees has a binomial distribution with parameters N and 0.4. Determine the mean plus the standard deviation of the number of cigarettes smoked during a workday in a randomly selected office of eight employees.

6.12 For a certain group, aggregate claims are uniformly distributed over $(0, 10)$. Insurer A proposes stop-loss coverage with a deductible of 6 for a

Table 6.6 Data for Exercise 6.11

	Male	Female
Mean	6	3
Variance	64	31

premium equal to the expected stop-loss claims. Insurer B proposes group coverage with a premium of 7 and a dividend (a premium refund) equal to the excess, if any, of $7k$ over claims. Calculate k such that the expected cost to the group is equal under both proposals.

6.13 For a group health contract, aggregate claims are assumed to have an exponential distribution where the mean of the distribution is estimated by the group underwriter. Aggregate stop-loss insurance for total claims in excess of 125% of the expected claims is provided by a gross premium that is twice the expected stop-loss claims. You have discovered an error in the underwriter's method of calculating expected claims. The underwriter's estimate is 90% of the correct estimate. Determine the actual percentage loading in the premium.

6.14 A random loss, X , has the following probability function. You are given that $E(X) = 4$ and $E[(X - d)_+] = 2$. Determine d .

x	$f(x)$
0	0.05
1	0.06
2	0.25
3	0.22
4	0.10
5	0.05
6	0.05
7	0.05
8	0.05
9	0.12

6.15 A reinsurer pays aggregate claim amounts in excess of d , and in return it receives a stop-loss premium $E[(S - d)_+]$. You are given $E[(S - 100)_+] = 15$, $E[(S - 120)_+] = 10$, and the probability that the aggregate claim amounts are greater than 80 and less than or equal to 120 is 0. Determine the probability that the aggregate claim amounts are less than or equal to 80.

6.16 A loss random variable X has pdf $f(x) = \frac{1}{100}$, $0 < x < 100$. Two policies can be purchased to alleviate the financial impact of the loss.

$$A = \begin{cases} 0, & x < 50k, \\ \frac{x}{k} - 50, & x \geq 50k, \end{cases}$$

and

$$B = kx, \quad 0 < x < 100,$$

where A and B are the amounts paid when the loss is x . Both policies have the same net premium, that is, $E(A) = E(B)$. Determine k .

6.17 For a nursing home insurance policy, you are given that the average length of stay is 440 days and 30% of the stays are terminated in the first 30 days. These terminations are distributed uniformly during that period. The policy pays 20 per day for the first 30 days and 100 per day thereafter. Determine the expected benefits payable for a single stay.

6.18 An insurance portfolio produces N claims, where

n	$\Pr(N = n)$
0	0.5
1	0.4
3	0.1

Individual claim amounts have the following distribution:

x	$f_X(x)$
1	0.9
10	0.1

Individual claim amounts and N are mutually independent. Calculate the probability that the ratio of aggregate claims to expected claims will exceed 3.0.

6.19 A company sells group travel-accident life insurance with m payable in the event of a covered individual's death in a travel accident. The gross premium for a group is set equal to the expected value plus the standard deviation of the group's aggregate claims. The standard premium is based on the following assumptions:

1. All individual claims within the group are mutually independent and

2. $m^2q(1-q) = 2,500$, where q is the probability of death by travel accident for an individual.

In a certain group of 100 lives, the independence assumption fails because three specific individuals always travel together. If one dies in an accident, all three are assumed to die. Determine the difference between this group's premium and the standard premium.

6.20 A life insurance company covers 16,000 lives for one-year term life insurance, as shown below.

Benefit amount	Number covered	Probability of claim
1	8,000	0.025
2	3,500	0.025
4	4,500	0.025

All claims are mutually independent. The insurance company's retention limit is 2 units per life. Reinsurance is purchased for 0.03 per unit. The probability that the insurance company's retained claims, S , plus cost of reinsurance will exceed 1,000 is $\Pr \left[\frac{S - E(S)}{\sqrt{\text{Var}(S)}} > K \right]$. Determine K using a normal approximation.

6.21 The probability density function of individual losses Y is

$$f(y) = \begin{cases} 0.02 \left(1 - \frac{y}{100} \right), & 0 < y < 100, \\ 0, & \text{elsewhere.} \end{cases}$$

The amount paid, Z , is 80% of that portion of the loss that exceeds a deductible of 10. Determine $E(Z)$.

6.22 An individual loss distribution is normal with $\mu = 100$ and $\sigma^2 = 9$. The distribution for the number of claims, N , is given in Table 6.7. Determine the probability that aggregate claims exceed 100.

Table 6.7 Distribution for Exercise 6.22

n	$\Pr(N = n)$
0	0.5
1	0.2
2	0.2
3	0.1

Table 6.8 Distribution for Exercise 6.23

n	$f(n)$
0	1/16
1	1/4
2	3/8
3	1/4
4	1/16

6.23 An employer self-insures a life insurance program with the following characteristics:

- Given that a claim has occurred, the claim amount is 2,000 with probability 0.4 and 3,000 with probability 0.6,
- The number of claims has the distribution given in Table 6.8.

The employer purchases aggregate stop-loss coverage that limits the employer's annual claims cost to 5,000. The aggregate stop-loss coverage costs 1,472. Determine the employer's expected annual cost of the program, including the cost of stop-loss coverage.

6.24 The probability that an individual admitted to the hospital will stay k days or less is $1 - 0.8^k$ for $k = 0, 1, 2, \dots$. A hospital indemnity policy provides a fixed amount per day for the 4th day through the 10th day (that is, for a maximum of 7 days). Determine the percentage increase in the expected cost per admission if the maximum number of days paid is increased from 7 to 14.

6.25 The probability density function of aggregate claims, S , is given by $f_S(x) = 3x^{-4}$, $x \geq 1$. The relative loading θ and the value λ are selected so that

$$\Pr[S \leq (1 + \theta)E(S)] = \Pr[S \leq E(S) + \lambda\sqrt{\text{Var}(S)}] = 0.90.$$

Calculate λ and θ .

X 6.4 ANALYTIC RESULTS

For most choices of distributions of N and the X_j s, the compound distributional values can only be obtained numerically. Subsequent sections in this chapter are devoted to such numerical procedures.

However, for certain combinations of choices, simple analytic results are available, thus reducing the computational problems considerably.

Example 6.12 (Compound geometric-exponential) Suppose X_1, X_2, \dots are i.i.d. with common exponential distribution with mean θ and mgf $M_X(z) = (1 - \theta z)^{-1}$. Suppose that N has a geometric distribution with parameter β and pgf $P_N(z) = [1 - \beta(z - 1)]^{-1}$ (see Appendix B). Determine the distribution of S .

The mgf of S is

$$\begin{aligned} M_S(z) &= P_N[M_X(z)] \\ &= \{1 - \beta[(1 - \theta z)^{-1} - 1]\}^{-1} \\ &= \frac{1}{1 + \beta} + \frac{\beta}{1 + \beta} [1 - \theta(1 + \beta)z]^{-1} \end{aligned}$$

with a bit of algebra.

This is a two-point mixture of a degenerate distribution with probability 1 at zero and an exponential distribution with mean $\theta(1 + \beta)$. Hence, $\Pr(S = 0) = (1 + \beta)^{-1}$, and for $x > 0$, S has pdf

$$f_S(x) = \frac{\beta}{\theta(1 + \beta)^2} \exp\left[-\frac{x}{\theta(1 + \beta)}\right].$$

It has a point mass of $(1 + \beta)^{-1}$ at zero and an exponentially decaying density over the positive axis. Its cdf can be written as

$$F_S(x) = 1 - \frac{\beta}{1 + \beta} \exp\left[-\frac{x}{\theta(1 + \beta)}\right], \quad x \geq 0.$$

It has a jump at zero and is continuous otherwise. This example will arise again in Chapter 8 in connection with ruin theory. \square

Example 6.13 (Exponential severities) Determine the cdf of S for any compound distribution with exponential severities.

The mgf of the sum of n independent exponential random variables each with mean θ is

$$M_{X_1 + X_2 + \dots + X_n}(z) = (1 - \theta z)^{-n},$$

which is the mgf of the gamma distribution with cdf

$$F_X^{*n}(x) = \Gamma\left(n; \frac{x}{\theta}\right)$$

(see Appendix A). For integer values of α , the values of $\Gamma(\alpha; x)$ can be calculated exactly (see Appendix A for the derivation) as

$$\Gamma(n; x) = 1 - \sum_{j=0}^{n-1} \frac{x^j e^{-x}}{j!}, \quad n = 1, 2, 3, \dots \quad (6.8)$$

From (6.3)

$$F_S(x) = p_0 + \sum_{n=1}^{\infty} p_n \Gamma\left(n; \frac{x}{\theta}\right).$$

Substituting in (6.8) yields

$$F_S(x) = 1 - \sum_{n=1}^{\infty} p_n \sum_{j=0}^{n-1} \frac{(x/\theta)^j e^{-x/\theta}}{j!}, \quad x \geq 0. \quad (6.9)$$

Interchanging the order of summation yields

$$\begin{aligned} F_S(x) &= 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \frac{(x/\theta)^j}{j!} \sum_{n=j+1}^{\infty} p_n \\ &= 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \bar{P}_j \frac{(x/\theta)^j}{j!}, \quad x \geq 0, \end{aligned}$$

where $\bar{P}_j = \sum_{n=j+1}^{\infty} p_n$ for $j = 0, 1, \dots$ \square

The approach of Example 6.13 may be extended to the larger class of mixed Erlang severity distributions, as shown in Exercise 6.35.

For frequency distributions which assign positive probability to all nonnegative integers, (6.9) can be evaluated by taking sufficient terms in the first summation. For distributions for which $\Pr(N > n^*) = 0$, the first summation becomes finite. For example, for the binomial frequency distribution, (6.9) becomes

$$F_S(x) = 1 - \sum_{n=1}^m \binom{m}{n} q^n (1 - q)^{m-n} \sum_{j=0}^{n-1} \frac{(x/\theta)^j e^{-x/\theta}}{j!}. \quad (6.10)$$

Example 6.14 (Compound negative binomial-exponential) Determine the distribution of S when the frequency distribution is negative binomial with an integer value for the parameter r and the severity distribution is exponential.

The mgf of S is

$$\begin{aligned} M_S(z) &= P_N[M_X(z)] \\ &= P_N[(1 - \theta z)^{-1}] \\ &= \{1 - \beta[(1 - \theta z)^{-1} - 1]\}^{-r}. \end{aligned}$$

With a bit of algebra, this can be rewritten as

$$M_S(z) = \left(1 + \frac{\beta}{1 + \beta} \{[1 - \theta(1 + \beta)z]^{-1} - 1\}\right)^r$$

which is of the form

$$M_S(z) = P_N^*[M_X^*(z)]$$

where

$$P_N^*(z) = \left[1 + \frac{\beta}{1+\beta}(z-1)\right]^r,$$

the pgf of the binomial distribution with parameters r and $\beta/(1+\beta)$, and $M_X^*(z)$ is the mgf of the exponential distribution with mean $\theta(1+\beta)$.

This transformation reduces the computation of the distribution function to the finite sum of the form of (6.10), that is,

$$F_S(x) = 1 - \sum_{n=1}^r \binom{r}{n} \left(\frac{\beta}{1+\beta}\right)^n \left(\frac{1}{1+\beta}\right)^{r-n} \times \sum_{j=0}^{n-1} \frac{[x\theta^{-1}(1+\beta)^{-1}]^j e^{-x\theta^{-1}(1+\beta)^{-1}}}{j!}.$$

□

Example 6.15 (Severity distributions closed under convolution) A distribution is said to be *closed under convolution* if adding i.i.d. members of a family produces another member of that family. Further assume that adding n members of a family produces a member with all but one parameter unchanged and the remaining parameter is multiplied by n . Determine the distribution of S when the severity distribution has this property.

The condition means that, if $f_X(x; a)$ is the pf of each X_j , then the pf of $X_1 + X_2 + \cdots + X_n$ is $f_X(x; na)$. This means that

$$\begin{aligned} f_S(x) &= \sum_{n=1}^{\infty} p_n f_X^{*n}(x; a) \\ &= \sum_{n=1}^{\infty} p_n f_X(x; na), \end{aligned}$$

eliminating the need to carry out evaluation of the convolution. Severity distributions that are closed under convolution include the gamma and inverse Gaussian distributions. See Exercise 6.26. □

6.4.1 Exercises

6.26 The following questions concern closure under convolution.

- (a) Show that the gamma and inverse Gaussian distributions are closed under convolution. Show that the gamma distribution has the additional property mentioned in Example 6.15.

- (b) Discrete distributions can also be used as severity distributions. Which of the distributions in Appendix B are closed under convolution? How can this information be used in simplifying calculation of compound probabilities of the form (4.20)?

6.27 A compound negative binomial distribution has parameters $\beta = 1$, $r = 2$, and severity distribution $\{f_X(x); x = 0, 1, 2, \dots\}$. How do the parameters of the distribution change if the severity distribution is $\{g_X(x) = f_X(x)/[1 - f_X(0)]; x = 1, 2, \dots\}$ but the aggregate claims distribution remains unchanged?

6.28 Consider the compound logarithmic distribution with exponential severity distribution.

- (a) Show that the density of aggregate losses may be expressed as

$$f_S(x) = \frac{1}{\ln(1+\beta)} \sum_{n=1}^{\infty} \frac{1}{n!} \left[\frac{\beta}{\theta(1+\beta)} \right]^n x^{n-1} e^{-x/\theta}.$$

- (b) Reduce this to

$$f_S(x) = \frac{\exp\{-x/[\theta(1+\beta)]\} - \exp(-x/\theta)}{x \ln(1+\beta)}.$$

6.29 An insurance policy reimburses aggregate incurred expenses at the rate of 80% of the first 1,000 in excess of 100, 90% of the next 1,000, and 100% thereafter. Express the expected cost of this coverage in terms of $R_d = E[(S - d)_+]$ for different values of d .

6.30 The number of accidents incurred by an insured driver in a single year has a Poisson distribution with parameter $\lambda = 2$. If an accident occurs, the probability that the damage amount exceeds the deductible is 0.25. The number of claims and the damage amounts are independent. What is the probability that there will be no damages exceeding the deductible in a single year?

6.31 The aggregate loss distribution is modeled by an insurance company using an exponential distribution. However, the mean is uncertain. The company uses a uniform distribution (2,000,000, 4,000,000) to express its view of what the mean should be. Determine the expected aggregate losses.

6.32 A group hospital indemnity policy provides benefits at a continuous rate of 100 per day of hospital confinement for a maximum of 30 days. Benefits for partial days of confinement are prorated. The length of hospital confinement

in days, T , has the following continuance (survival) function for $0 \leq t \leq 30$:

$$\Pr(T \geq t) = \begin{cases} 1 - 0.04t, & 0 \leq t \leq 10, \\ 0.95 - 0.035t, & 10 < t \leq 20, \\ 0.65 - 0.02t, & 20 < t \leq 30. \end{cases}$$

For a policy period, each member's probability of a single hospital admission is 0.1 and of more than one admission is 0. Determine the pure premium per member, ignoring the time value of money.

6.33 Medical and dental claims are assumed to be independent with compound Poisson distributions as follows:

Claim type	Claim amount distribution	λ
Medical claims	Uniform (0, 1,000)	2
Dental claims	Uniform (0, 200)	3

Let X be the amount of a given claim under a policy which covers both medical and dental claims. Determine $E[(X - 100)_+]$, the expected cost (in excess of 100) of any given claim.

6.34 For a certain insured, the distribution of aggregate claims is binomial with parameters $m = 12$ and $q = 0.25$. The insurer will pay a dividend, D , equal to the excess of 80% of the premium over claims, if positive. The premium is 5. Determine $E[D]$.

6.35 Consider a severity distribution which is a finite mixture of gamma distributions with integer shape parameters (such gamma distributions are called Erlang distributions), that is, which may be expressed as

$$f_X(x) = \sum_{k=1}^r q_k \frac{\theta^{-k} x^{k-1} e^{-x/\theta}}{(k-1)!}, \quad x > 0.$$

(a) Show that the moment generating function may be written as

$$M_X(z) = Q\{(1 - \theta z)^{-1}\},$$

where

$$Q(z) = \sum_{k=1}^r q_k z^k$$

is the pgf of the distribution $\{q_1, q_2, \dots, q_r\}$. Thus interpret $f_X(x)$ as the pf of a compound distribution.

(b) Show that the mgf of S is

$$M_S(z) = C\{(1 - \theta z)^{-1}\},$$

where

$$C(z) = \sum_{k=0}^{\infty} c_k z^k = P_N\{Q(z)\}.$$

(c) Describe how the distribution $\{c_k; k = 0, 1, 2, \dots\}$ may be calculated recursively if the number of claims distribution is a member of the $(a, b, 1)$ class (Section 4.6.6).

(d) Show that the distribution function of S is given by

$$\begin{aligned} F_S(x) &= 1 - \sum_{n=1}^{\infty} c_n \sum_{j=0}^{n-1} \frac{(x/\theta)^j e^{-x/\theta}}{j!} \\ &= 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \bar{C}_j \frac{(x/\theta)^j}{j!}, \quad x \geq 0, \end{aligned}$$

$$\text{where } \bar{C}_j = \sum_{n=j+1}^{\infty} c_n.$$

6.5 COMPUTING THE AGGREGATE CLAIMS DISTRIBUTION

The computation of the compound distribution function

$$F_S(x) = \sum_{n=0}^{\infty} p_n F_X^{*n}(x) \quad (6.11)$$

or the corresponding probability (density) function is generally not an easy task, even in the simplest of cases. In this section we discuss a number of approaches to numerical evaluation of (6.11) for specific choices of the frequency and severity distributions as well as for arbitrary choices of one or both distributions.

One approach is to use an **approximating distribution** to avoid direct calculation of (6.11). This approach was used in Example 6.5 where the method of moments was used to estimate the parameters of the approximating distribution. The advantage of this method is that it is simple and easy to apply. However, the disadvantages are significant. First, there is no way of knowing how good the approximation is. Choosing different approximating distributions can result in very different results, particularly in the right-hand tail of the distribution. Of course, the approximation should improve as more moments are used; but after four moments, one quickly runs out of distributions!

The approximating distribution may also fail to accommodate special features of the true distribution. For example, when the loss distribution is of

the continuous type and there is a maximum possible claim (for example, when there is a policy limit), the severity distribution may have a point mass ("atom" or "spike") at the maximum. The true aggregate claims distribution is of the mixed type with spikes at integral multiples of the maximum corresponding to 1, 2, 3, ... claims at the maximum. These spikes, if large, can have a significant effect on the probabilities near such multiples. These jumps in the aggregate claims distribution function cannot be replicated by a smooth approximating distribution.

The second method to evaluate (6.11) or the corresponding pdf is **direct calculation**. The most difficult (or computer intensive) part is the evaluation of the n -fold convolutions of the severity distribution for $n = 2, 3, 4, \dots$. In some situations, there is an analytic form—for example, when the severity distribution is closed under convolution, as defined in Example 6.15 and illustrated in Examples 6.12–6.14. Otherwise the convolutions need to be evaluated numerically using

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) dF_X(y). \quad (6.12)$$

When the losses are limited to nonnegative values (as is usually the case), the range of integration becomes finite, reducing (6.12) to

$$F_X^{*k}(x) = \int_{0-x}^x F_X^{*(k-1)}(x-y) dF_X(y). \quad (6.13)$$

These integrals are written in Lebesgue–Stieltjes form because of possible jumps in the cdf $F_X(x)$ at zero and at other points.¹ Numerical evaluation of (6.13) requires numerical integration methods. Because of the first term inside the integral, (6.13) needs to be evaluated for all possible values of x . This quickly becomes technically overpowering!

A simple way to avoid these technical problems is to replace the severity distribution by a discrete distribution defined at multiples 0, 1, 2, ... of some convenient monetary unit such as 1,000. This reduces (6.13) to (in terms of the new monetary unit)

$$F_X^{*k}(x) = \sum_{y=0}^x F_X^{*(k-1)}(x-y) f_X(y).$$

The corresponding pf is

$$f_X^{*k}(x) = \sum_{y=0}^x f_X^{*(k-1)}(x-y) f_X(y).$$

¹Without going into the formal definition of the Lebesgue–Stieltjes integral, it suffices to interpret $\int g(y) dF_X(y)$ as to be evaluated by integrating $g(y)f_X(y)$ over those y values for which X has a continuous distribution and then adding $g(y_i)\Pr(X=y_i)$ over those points where $\Pr(X=y_i) > 0$. This allows for a single notation to be used for continuous, discrete, and mixed random variables.

In practice, the monetary unit can be made sufficiently small to accommodate spikes at maximum insurance amounts. One needs only the spike to be a multiple of the monetary unit to have it located at exactly the right point. As the monetary unit of measurement becomes small, the discrete distribution function needs to approach the true distribution function. The simplest approach is to round all amounts to the nearest multiple of the monetary unit; for example, round all losses or claims to the nearest 1,000. More sophisticated methods will be discussed later in this chapter.

When the severity distribution is defined on nonnegative integers 0, 1, 2, ..., calculating $f_X^{*k}(x)$ for integral x requires $x+1$ multiplications. Then carrying out these calculations for all possible values of k and x up to n requires a number of multiplications that are of order n^3 , written as $O(n^3)$, to obtain the distribution of (6.11) for $x = 0$ to $x = n$. When the maximum value, n , for which the aggregate claims distribution is calculated is large, the number of computations quickly becomes prohibitive, even for fast computers. For example, in real applications n can easily be as large as 1,000. This requires about 10^9 multiplications. Further, if $\Pr(X=0) > 0$, an infinite number of calculations are required to obtain any single probability. This is because $F_X^{*n}(x) > 0$ for all n and all x and so the sum in (6.11) contains an infinite number of terms. When $\Pr(X=0) = 0$, we have $F_X^{*n}(x) = 0$ for $n > x$ and so (6.11) will have no more than $x+1$ positive terms. Table 6.3 provides an example of this latter case.

Alternative methods to more quickly evaluate the aggregate claims distribution are discussed in the next two sections. The first such method, the **recursive method**, reduces the number of computations discussed above to $O(n^2)$, which is a considerable savings in computer time, a reduction of about 99.9% when $n = 1,000$ compared to direct calculation. However, the method is limited to certain frequency distributions. Fortunately, it includes all frequency distributions discussed in Section 4.6 and Appendix B.

The second such method, the **inversion method**, numerically inverts a transform, such as the characteristic function, using a general or specialized inversion software package. Two versions of this method are discussed in this chapter.

6.6 THE RECURSIVE METHOD

Suppose that the severity distribution $f_X(x)$ is defined on 0, 1, 2, ..., m representing multiples of some convenient monetary unit. The number m represents the largest possible payment and could be infinite. Further, suppose that the frequency distribution, p_k , is a member of the $(a, b, 1)$ class and therefore satisfies

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}, \quad k = 2, 3, 4, \dots$$

Then the following result holds.

Theorem 6.16 For the $(a, b, 1)$ class,

$$f_S(x) = \frac{[p_1 - (a+b)p_0]f_X(x) + \sum_{y=1}^{x \wedge m} (a + by/x)f_X(y)f_S(x-y)}{1 - af_X(0)}, \quad (6.14)$$

noting that $x \wedge m$ is notation for $\min(x, m)$.

Proof: This result is identical to Theorem 4.49 with appropriate substitution of notation and recognition that the argument of $f_X(x)$ cannot exceed m . \square

Corollary 6.17 For the $(a, b, 0)$ class, the result (6.14) reduces to

$$f_S(x) = \frac{\sum_{y=1}^{x \wedge m} (a + by/x) f_X(y) f_S(x-y)}{1 - af_X(0)}. \quad (6.15)$$

Note that when the severity distribution has no probability at zero, the denominator of (6.14) and (6.15) equals 1. Further, in the case of the Poisson distribution, (6.15) reduces to

$$f_S(x) = \frac{\lambda}{x} \sum_{y=1}^{x \wedge m} y f_X(y) f_S(x-y), \quad x = 1, 2, \dots \quad (6.16)$$

The starting value of the recursive schemes (6.14) and (6.15) is $f_S(0) = P_N[f_X(0)]$ following Theorem 4.51 with an appropriate change of notation. In the case of the Poisson distribution we have

$$f_S(0) = e^{-\lambda[1-f_X(0)]}.$$

Starting values for other frequency distributions are found in Appendix D.

6.6.1 Applications to compound frequency models

When the frequency distribution can be represented as a compound distribution (e.g., Neyman Type A, Poisson-inverse Gaussian) involving only distributions from the $(a, b, 0)$ or $(a, b, 1)$ classes, the recursive formula (6.14) can be used two or more times to obtain the aggregate claims distribution. If the frequency distribution can be written as

$$P_N(z) = P_1[P_2(z)],$$

then the aggregate claims distribution has pgf

$$\begin{aligned} P_S(z) &= P_N[P_X(z)] \\ &= P_1\{P_2[P_X(z)]\}, \end{aligned}$$

which can be rewritten as

$$P_S(z) = P_1[P_{S_1}(z)] \quad (6.17)$$

where

$$P_{S_1}(z) = P_2[P_X(z)]. \quad (6.18)$$

Now (6.18) is the same form as an aggregate claims distribution. Thus, if $P_2(z)$ is in the $(a, b, 0)$ or $(a, b, 1)$ class, the distribution of S_1 can be calculated using (6.14). The resulting distribution is the "severity" distribution in (6.18). Thus, a second application of (6.14) to (6.17) results in the distribution of S .

The following example illustrates the use of this algorithm.

Example 6.18 The number of claims has a Poisson-ETNB distribution with Poisson parameter $\lambda = 2$ and ETNB parameters $\beta = 3$ and $r = 0.2$. The claim size distribution has probabilities 0.3, 0.5, and 0.2 at 0, 10, and 20, respectively. Determine the total claims distribution recursively.

In the above terminology, N has pgf $P_N(z) = P_1[P_2(z)]$, where $P_1(z)$ and $P_2(z)$ are the Poisson and ETNB pgfs, respectively. Then the total dollars of claims has pgf $P_S(z) = P_1[P_{S_1}(z)]$, where $P_{S_1}(z) = P_2[P_X(z)]$ is a compound ETNB pgf. We will first compute the distribution of S_1 . We have (in monetary units of 10) $f_X(0) = 0.3$, $f_X(1) = 0.5$, and $f_X(2) = 0.2$. In order to use the compound ETNB recursion, we start with

$$\begin{aligned} f_{S_1}(0) &= P_2[f_X(0)] \\ &= \frac{\{1 + \beta[1 - f_X(0)]\}^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}} \\ &= \frac{\{1 + 3(1 - 0.3)\}^{-0.2} - (1 + 3)^{-0.2}}{1 - (1 + 3)^{-0.2}} \\ &= 0.16369. \end{aligned}$$

The remaining values of $f_{S_1}(x)$ may be obtained from (6.14) with S replaced by S_1 . In this case we have $a = 3/(1+3) = 0.75$, $b = (0.2-1)a = -0.6$, $p_0 = 0$ and $p_1 = (0.2)(3)/[(1+3)^{0.2+1} - (1+3)] = 0.46947$. Then (6.14) becomes

$$\begin{aligned} f_{S_1}(x) &= \frac{[0.46947 - (0.75 - 0.6)(0)]f_X(x) + \sum_{y=1}^x (0.75 - 0.6y/x)f_X(y)f_{S_1}(x-y)}{1 - (0.75)(0.3)} \\ &= 0.60577f_X(x) + 1.29032 \sum_{y=1}^x \left(0.75 - 0.6\frac{y}{x}\right) f_X(y)f_{S_1}(x-y). \end{aligned}$$

The first few probabilities are

$$\begin{aligned}
 f_{S_1}(1) &= 0.60577(0.5) + 1.29032 \left[0.75 - 0.6 \left(\frac{1}{1} \right) \right] (0.5)(0.16369) \\
 &= 0.31873, \\
 f_{S_1}(2) &= 0.60577(0.2) + 1.29032 \left\{ \left[0.75 - 0.6 \left(\frac{1}{2} \right) \right] (0.5)(0.31873) \right. \\
 &\quad \left. + \left[0.75 - 0.6 \left(\frac{2}{2} \right) \right] (0.2)(0.16369) \right\} = 0.22002, \\
 f_{S_1}(3) &= 1.29032 \left\{ \left[0.75 - 0.6 \left(\frac{1}{3} \right) \right] (0.5)(0.22002) \right. \\
 &\quad \left. + \left[0.75 - 0.6 \left(\frac{2}{3} \right) \right] (0.2)(0.31873) \right\} = 0.10686, \\
 f_{S_1}(4) &= 1.29032 \left\{ \left[0.75 - 0.6 \left(\frac{1}{4} \right) \right] (0.5)(0.10686) \right. \\
 &\quad \left. + \left[0.75 - 0.6 \left(\frac{2}{4} \right) \right] (0.2)(0.22002) \right\} = 0.06692.
 \end{aligned}$$

We now turn to evaluation of the distribution of S with compound Poisson pgf

$$P_S(z) = P_1[P_{S_1}(z)] = e^{\lambda[P_{S_1}(z)-1]}.$$

Thus the distribution

$$\{f_{S_1}(x); x = 0, 1, 2, \dots\}$$

becomes the "secondary" or "claim size" distribution in an application of the compound Poisson recursive formula. Therefore,

$$f_S(0) = P_S(0) = e^{\lambda[P_{S_1}(0)-1]} = e^{\lambda[f_{S_1}(0)-1]} = e^{2(0.16369-1)} = 0.18775.$$

The remaining probabilities may be found from the recursive formula

$$f_S(x) = \frac{2}{x} \sum_{y=1}^x y f_{S_1}(y) f_S(x-y), \quad x = 1, 2, \dots$$

The first few probabilities are

$$\begin{aligned}
 f_S(1) &= 2 \left(\frac{1}{1} \right) (0.31873)(0.18775) = 0.11968, \\
 f_S(2) &= 2 \left(\frac{1}{2} \right) (0.31873)(0.11968) + 2 \left(\frac{2}{2} \right) (0.22002)(0.18775) = 0.12076, \\
 f_S(3) &= 2 \left(\frac{1}{3} \right) (0.31873)(0.12076) + 2 \left(\frac{2}{3} \right) (0.22002)(0.11968) \\
 &\quad + 2 \left(\frac{3}{3} \right) (0.10686)(0.18775) = 0.10090, \\
 f_S(4) &= 2 \left(\frac{1}{4} \right) (0.31873)(0.10090) + 2 \left(\frac{2}{4} \right) (0.22002)(0.12076) \\
 &\quad + 2 \left(\frac{3}{4} \right) (0.10686)(0.11968) + 2 \left(\frac{4}{4} \right) (0.06692)(0.18775) \\
 &= 0.08696.
 \end{aligned}$$

This simple idea can be extended to higher levels of compounding by repeatedly applying the same concepts. The computer time required to carry out two applications will be about twice that of one application of (6.14).

However, the total number of computations is still of order $O(x^2)$ rather than $O(x^3)$ as in the direct method.

When the severity distribution has a maximum possible value at m , the computations are speeded up even more because the sum in (6.14) will be restricted to at most m nonzero terms. In this case, then, the computations can be considered to be of order $O(x)$.

6.6.2 Underflow/overflow problems

The recursion (6.14) starts with the calculated value of $P(S=0) = P_N[f_X(0)]$. For large insurance portfolios, this probability is very small, sometimes smaller than the smallest number that can be represented on the computer. When this occurs, this initial value is represented on the computer as zero and the recursion (6.14) fails. This problem can be overcome in several different ways (see Panjer and Willmot [105]). One of the easiest ways is to start with an arbitrary set of values for $f_S(0), f_S(1), \dots, f_S(k)$ such as $(0, 0, 0, \dots, 0, 1)$, where k is sufficiently far to the left in the distribution so that $F_S(k)$ is still negligible. Setting k to a point that lies six standard deviations to the left of the mean is usually sufficient. Recursion (6.14) is used to generate values of the distribution with this set of starting values until the values are consistently less than $f_S(k)$. The "probabilities" are then summed and divided by the sum so that the "true" probabilities add to 1. Trial and error will dictate how small k should be for a particular problem.

Another method to obtain probabilities when the starting value is too small is to carry out the calculations for a subset of the portfolio. For example, for the Poisson distribution with mean λ , find a value of $\lambda^* = \lambda/2^n$ so that the probability at zero is representable on the computer when λ^* is used as the Poisson mean. Equation (6.14) is now used to obtain the aggregate claims distribution when λ^* is used as the Poisson mean. If $P_*(z)$ is the pgf of the aggregate claims using Poisson mean λ^* , then $P_S(z) = [P_*(z)]^{2^n}$. Hence one can obtain successively the distributions with pgfs $[P_*(z)]^2, [P_*(z)]^4, [P_*(z)]^8, \dots, [P_*(z)]^{2^n}$ by convoluting the result at each stage with itself. This requires an additional n convolutions in carrying out the calculations but involves no approximations. This procedure can be carried out for any frequency distributions that are closed under convolution. For the negative binomial distribution, the analogous procedure starts with $r^* = r/2^n$. For the binomial distribution, the parameter m must be integer valued. A slight modification can be used. Let $m^* = \lfloor m/2^n \rfloor$ when $\lfloor \cdot \rfloor$ indicates the integer part of function. When the n convolutions are carried out, one still needs to carry out the calculations using (6.14) for parameter $m - m^*2^n$. This result is then convoluted with the result of the n convolutions. For compound frequency distributions, only the primary distribution needs to be closed under convolution.

6.6.3 Numerical stability

Any recursive formula requires accurate computation of values because each such value will be used in computing subsequent values. Recursive schemes suffer the risk of errors propagating through all subsequent values and potentially blowing up. In the recursive formula (6.14), errors are introduced through rounding or truncation at each stage because computers represent numbers with a finite number of significant digits. The question about stability is, "How fast do the errors in the calculations grow as the computed values are used in successive computations?"

The question of error propagation in recursive formulas has been a subject of study of numerical analysts. This work has been extended by Panjer and Wang [104] to study the recursive formula (6.14). The analysis is quite complicated and well beyond the scope of this book. However, some general conclusions can be made here.

Errors are introduced in subsequent values through the summation

$$\sum_{y=1}^x \left(a + \frac{by}{x} \right) f_X(y) f_S(x-y)$$

in recursion (6.14). In the extreme right-hand tail of the distribution of S , this sum is positive (or at least nonnegative), and subsequent values of the sum will be decreasing. The sum will stay positive, even with rounding errors, when each of the three factors in each term in the sum is positive. In this case, the recursive formula is stable, producing relative errors that do not grow fast. For the Poisson and negative binomial based distributions, the factors in each term are always positive.

On the other hand, for the binomial distribution, the sum can have negative terms because a is negative, b is positive, and y/x is a positive function not exceeding 1. In this case, the negative terms can cause the successive values to blow up with alternating signs. When this occurs, the nonsensical results are immediately obvious. Although this does not happen frequently in practice, the reader should be aware of this possibility in models based on the binomial distribution.

6.6.4 Continuous severity

The recursive method has been developed for discrete severity distributions, while it is customary to use continuous distributions for severity. In the case of continuous severities, the analog of the recursion (6.14) is an integral equation, the solution of which is the aggregate claims distribution.

Theorem 6.19 *For the $(a, b, 1)$ class of frequency distributions and any continuous severity distribution with probability on the positive real line, the fol-*

lowing integral equation holds:

$$f_S(x) = p_1 f_X(x) + \int_0^x \left(a + \frac{by}{x} \right) f_X(y) f_S(x-y) dy. \quad (6.19)$$

The proof of this result is beyond the scope of this book. For a detailed proof, see Theorems 6.14.1 and 6.16.1 of Panjer and Willmot [106], along with the associated corollaries. They consider the more general (a, b, m) class of distributions, which allow for arbitrary modification of m initial values of the distribution. Note that the initial term is $p_1 f_X(x)$, not $[p_1 - (a+b)p_0] f_X(x)$ as in (6.14). Also, (6.19) holds for members of the $(a, b, 0)$ class as well.

Integral equations of the form (6.19) are Volterra integral equations of the second kind. Numerical solution of this type of integral equation has been studied in the text by Baker [10]. We will develop a method using a discrete approximation of the severity distribution in order to use the recursive method (6.14) and avoid the more complicated methods of Baker [10].

6.6.5 Constructing arithmetic distributions

In order to implement recursive methods, the easiest approach is to construct a discrete severity distribution on multiples of a convenient unit of measurement h , the **span**. Such a distribution is called arithmetic because it is defined on the nonnegative integers. In order to arithmetize a distribution, it is important to preserve the properties of the original distribution both locally through the range of the distribution and globally—that is, for the entire distribution. This should preserve the general shape of the distribution and at the same time preserve global quantities such as moments.

The methods suggested here apply to the discretization (arithmetization) of continuous, mixed, and nonarithmetic discrete distributions.

6.6.5.1 Method of rounding (mass dispersal) Let f_j denote the probability placed at jh , $j = 0, 1, 2, \dots$. Then set²

$$\begin{aligned} f_0 &= \Pr \left(X < \frac{h}{2} \right) = F_X \left(\frac{h}{2} - 0 \right), \\ f_j &= \Pr \left(jh - \frac{h}{2} \leq X < jh + \frac{h}{2} \right) \\ &= F_X \left(jh + \frac{h}{2} - 0 \right) - F_X \left(jh - \frac{h}{2} - 0 \right), \quad j = 1, 2, \dots \end{aligned}$$

This method splits the probability between $(j+1)h$ and jh and assigns it to $j+1$ and j . This, in effect, rounds all amounts to the nearest convenient monetary unit, h , the span of the distribution.

²The notation $F_X(x-0)$ indicates that discrete probability at x should not be included. For continuous distributions this will make no difference.

6.6.5.2 Method of local moment matching In this method we construct an arithmetic distribution that matches p moments of the arithmetic and the true severity distributions. Consider an arbitrary interval of length ph , denoted by $[x_k, x_k + ph)$. We will locate point masses $m_0^k, m_1^k, \dots, m_p^k$ at points $x_k, x_k + h, \dots, x_k + ph$ so that the first p moments are preserved. The system of $p + 1$ equations reflecting these conditions is

$$\sum_{j=0}^p (x_k + jh)^r m_j^k = \int_{x_k-0}^{x_k+ph-0} x^r dF_X(x), \quad r = 0, 1, 2, \dots, p, \quad (6.20)$$

where the notation “ -0 ” at the limits of the integral indicates that discrete probability at x_k is to be included but discrete probability at $x_k + ph$ is to be excluded.

Arrange the intervals so that $x_{k+1} = x_k + ph$ and so the endpoints coincide. Then the point masses at the endpoints are added together. With $x_0 = 0$, the resulting discrete distribution has successive probabilities:

$$\begin{aligned} f_0 &= m_0^0, & f_1 &= m_1^0, & f_2 &= m_2^0, \dots, \\ f_p &= m_p^0 + m_0^1, & f_{p+1} &= m_1^1, & f_{p+2} &= m_2^1, \dots \end{aligned} \quad (6.21)$$

By summing (6.20) for all possible values of k , with $x_0 = 0$, it is clear that the first p moments are preserved for the entire distribution and that the probabilities add to 1 exactly. It only remains to solve the system of equations (6.20).

Theorem 6.20 *The solution of (6.20) is*

$$m_j^k = \int_{x_k-0}^{x_k+ph-0} \prod_{i \neq j} \frac{x - x_k - ih}{(j - i)h} dF_X(x), \quad j = 0, 1, \dots, p. \quad (6.22)$$

Proof: The Lagrange formula for collocation of a polynomial $f(y)$ at points y_0, y_1, \dots, y_n is

$$f(y) = \sum_{j=0}^n f(y_j) \prod_{i \neq j} \frac{y - y_i}{y_j - y_i}.$$

Applying this formula to the polynomial $f(y) = y^r$ over the points $x_k, x_k + h, \dots, x_k + ph$ yields

$$x^r = \sum_{j=0}^p (x_k + jh)^r \prod_{i \neq j} \frac{x - x_k - ih}{(j - i)h}, \quad r = 0, 1, \dots, p.$$

Integrating over the interval $[x_k, x_k + ph)$ with respect to the severity distribution results in

$$\int_{x_k-0}^{x_k+ph-0} x^r dF_X(x) = \sum_{j=0}^p (x_k + jh)^r m_j^k,$$

where m_j^k is given by (6.22). Hence, the solution (6.22) preserves the first p moments, as required. \square

Example 6.21 Suppose X has the exponential distribution with pdf $f(x) = 0.1e^{-0.1x}$. Use a span of $h = 2$ to discretize this distribution by the method of rounding and by matching the first moment.

For the method of rounding, the general formulas are

$$\begin{aligned} f_0 &= F(1) = 1 - e^{-0.1(1)} = 0.09516, \\ f_j &= F(2j+1) - F(2j-1) = e^{-0.1(2j-1)} - e^{-0.1(2j+1)}. \end{aligned}$$

The first few values are given in Table 6.9.

For matching the first moment we have $p = 1$ and $x_k = 2k$. The key equations become

$$\begin{aligned} m_0^k &= \int_{2k}^{2k+2} \frac{x - 2k - 2}{-2} (0.1)e^{-0.1x} dx = 5e^{-0.1(2k+2)} - 4e^{-0.1(2k)}, \\ m_1^k &= \int_{2k}^{2k+2} \frac{x - 2k}{2} (0.1)e^{-0.1x} dx = -6e^{-0.1(2k+2)} + 5e^{-0.1(2k)}, \end{aligned}$$

and then

$$\begin{aligned} f_0 &= m_0^0 = 5e^{-0.2} - 4 = 0.09365, \\ f_j &= m_1^{j-1} + m_0^j = 5e^{-0.1(2j-2)} - 10e^{-0.1(2j)} + 5e^{-0.1(2j+2)}. \end{aligned}$$

The first few values also are given in Table 6.9. A more direct solution for matching the first moment is provided in Exercise 6.36. \square

This method of local moment matching was introduced by Gerber and Jones [44] and Gerber [43] and further studied by Panjer and Lutek [103] for a variety of empirical and analytical severity distributions. In assessing the impact of errors on aggregate stop-loss net premiums (aggregate excess-of-loss pure premiums), Panjer and Lutek [103] found that two moments were usually sufficient and that adding a third moment requirement adds only marginally to the accuracy. Furthermore, the rounding method and the first-moment method ($p = 1$) had similar errors while the second-moment method ($p = 2$) provided significant improvement. The specific formulas for the method of rounding and the method of matching the first moment are given in Appendix E. A reason to favor matching zero or one moment is that the resulting probabilities will always be nonnegative. When matching two or more moments, this cannot be guaranteed.

The methods described here are qualitatively similar to numerical methods used to solve Volterra integral equations such as (6.19) developed in numerical analysis (see, for example, Baker [10]).

Table 6.9 Discretization of the exponential distribution by two methods

j	f_j rounding	f_j matching
0	0.09516	0.09365
1	0.16402	0.16429
2	0.13429	0.13451
3	0.10995	0.11013
4	0.09002	0.09017
5	0.07370	0.07382
6	0.06034	0.06044
7	0.04940	0.04948
8	0.04045	0.04051
9	0.03311	0.03317
10	0.02711	0.02716

6.6.6 Exercises

6.36 Show that the method of local moment matching with $k = 1$ (matching total probability and the mean) using (6.21) and (6.22) results in

$$f_0 = 1 - \frac{E[X \wedge h]}{h}$$

$$f_{i+1} = \frac{2E[X \wedge ih] - E[X \wedge (i-1)h] - E[X \wedge (i+1)h]}{h}, \quad i = 0, 1, 2, \dots,$$

and that $\{f_i; i = 0, 1, 2, \dots\}$ forms a valid distribution with the same mean as the original severity distribution. Using the formula given here, verify the formula given in Example 6.21.

6.37 You are the agent for a baseball player who desires an incentive contract that will pay the amounts given in Table 6.10. The number of times at bat has a Poisson distribution with $k = 200$. The parameter x is determined so that the probability of the player earning at least 4,000,000 is 0.95. Determine the player's expected compensation.

Table 6.10 Data for Exercise 6.37

Type of hit	Probability of hit per time at bat	Compensation per hit
Single	0.14	x
Double	0.05	$2x$
Triple	0.02	$3x$
Home run	0.03	$4x$

6.38 A weighted average of two Poisson distributions

$$p_k = w \frac{e^{-\lambda_1} \lambda_1^k}{k!} + (1-w) \frac{e^{-\lambda_2} \lambda_2^k}{k!}$$

has been used by some authors e.g., Tröbliger [130] to treat drivers as either "good" or "bad" (see Example 4.64).

- Find the pgf $P_N(z)$ of the number of losses in terms of the two pgfs $P_1(z)$ and $P_2(z)$ of the number of losses of the two types of drivers.
- Let $f_X(x)$ denote a severity distribution defined on the nonnegative integers. How can (6.16) be used to compute the distribution of aggregate claims for the entire group?
- Can this be extended to other frequency distributions?

6.39 A compound Poisson aggregate loss model has five expected claims per year. The severity distribution is defined on positive multiples of 1,000. Given that $f_S(1) = e^{-5}$ and $f_S(2) = \frac{5}{2}e^{-5}$, determine $f_X(2)$.

6.40 For a compound Poisson distribution, $\lambda = 6$ and individual losses have pf $f_X(1) = f_X(2) = f_X(4) = \frac{1}{3}$. Some of the pf values for the aggregate distribution S are given in Table 6.11. Determine $f_S(6)$.

6.41 Consider the $(a, b, 0)$ class of frequency distributions and any severity distribution defined on the positive integers $\{1, 2, \dots, M < \infty\}$, where M is the maximum possible single loss.

- Show that for the compound distribution the following backward recursion holds:

$$f_X(x) = \frac{f_S(x+M) - \sum_{y=1}^{M-1} \left(a + b \frac{M-y}{x+M}\right) f_X(M-y) f_S(x+y)}{\left(a + b \frac{M}{x+M}\right) f_X(M)}.$$

Table 6.11 Data for Exercise 6.40

x	$f_S(x)$
3	0.0132
4	0.0215
5	0.0271
6	$f_S(6)$
7	0.0410

- (b) For the binomial (m, q) frequency distribution, how can the above formula be used to obtain the distribution of aggregate losses? See Panjer and Wang [104].

6.42 Aggregate claims are compound Poisson with $\lambda = 2$, $f_X(1) = \frac{1}{4}$, and $f_X(2) = \frac{3}{4}$. For a premium of 6 an insurer covers aggregate claims and agrees to pay a dividend (a refund of premium) equal to the excess, if any, of 75% of the premium over 100% of the claims. Determine the excess of premium over expected claims and dividends.

6.43 On a given day, a physician provides medical care to N_A adults and N_C children. Assume N_A and N_C have Poisson distributions with parameters 3 and 2, respectively. The distributions of length of care per patient are as follows:

	Adult	Child
1 hour	0.4	0.9
2 hour	0.6	0.1

Let N_A , N_C , and the lengths of care for all individuals be independent. The physician charges 200 per hour of patient care. Determine the probability that the office income on a given day is less than or equal to 800.

6.44 A group policyholder's aggregate claims, S , has a compound Poisson distribution with $\lambda = 1$ and all claim amounts equal to 2. The insurer pays the group the following dividend:

$$D = \begin{cases} 6 - S, & S < 6, \\ 0, & S \geq 6. \end{cases}$$

Determine $E[D]$.

6.45 You are given two independent compound Poisson random variables S_1 and S_2 , where $f_j(x)$, $j = 1, 2$, are the two single-claim size distributions. You are given $\lambda_1 = \lambda_2 = 1$, $f_1(1) = 1$, and $f_2(1) = f_2(2) = 0.5$. Let $F_X(x)$ be the single-claim size distribution function associated with the compound distribution $S = S_1 + S_2$. Calculate $F_X^{*4}(6)$.

6.46 The variable S has a compound Poisson claims distribution with the following:

1. Individual claim amounts equal to 1, 2, or 3.
2. $E(S) = 56$.
3. $\text{Var}(S) = 126$.

Table 6.12 Data for Exercise 6.48

Deductible	Net premium
4	0.20
5	0.10
6	0.04
7	0.02

4. $\lambda = 29$.

Determine the expected number of claims of size 2.

6.47 For a compound Poisson distribution with positive integer claim amounts, the probability function follows:

$$f_S(x) = \frac{1}{x} [0.16f_S(x-1) + kf_S(x-2) + 0.72f_S(x-3)], \quad x = 1, 2, 3, \dots$$

The expected value of aggregate claims is 1.68. Determine the expected number of claims.

6.48 For a portfolio of policies you are given the following:

1. The number of claims has a Poisson distribution.
2. Claim amounts can be 1, 2, or 3.
3. A stop-loss reinsurance contract has net premiums for various deductibles as given in Table 6.12.

Determine the probability that aggregate claims will be either 5 or 6.

6.49 For group disability income insurance, the expected number of disabilities per year is 1 per 100 lives covered. The continuance (survival) function for the length of a disability in days, Y , is

$$\Pr(Y > y) = 1 - \frac{y}{10}, \quad y = 0, 1, \dots, 10.$$

The benefit is 20 per day following a five-day waiting period. Using a compound Poisson distribution, determine the variance of aggregate claims for a group of 1,500 independent lives.

6.50 A population has two classes of drivers. The number of accidents per individual driver has a geometric distribution. For a driver selected at random from Class I, the geometric distribution parameter has a uniform distribution over the interval $(0, 1)$. Twenty-five percent of the drivers are in Class I. All drivers in Class II have expected number of claims 0.25. For a driver selected

at random from this population, determine the probability of exactly two accidents.

Note: The following two Exercises require the use of a computer.

6.51 A policy covers physical damage incurred by the trucks in a company's fleet. The number of losses in a year has a Poisson distribution with $\lambda = 5$. The amount of a single loss has a gamma distribution with $\alpha = 0.5$ and $\theta = 2,500$. The insurance contract pays a maximum annual benefit of 20,000. Determine the probability that the maximum benefit will be paid. Use a span of 100 and the method of rounding.

6.52 An individual has purchased health insurance for which he pays 10 for each physician visit and 5 for each prescription. The probability that a payment will be 10 is 0.25, and the probability that it will be 5 is 0.75. The total number of payments per year has the Poisson-Poisson (Neyman Type A) distribution with $\lambda_1 = 10$ and $\lambda_2 = 4$. Determine the probability that total payments in one year will exceed 400. Compare your answer to a normal approximation.

6.53 Demonstrate that if the exponential distribution is discretized by the method of rounding, the resulting discrete distribution is a ZM geometric distribution.

6.7 THE IMPACT OF INDIVIDUAL POLICY MODIFICATIONS ON AGGREGATE PAYMENTS

In Section 5.6 the manner in which individual deductibles (both ordinary and franchise) affect both the individual loss amounts and the claim frequency distribution was discussed. In this section we will consider the impact on aggregate losses. It is worth noting that both individual coinsurance and individual policy limits have an impact on the individual losses but not on the frequency of such losses, so we will focus primarily on the deductible issues in what follows. We also remark that we continue to assume that the presence of policy modifications does not have an underwriting impact on the individual loss distribution through an effect on the risk characteristics of the insured population, an issue which was discussed in Section 5.6. That is, the *ground-up* distribution of the individual loss amount X is assumed to be unaffected by the policy modifications, and only the payments themselves are affected.

From the standpoint of the aggregate losses, the relevant facts are now described. Regardless of whether the deductible is of the ordinary or franchise type, we shall assume that an individual loss results in a payment with probability v . The individual ground-up loss random variable X has policy modifications (including deductibles) applied, so that a payment is then

made. Individual payments may then be viewed on a *per-loss* basis, where the amount of such payment, denoted by Y^L , will be 0 if the loss results in no payment. Thus, on a per-loss basis, the payment amount is determined on each and every loss. Alternatively, individual payments may also be viewed on a *per-payment* basis. In this case, the amount of payment is denoted by Y^P , and on this basis payment amounts are only determined on losses which actually result in a nonzero payment being made. Therefore, by definition, $\Pr(Y^P = 0) = 0$, and the distribution of Y^P is the conditional distribution of Y^L given that $Y^L > 0$. Notationally, we write $Y^P = Y^L | Y^L > 0$. Therefore, the cumulative distribution functions are related by

$$F_{Y^L}(y) = (1 - v) + vF_{Y^P}(y), \quad y \geq 0,$$

because $1 - v = \Pr(Y^L = 0) = F_{Y^L}(0)$ (recall that Y^L has a discrete probability mass point $1 - v$ at 0, even if X and hence Y^P and Y^L have continuous probability density functions for $y > 0$). The moment generating functions of Y^L and Y^P are thus related by

$$M_{Y^L}(t) = (1 - v) + vM_{Y^P}(t), \quad (6.23)$$

which may be restated in terms of expectations as

$$E(e^{tY^L}) = E(e^{tY^L} | Y^L = 0) \Pr(Y^L = 0) + E(e^{tY^L} | Y^L > 0) \Pr(Y^L > 0).$$

It follows from Section 5.6 that the number of losses N^L and the number of payments N^P are related through their probability generating functions by

$$P_{N^P}(z) = P_{N^L}(1 - v + vz), \quad (6.24)$$

where $P_{N^P}(z) = E(z^{N^P})$ and $P_{N^L}(z) = E(z^{N^L})$.

We now turn to the analysis of the aggregate payments. On a per loss basis, the total payments may be expressed as $S = Y_1^L + Y_2^L + \cdots + Y_{N^L}^L$ with $S = 0$ if $N^L = 0$ and where Y_j^L is the payment amount on the j th loss. Alternatively, ignoring losses on which no payment is made, we may express the total payments on a per-payment basis as $S = Y_1^P + Y_2^P + \cdots + Y_{N^P}^P$ with $S = 0$ if $N^P = 0$, and Y_j^P is the payment amount on the j th loss, which results in a nonzero payment. Clearly, S may be represented in two distinct ways on an aggregate basis. Of course, the moment generating function of S on a per-loss basis is

$$M_S(t) = E(e^{tS}) = P_{N^L} [M_{Y^L}(t)], \quad (6.25)$$

whereas on a per-payment basis we have

$$M_S(t) = E(e^{tS}) = P_{N^P} [M_{Y^P}(t)]. \quad (6.26)$$

Obviously, (6.25) and (6.26) are equal, as may be seen from (6.23) and (6.24). That is,

$$P_{N^L} [M_{Y^L}(t)] = P_{N^L} [1 - v + vM_{Y^P}(t)] = P_{N^P} [M_{Y^P}(t)].$$

Consequently, any analysis of the aggregate payments S may be done on either a per-loss basis [with compound representation (6.25) for the moment generating function] or on a per payment basis [with (6.26) as the compound moment generating function]. The basis selected should obviously be determined by whatever is more suitable for the particular situation at hand. While by no means a hard-and-fast rule, the authors have found it more convenient to use the per-loss basis to evaluate moments of S . In particular, the formulas given in Section 5.5 for the individual mean and variance are on a per-loss basis, and the mean and variance of the aggregate payments S may be computed using these and (6.6) but with N replaced by N^L and X by Y^L .

On the other hand, if the (approximated) distribution of S is of interest, then a payment basis is normally to be preferred. The reason for this choice is that on a per-loss basis underflow problems may result if $E(N^L)$ is large, and computer storage problems may occur due to the presence of a large number of zero probabilities in the distribution of Y^L , particularly if a franchise deductible is employed. Also, for convenience, we normally elect to apply policy modifications to the individual loss distribution first and then discretize (if necessary), rather than discretizing and then applying policy modifications to the discretized distributions. This issue is only relevant if the deductible and policy limit are not integer multiples of the discretization span, however. The following example illustrates these ideas.

Example 6.22 *The number of ground-up losses is Poisson distributed with mean $\lambda = 3$. The individual loss distribution is Pareto with parameters $\alpha = 4$ and $\theta = 10$. An individual ordinary deductible of 6, coinsurance of 75%, and an individual loss limit of 24 (before application of the deductible and coinsurance) are all applied. Determine the mean, variance, and distribution of aggregate payments.*

We will first compute the mean and variance on a per-loss basis. The mean number of losses is $E(N^L) = 3$, and the mean individual payment on a per loss basis is (using Theorem 5.13 with $r = 0$ and the Pareto distribution)

$$E(Y^L) = 0.75[E(X \wedge 24) - E(X \wedge 6)] = 0.75(3.2485 - 2.5195) = 0.54675.$$

The mean of the aggregate payments is thus

$$E(S) = E(N^L)E(Y^L) = (3)(0.54675) = 1.64.$$

The second moment of the individual payments on a per-loss basis is, using Theorem 5.14 with $r = 0$ and the Pareto distribution,

$$\begin{aligned} E[(Y^L)^2] &= (0.75)^2 \{E[(X \wedge 24)^2] - E[(X \wedge 6)^2] \\ &\quad - 2(6)E(X \wedge 24) + 2(6)E(X \wedge 6)\} \\ &= (0.75)^2 [26.3790 - 10.5469 - 12(3.2485) + 12(2.5195)] \\ &= 3.98481. \end{aligned}$$

In order to compute the variance of aggregate payments, we do not need to explicitly determine $\text{Var}(Y^L)$ because S is compound Poisson distributed, which implies [using (4.27), for example] that

$$\text{Var}(S) = \lambda E[(Y^L)^2] = 3(3.98481) = 11.9544 = (3.46)^2.$$

In order to compute the (approximate) distribution of S , we will use the per payment basis. First note that $v = \Pr(X > 6) = [10/(10 + 6)]^4 = 0.15259$, and the number of payments N^P is Poisson distributed with mean $E(N^P) = \lambda v = 3(0.15259) = 0.45776$. Let $Z = X - 6 | X > 6$, so that Z is the individual payment random variable with only a deductible of 6. Then

$$\Pr(Z > z) = \frac{\Pr(X > z + 6)}{\Pr(X > 6)}.$$

With coinsurance of 75%, $Y^P = 0.75Z$ has cumulative distribution function

$$F_{Y^P}(y) = 1 - \Pr(0.75Z > y) = 1 - \frac{\Pr(X > 6 + y/0.75)}{\Pr(X > 6)}.$$

That is, for y less than the maximum payment of $(0.75)(24 - 6) = 13.5$,

$$F_{Y^P}(y) = \frac{\Pr(X > 6) - \Pr(X > 6 + y/0.75)}{\Pr(X > 6)}, \quad y < 13.5,$$

and $F_{Y^P}(y) = 1$ for $y \geq 13.5$. We then discretize the distribution of Y^P (we thus apply the policy modifications first and then discretize) using a span of 2.25 and the method of rounding. This yields $f_0 = F_{Y^P}(1.125) = 0.30124$, $f_1 = F_{Y^P}(3.375) - F_{Y^P}(1.125) = 0.32768$, and so on. In this situation care must be exercised in evaluation of f_6 , and we have $f_6 = F_{Y^P}(14.625) - F_{Y^P}(12.375) = 1 - 0.94126 = 0.05874$. Then $f_n = 1 - 1 = 0$ for $n = 7, 8, \dots$. The approximate distribution of S may then be computed using the compound Poisson recursive formula, namely, $g_0 = e^{-0.45776(1-0.30124)} = 0.72625$, and

$$g_k = \frac{0.45776}{k} \sum_{j=1}^k j f_j g_{k-j}, \quad k = 1, 2, 3, \dots$$

Thus, $g_1 = (0.45776)(1)(0.32768)(0.72625) = 0.10894$, for example. \square

6.7.1 Exercises

6.54 Suppose that the number of ground-up losses N^L has probability generating function $P_{N^L}(z) = B[\theta(z - 1)]$, where θ is a parameter and B is functionally independent of θ . The individual ground-up loss distribution is exponential with cumulative distribution function $F_X(x) = 1 - e^{-\mu x}$, $x \geq 0$. Individual losses are subject to an ordinary deductible of d and coinsurance of

α . Demonstrate that the aggregate payments, on a per-payment basis, have compound moment generating function given by (6.26), where N^P has the same distribution as N^L but with θ replaced by $\theta e^{-\mu d}$ and Y^P has the same distribution as X but with μ replaced by μ/α .

6.55 A ground-up model of individual losses has the gamma distribution with parameters $\alpha = 2$ and $\theta = 100$. The number of losses has the negative binomial distribution with $r = 2$ and $\beta = 1.5$. An ordinary deductible of 50 and a loss limit of 175 (before imposition of the deductible) are applied to each individual loss.

- Determine the mean and variance of the aggregate payments on a per-loss basis.
- Determine the distribution of the number of payments.
- Determine the cumulative distribution function of the amount Y^P of a payment given that a payment is made.
- Discretize the severity distribution from (c) using the method of rounding and a span of 40.
- Use the recursive formula to calculate the discretized distribution of aggregate payments up to a discretized amount paid of 120.

6.8 CALCULATIONS WITH APPROXIMATE DISTRIBUTIONS

Whenever the severity distribution is calculated using an approximate method, the result is, of course, an approximation to the true aggregate distribution. In particular, the true aggregate distribution is often continuous (except, perhaps, with discrete probability at zero or at an aggregate censoring limit) while the approximate distribution either is discrete with probability at equally spaced values as with recursion and Fast Fourier Transform (FFT), is discrete with probability $1/n$ at arbitrary values as with simulation, or has a piecewise linear distribution function as with Heckman-Meyers. In this section we introduce reasonable ways to obtain values of $F_S(x)$ and $E[(S \wedge x)^k]$ from those approximating distributions. In all cases we assume that the true distribution of aggregate payments is continuous, except perhaps with discrete probability at $S = 0$.

6.8.1 Arithmetic distributions

For recursion and the FFT, the approximating distribution can be written as p_0, p_1, \dots , where $p_j = \Pr(S^* = jh)$ and S^* refers to the approximating distribution. While several methods of undiscretizing this distribution are possible, we will introduce only one. It assumes we can obtain $g_0 = \Pr(S = 0)$, the true probability that aggregate payments are zero. The method is based

Table 6.13 Discrete approximation to the aggregate payments distribution

j	x	$f_X(x)$	$p_j = f_{S^*}(x)$
0	0	0.009934	0.335556
1	2	0.019605	0.004415
2	4	0.019216	0.004386
3	6	0.018836	0.004356
4	8	0.018463	0.004327
5	10	0.018097	0.004299
6	12	0.017739	0.004270
7	14	0.017388	0.004242
8	16	0.017043	0.004214
9	18	0.016706	0.004186
10	20	0.016375	0.004158

on constructing a continuous approximation to S^* by assuming the probability p_j is uniformly spread over the interval $(j - \frac{1}{2})h$ to $(j + \frac{1}{2})h$ for $j = 1, 2, \dots$. For the interval from 0 to $h/2$, a discrete probability of g_0 is placed at zero and the remaining probability, $p_0 - g_0$, is spread uniformly over the interval. Let S^{**} be the random variable with this mixed distribution. All quantities of interest are then computed using S^{**} .

Example 6.23 Let N have the geometric distribution with $\beta = 2$ and let X have the exponential distribution with $\theta = 100$. Use recursion with a span of 2 to approximate the aggregate distribution and then obtain a continuous approximation.

The exponential distribution was discretized using the method which preserves the first moment. The probabilities appear in Table 6.13. Also presented are the aggregate probabilities computed using the recursive formula. We also note that $g_0 = \Pr(N = 0) = (1 + \beta)^{-1} = \frac{1}{3}$. For $j = 1, 2, \dots$ the continuous approximation has pdf $f_{S^{**}}(x) = f_{S^*}(2j)/2$, $2j - 1 < x \leq 2j + 1$. We also have $\Pr(S^{**} = 0) = \frac{1}{3}$ and $f_{S^{**}}(x) = (0.335556 - \frac{1}{3})/1 = 0.002223$, $0 < x \leq 1$. \square

Returning to the original problem, it is possible to work out the general formulas for the basic quantities. For the cdf,

$$\begin{aligned}
 F_{S^{**}}(x) &= g_0 + \int_0^x \frac{p_0 - g_0}{h/2} ds \\
 &= g_0 + \frac{2x}{h}(p_0 - g_0), \quad 0 \leq x \leq \frac{h}{2},
 \end{aligned}$$

and

$$\begin{aligned} F_{S^{**}}(x) &= \sum_{i=0}^{j-1} p_i + \int_{(j-1/2)h}^x \frac{p_j}{h} ds \\ &= \sum_{i=0}^{j-1} p_i + \frac{x - (j-1/2)h}{h} p_j, \quad \left(j - \frac{1}{2}\right)h < x \leq \left(j + \frac{1}{2}\right)h. \end{aligned}$$

For the limited expected value (LEV),

$$\begin{aligned} E[(S^{**} \wedge x)^k] &= 0^k g_0 + \int_0^x s^k \frac{p_0 - g_0}{h/2} ds + x^k [1 - F_{S^{**}}(x)] \\ &= \frac{2x^{k+1}(p_0 - g_0)}{h(k+1)} + x^k [1 - F_{S^{**}}(x)], \quad 0 < x \leq \frac{h}{2}, \end{aligned}$$

and

$$\begin{aligned} E[(S^{**} \wedge x)^k] &= 0^k g_0 + \int_0^{h/2} s^k \frac{p_0 - g_0}{h/2} ds + \sum_{i=1}^{j-1} \int_{(i-1/2)h}^{(i+1/2)h} s^k \frac{p_i}{h} ds \\ &\quad + \int_{(j-1/2)h}^x s^k \frac{p_j}{h} ds + x^k [1 - F_{S^{**}}(x)] \\ &= \frac{(h/2)^k (p_0 - g_0)}{k+1} + \sum_{i=1}^{j-1} \frac{h^k [(i+1/2)^{k+1} - (i-1/2)^{k+1}]}{k+1} p_i \\ &\quad + \frac{x^{k+1} - [(j-1/2)h]^{k+1}}{h(k+1)} p_j \\ &\quad + x^k [1 - F_{S^{**}}(x)], \quad \left(j - \frac{1}{2}\right)h < x \leq \left(j + \frac{1}{2}\right)h. \end{aligned}$$

For $k = 1$ this reduces to

$$E(S^{**} \wedge x) = \begin{cases} x(1 - g_0) - \frac{x^2}{h}(p_0 - g_0), & 0 < x \leq \frac{h}{2}, \\ \frac{h}{4}(p_0 - g_0) + \sum_{i=1}^{j-1} i h p_i + \frac{x^2 - [(j-1/2)h]^2}{2h} p_j \\ \quad + x[1 - F_{S^{**}}(x)], & \left(j - \frac{1}{2}\right)h < x \leq \left(j + \frac{1}{2}\right)h. \end{cases} \quad (6.27)$$

These formulas are summarized in Appendix E.

Example 6.24 (Example 6.23 continued) *Compute the cdf and LEV at integral values from 1 to 10 using S^* , S^{**} , and the exact distribution of aggregate losses.*

The exact distribution is available for this example. It was developed in Example 6.12 where it was determined that $\Pr(S = 0) = (1 + \beta)^{-1} = \frac{1}{3}$ and

Table 6.14 Comparison of true aggregate payment values and two approximations

x	cdf			LEV		
	S	S^*	S^{**}	S	S^*	S^{**}
1	0.335552	0.335556	0.335556	0.66556	0.66444	0.66556
2	0.337763	0.339971	0.337763	1.32890	1.32889	1.32890
3	0.339967	0.339971	0.339970	1.99003	1.98892	1.99003
4	0.342163	0.344357	0.342163	2.64897	2.64895	2.64896
5	0.344352	0.344357	0.344356	3.30571	3.30459	3.30570
6	0.346534	0.348713	0.346534	3.96027	3.96023	3.96025
7	0.348709	0.348713	0.348712	4.61264	4.61152	4.61263
8	0.350876	0.353040	0.350876	5.26285	5.26281	5.26284
9	0.353036	0.353040	0.353039	5.91089	5.90977	5.91088
10	0.355189	0.357339	0.355189	6.55678	6.55673	6.55676

the pdf for the continuous part is

$$f_S(x) = \frac{\beta}{\theta(1+\beta)^2} \exp\left[-\frac{x}{\theta(1+\beta)}\right] = \frac{2}{900} e^{-x/300}, \quad x > 0.$$

From this we have

$$F_S(x) = \frac{1}{3} + \int_0^x \frac{2}{900} e^{-s/300} ds = 1 - \frac{2}{3} e^{-x/300}$$

and

$$E(S \wedge x) = \int_0^x \frac{2s}{900} e^{-s/300} ds + x \frac{2}{3} e^{-x/300} = 200(1 - e^{-x/300}).$$

The requested values are given in Table 6.14. \square

6.8.2 Empirical distributions

When the approximate distribution is obtained by simulation (the simulation process is discussed in Chapter 17), the result is an empirical distribution. Unlike approximations produced by recursion or the FFT, simulation does not place the probabilities at equally spaced values. This makes it less clear how the approximate distribution should be smoothed. On the other hand, simulation usually involves tens or hundreds of thousands of points, and therefore the individual points are likely to be close to each other. For these reasons it seems sufficient to simply use the empirical distribution as the answer. That is, all calculations should be done using the approximate empirical random variable, S^* . The formulas for the commonly required quantities are very simple. Let x_1, x_2, \dots, x_n be the simulated values. Then

$$F_{S^*}(x) = \frac{\text{number of } x_j \leq x}{n}$$

Table 6.15 Simulated values of aggregate losses

j	x_j	j	x_j
1-331	0	346	6.15
332	0.04	347	6.26
333	0.12	348	6.58
334	0.89	349	6.68
335	1.76	350	6.71
336	2.16	351	6.82
337	3.13	352	7.76
338	3.40	353	8.23
339	4.38	354	8.67
340	4.78	355	8.77
341	4.95	356	8.85
342	5.04	357	9.18
343	5.07	358	9.88
344	5.81	359	10.12
345	5.94		

and

$$E[(S^* \wedge x)^k] = \frac{1}{n} \sum_{x_j < x} x_j^k + x^k [1 - F_{S^*}(x)].$$

Example 6.25 (Example 6.23 continued) *Simulate 1,000 observations from the compound model with geometric frequency and exponential severity. Use the results to obtain values of the cdf and LEV for the integers from 1 to 10. The small sample size was selected so that only about 30 values between zero and 10 (not including zero) are expected.*

The simulations produced an aggregate payment of zero 331 times. The set of nonzero values that were less than 10 plus the first value past 10 are presented in Table 6.15. Other than zero, none of the values appeared more than once in the simulation. The requested values from the empirical distribution along with the true values are given in Table 6.16. \square

6.8.3 Piecewise linear cdf

When using the Heckman-Meyers inversion method (to be introduced in Section 6.9.2), the output is approximate values of the cdf $F_S(x)$ at any set of desired values. The values are approximate because the severity distribution function is required to be piecewise linear and because approximate integration is used. Let $S^\#$ denote an arbitrary random variable with cdf values as given by the Heckman-Meyers method at arbitrarily selected points $0 = x_1 < x_2 < \dots < x_n$ and let $F_j = F_{S^\#}(x_j)$. Also, set $F_n = 1$ so that

Table 6.16 Empirical and smoothed values from a simulation

x	$F_{S^*}(x)$	$F_S(x)$	$E(S^* \wedge x)$	$E(S \wedge x)$
0	0.331	0.333	0.0000	0.0000
1	0.334	0.336	0.6671	0.6656
2	0.335	0.338	1.3328	1.3289
3	0.336	0.340	1.9970	1.9900
4	0.338	0.342	2.6595	2.6490
5	0.341	0.344	3.3206	3.3057
6	0.345	0.347	3.9775	3.9603
7	0.351	0.349	4.6297	4.6126
8	0.352	0.351	5.2784	5.2629
9	0.356	0.353	5.9250	5.9109
10	0.358	0.355	6.5680	6.5568

no probability is lost. The easiest way to complete the description of the smoothed distribution is to connect these points with straight lines. Let $S^{\#\#}$ be the random variable with this particular cdf. Intermediate values of the cdf of $S^{\#\#}$ are found by interpolation.

$$F_{S^{\#\#}}(x) = \frac{(x - x_{j-1})F_j + (x_j - x)F_{j-1}}{x_j - x_{j-1}}.$$

The formula for the limited expected value is (for $x_{j-1} < x \leq x_j$)

$$\begin{aligned} E[(S^{\#\#} \wedge x)^k] &= \sum_{i=2}^{j-1} \int_{x_{i-1}}^{x_i} s^k \frac{F_i - F_{i-1}}{x_i - x_{i-1}} ds \\ &\quad + \int_{x_{j-1}}^x s^k \frac{F_j - F_{j-1}}{x_j - x_{j-1}} ds + x^k [1 - F_{S^{\#\#}}(x)] \\ &= \sum_{i=2}^{j-1} \frac{(x_i^{k+1} - x_{i-1}^{k+1})(F_i - F_{i-1})}{(k+1)(x_i - x_{i-1})} \\ &\quad + \frac{(x^{k+1} - x_{j-1}^{k+1})(F_j - F_{j-1})}{(k+1)(x_j - x_{j-1})} \\ &\quad + x^k \left[1 - \frac{(x - x_{j-1})F_j + (x_j - x)F_{j-1}}{x_j - x_{j-1}} \right], \end{aligned}$$

and when $k = 1$,

$$\begin{aligned} E(S^{\#\#} \wedge x) &= \sum_{i=2}^{j-1} \frac{(x_i + x_{i-1})(F_i - F_{i-1})}{2} + \frac{(x^2 - x_{j-1}^2)(F_j - F_{j-1})}{2(x_j - x_{j-1})} \\ &\quad + x \left[1 - \frac{(x - x_{j-1})F_j + (x_j - x)F_{j-1}}{x_j - x_{j-1}} \right]. \end{aligned}$$

6.8.4 Exercises

6.56 Let the frequency (of losses) distribution be negative binomial with $r = 2$ and $\beta = 2$. Let the severity distribution (of losses) have the gamma distribution with $\alpha = 4$ and $\theta = 25$. Determine $F_S(200)$ and $E(S \wedge 200)$ for an ordinary per-loss deductible of 25. Use the recursive formula to obtain the aggregate distribution and use a discretization interval of 5 with the method of rounding to discretize the severity distribution.

6.57 (*Exercise 6.51 continued*) Recall that the number of claims has a Poisson distribution with $\lambda = 5$ and the amount of a single claim has a gamma distribution with $\alpha = 0.5$ and $\theta = 2,500$. Determine the mean, standard deviation, and 90th percentile of payments by the insurance company under each of the following coverages. Any computational method may be used.

- (a) A maximum aggregate payment of 20,000.
- (b) A per-claim ordinary deductible of 100 and a per claim maximum payment of 10,000. There is no aggregate maximum payment.
- (c) A per-claim ordinary deductible of 100 with no maximum payment. There is an aggregate ordinary deductible of 15,000, an aggregate coinsurance factor of 0.8, and a maximum insurance payment of 20,000. This corresponds to an aggregate reinsurance provision.

6.58 (*Exercise 6.52 continued*) Recall that the number of payments has the Poisson-Poisson distribution with $\lambda_1 = 10$ and $\lambda_2 = 4$ while the payment per claim by the insured is 5 with probability 0.75 and 10 with probability 0.25. Determine the expected payment by the insured under each of the following situations. Any computational method may be used.

- (a) A maximum payment of 400.
- (b) A coinsurance arrangement where the insured pays 100% up to an aggregate total of 300 and then pays 20% of aggregate payments above 300.

X 6.9 INVERSION METHODS

Inversion methods discussed in this section are used to obtain numerically the probability function, or some related function such as a net stop-loss premium (aggregate excess-of-loss pure premium), from a known expression for a transform, such as the pgf, mgf, or cf of the desired function.

Compound distributions lend themselves naturally to this approach because their transforms are compound functions and are easily evaluated when both frequency and severity components are known. The pgf and cf of the aggregate loss distribution are

$$P_S(z) = P_N[P_X(z)]$$

and

$$\varphi_S(z) = E[e^{iSz}] = P_N[\varphi_X(z)], \quad (6.28)$$

respectively. The characteristic function always exists and is unique. Conversely, for a given characteristic function, there always exists a unique distribution. The objective of inversion methods is to obtain the distribution numerically from the characteristic function (6.28).

It is worth mentioning that there has recently been much research in other areas of applied probability on obtaining the distribution numerically from the associated Laplace-Stieltjes transform. These techniques are applicable to the evaluation of compound distributions in the present context but will not be discussed further here. A good survey is [2], pp. 257–323.

6.9.1 Fast Fourier transform

The FFT is an algorithm that can be used for inverting characteristic functions to obtain densities of discrete random variables. The FFT comes from the field of signal processing. It was first used for the inversion of characteristic functions of compound distributions by Bertram [14] and is explained in detail with applications to aggregate loss calculation by Robertson [111].

Definition 6.26 For any continuous function $f(x)$, the *Fourier transform* is the mapping

$$\tilde{f}(z) = \int_{-\infty}^{\infty} f(x)e^{izx} dx. \quad (6.29)$$

The original function can be recovered from its Fourier transform as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(z)e^{-izx} dz.$$

When $f(x)$ is a probability density function, $\tilde{f}(z)$ is its characteristic function. For our applications, $f(x)$ will be real valued. From (6.29), $\tilde{f}(z)$ is complex valued. When $f(x)$ is a probability function of a discrete (or mixed) distribution, the definitions can be easily generalized (see, for example, Fisz [38]).

Definition 6.27 Let f_x denote a function defined for all integer values of x that is periodic with period length n (that is, $f_{x+n} = f_x$ for all x). For the vector $(f_0, f_1, \dots, f_{n-1})$, the *discrete Fourier transform* is the mapping \tilde{f}_k , $k = \dots, -1, 0, 1, \dots$, defined by

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n} jk\right), \quad k = \dots, -1, 0, 1, \dots \quad (6.30)$$

This mapping is bijective. In addition, \tilde{f}_k is also periodic with period length n . The inverse mapping is

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp\left(-\frac{2\pi i}{n} kj\right), \quad j = \dots, -1, 0, 1, \dots \quad (6.31)$$

This inverse mapping recovers the values of the original function.

Because of the periodic nature of f and \tilde{f} , we can think of the discrete Fourier transform as a bijective mapping of n points into n points. From (6.30), it is clear that, in order to obtain n values of \tilde{f}_k , the number of terms that need to be evaluated is of order n^2 , that is, $O(n^2)$.

The **Fast Fourier Transform (FFT)** is an algorithm that reduces the number of computations required to be of order $O(n \ln_2 n)$. This can be a dramatic reduction in computations when n is large. The algorithm exploits the property that a discrete Fourier transform of length n can be rewritten as the sum of two discrete transforms, each of length $n/2$, the first consisting of the even-numbered points and the second consisting of the odd-numbered points.

$$\begin{aligned} \tilde{f}_k &= \sum_{j=0}^{n-1} f_j \exp\left(\frac{2\pi i}{n} jk\right) \\ &= \sum_{j=0}^{n/2-1} f_{2j} \exp\left(\frac{2\pi i}{n} 2jk\right) + \sum_{j=0}^{n/2-1} f_{2j+1} \exp\left[\frac{2\pi i}{n} (2j+1)k\right] \\ &= \sum_{j=0}^{m-1} f_{2j} \exp\left(\frac{2\pi i}{m} jk\right) + \exp\left(\frac{2\pi i}{n} k\right) \sum_{j=0}^{m-1} f_{2j+1} \exp\left(\frac{2\pi i}{m} jk\right), \end{aligned}$$

when $m = n/2$. Hence

$$\tilde{f}_k = \tilde{f}_k^a + \exp\left(\frac{2\pi i}{n} k\right) \tilde{f}_k^b. \quad (6.32)$$

These can, in turn, be written as the sum of two transforms of length $m/2$. This can be continued successively. For the lengths $n/2, m/2, \dots$ to be integers, the FFT algorithm begins with a vector of length $n = 2^r$. The successive writing of the transforms into transforms of half the length will result, after r times, in transforms of length 1. Knowing the transform of length 1 will allow one to successively compose the transforms of length 2, $2^2, 2^3, \dots, 2^r$ by simple addition using (6.32). Details of the methodology are found in Press et al. [107].

In our applications, we use the FFT to invert the characteristic function when discretization of the severity distribution is done. This is carried out as follows:

1. Discretize the severity distribution using some methods such as those described in the previous section, obtaining the discretized severity distribution

$$f_X(0), f_X(1), \dots, f_X(n-1),$$

where $n = 2^r$ for some integer r and n is the number of points desired in the distribution $f_S(x)$ of aggregate claims.

2. Apply the FFT to this vector of values, obtaining $\varphi_X(z)$, the characteristic function of the discretized distribution. The result is also a vector of $n = 2^r$ values.
3. Transform this vector using the pgf transformation of the claim frequency distribution, obtaining $\varphi_S(z) = P_N[\varphi_X(z)]$, which is the characteristic function, that is, the discrete Fourier transform of the aggregate claims distribution, a vector of $n = 2^r$ values.
4. Apply the Inverse Fast Fourier Transform (IFFT), which is identical to the FFT except for a sign change and a division by n [see (6.31)]. This gives a vector of length $n = 2^r$ values representing the exact distribution of aggregate claims for the discretized severity model.

The FFT procedure requires a discretization of the severity distribution. When the number of points in the severity distribution is less than $n = 2^r$, the severity distribution vector must be padded with zeros until it is of length n .

When the severity distribution places probability on values beyond $x = n$, as is the case with most distributions discussed in Chapter 4, the probability that is missed in the right-hand tail beyond n can introduce some minor error in the final solution because the function and its transform are both assumed to be periodic with period n , when in reality they are not. The authors suggest putting all the remaining probability at the final point at $x = n$ so that the probabilities add up to 1 exactly. This allows for periodicity to be used for the severity distribution in the FFT algorithm and ensures that the final set of aggregate probabilities will sum to 1. However, it is imperative that n be selected to be large enough so that most all the aggregate probability occurs by the n th point. The following example provides an extreme illustration.

Example 6.28 Suppose the random variable X takes on the values 1, 2, and 3 with probabilities 0.5, 0.4, and 0.1, respectively. Further suppose the number of claims has the Poisson distribution with parameter $\lambda = 3$. Use the FFT to obtain the distribution of S using $n = 8$ and $n = 4,096$.

In either case, the probability distribution of X is completed by adding one zero at the beginning (because S places probability at zero, the initial representation of X must also have the probability at zero given) and either 4 or 4,092 zeros at the end. The results from employing the FFT and IFFT

Table 6.17 Aggregate probabilities computed by the FFT and IFFT

s	$n = 8$ $f_S(s)$	$n = 4,096$ $f_S(s)$
0	0.11227	0.04979
1	0.11821	0.07468
2	0.14470	0.11575
3	0.15100	0.13256
4	0.14727	0.13597
5	0.13194	0.12525
6	0.10941	0.10558
7	0.08518	0.08305

appear in Table 6.17. For the case $n = 8$, the eight probabilities sum to 1. For the case $n = 4,096$, the probabilities also sum to 1, but there is not room here to show them all. It is easy to apply the recursive formula to this problem, which verifies that all of the entries for $n = 4,096$ are accurate to the five decimal places presented. On the other hand, with $n = 8$, the FFT gives values that are clearly distorted. If any generalization can be made, it is that more of the extra probability has been added to the smaller values of S . \square

Because the FFT and IFFT algorithms are available in many computer software packages and because the computer code is short, easy to write, and available (e.g., [107], pp. 411–412), no further technical details about the algorithm are given here. The reader can read any one of numerous books dealing with FFTs for a more detailed understanding of the algorithm. The technical details which allow the speeding up of the calculations from $O(n^2)$ to $O(n \ln_2 n)$ relate to the detailed properties of the discrete Fourier transform. Robertson [111] gives a good explanation of the FFT as applied to calculating the distribution of aggregate claims.

6.9.2 Direct numerical inversion

The inversion of the characteristic function (6.28) has been done using approximate integration methods by Heckman and Meyers [51] in the case of Poisson, binomial, and negative binomial claim frequencies and continuous severity distributions. The method is easily extended to other frequency distributions.

In this method, the severity distribution function is replaced by a piecewise linear distribution. It further uses a maximum single-loss amount so the cdf jumps to 1 at the maximum possible individual loss. The range of the severity random variable is divided into intervals of possibly unequal length. The

remaining steps parallel those of the FFT method. Consider the cdf of the severity distribution $F_X(x)$, $0 \leq x < \infty$. Let $0 = x_0 < x_1 < \dots < x_n$ be arbitrarily selected loss values. Then the probability that losses lie in the interval $(x_{k-1}, x_k]$ is given by $f_k = F_X(x_k) - F_X(x_{k-1})$. Using a uniform density d_k over this interval results in the approximating density function $f^*(x) = d_k = f_k/(x_k - x_{k-1})$ for $x_{k-1} < x \leq x_k$. Any remaining probability $f_{n+1} = 1 - F_X(x_n)$ is placed as a spike at x_n . This approximating pdf is selected to make evaluation of the cf easy. It is not required for direct inversion. The cf of the approximating severity distribution is

$$\begin{aligned}\varphi_X(z) &= \int_0^\infty e^{izx} dF_X(x) \\ &= \sum_{k=1}^n \int_{x_{k-1}}^{x_k} d_k e^{izx} dx + f_{n+1} e^{izx_n} \\ &= \sum_{k=1}^n d_k \frac{e^{izx_k} - e^{izx_{k-1}}}{iz} + f_{n+1} e^{izx_n}.\end{aligned}$$

The cf can be separated into real and imaginary parts by using Euler's formula

$$e^{i\theta} = \cos(\theta) + i \sin(\theta).$$

Then the real part of the cf is

$$\begin{aligned}a(z) = \operatorname{Re}[\varphi_X(z)] &= \frac{1}{z} \sum_{k=1}^n d_k [\sin(zx_k) - \sin(zx_{k-1})] \\ &\quad + f_{n+1} \cos(zx_n)\end{aligned}$$

and the imaginary part is

$$\begin{aligned}b(z) = \operatorname{Im}[\varphi_X(z)] &= \frac{1}{z} \sum_{k=1}^n d_k [\cos(zx_{k-1}) - \cos(zx_k)] \\ &\quad + f_{n+1} \sin(zx_n).\end{aligned}$$

The cf of aggregate losses (6.28) is obtained as

$$\varphi_S(z) = P_N[\varphi_X(z)] = P_N[a(z) + ib(z)],$$

which can be rewritten as

$$\varphi_S(z) = r(z) e^{i\theta(z)}$$

because it is complex valued.

The distribution of aggregate claims is obtained as

$$F_S(x) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{r(z/\sigma)}{z} \sin\left(\frac{zx}{\sigma} - \frac{\theta z}{\sigma}\right) dz, \quad (6.33)$$

where σ is the standard deviation of the distribution of aggregate losses. Approximate integration techniques are used to evaluate (6.33) for any value of x . The reader is referred to Heckman and Meyers [51] for details. They also obtain the net stop-loss (excess pure) premium for the aggregate loss distribution as

$$\begin{aligned} P(d) &= E[(S - d)_+] = \int_d^\infty (s - d) dF_S(s) \\ &= \frac{\sigma}{\pi} \int_0^\infty \frac{r(z/\sigma)}{z^2} \left[\cos\left(\frac{\theta z}{\sigma}\right) - \cos\left(\frac{zd}{\sigma} - \frac{\theta z}{\sigma}\right) \right] dz \\ &\quad + \mu - \frac{d}{2} \end{aligned} \quad (6.34)$$

from (6.33), where μ is the mean of the aggregate loss distribution and d is the deductible.

Equation (6.33) provides only a single value of the distribution, while (6.34) provides only one value of the premium, but it does so quickly. The error of approximation depends on the spacing of the numerical integration method but is controllable.

6.9.3 Exercise

6.59 Repeat Exercises 6.51 and 6.52 using the inversion method.

6.10 COMPARISON OF METHODS

The recursive method has some significant advantages. The time required to compute an entire distribution of n points is reduced to $O(n^2)$ from $O(n^3)$ for the direct convolution method. Furthermore, it provides exact values when the severity distribution is itself discrete (arithmetic). The only source of error is in the discretization of the severity distribution. Except for binomial models, the calculations are guaranteed to be numerically stable. This method is very easy to program in a few lines of computer code. However, it has a few disadvantages. The recursive method only works for the classes of frequency distributions described in Section 4.6. Using distributions not based on the $(a, b, 0)$ and $(a, b, 1)$ classes requires modification of the formula or developing a new recursion. Numerous other recursions have recently been developed in the actuarial and statistical literature.

The FFT method is easy to use in that it uses standard routines available with many software packages. It is faster than the recursive method when n is large because it requires calculations of order $n \ln_2 n$ rather than n^2 . However, if the severity distribution has a fixed (and not too large) number of points, the recursive method will require fewer computations because the sum in (6.14) will have at most m terms, reducing the order of required computations to

be of order n , rather than n^2 in the case of no upper limit of the severity. The FFT method can be extended to the case where the severity distribution can take on negative values. Like the recursive method, it produces the entire distribution.

The direct inversion method has been demonstrated to be very fast in calculating a single value of the aggregate distribution or the net stop-loss (excess pure) premium for a single deductible d . However, it requires a major computer programming effort. It has been developed by Heckman and Meyers [51] specifically for $(a, b, 0)$ frequency models. It is possible to generalize the computer code to handle any distribution with a pgf that is a relatively simple function. This method is much faster than the recursive method when the expected number of claims is large. The speed does not depend on the size of λ in the case of the Poisson frequency model. In addition to being complicated to program, the method involves approximate integration whose errors depend on the method and interval size.

Through the use of transforms, both the FFT and inversion methods are able to handle convolutions efficiently. For example, suppose a reinsurance agreement was to cover the aggregate losses of three groups, each with unique frequency and severity distributions. If S_i , $i = 1, 2, 3$, are the aggregate losses for each group, the characteristic function for the total aggregate losses $S = S_1 + S_2 + S_3$ is $\varphi_S(z) = \varphi_{S_1}(z)\varphi_{S_2}(z)\varphi_{S_3}(z)$ and so the only extra work is some multiplications prior to the inversion step. The recursive method does not accommodate convolutions as easily.

The Heckman-Meyers method has some technical difficulties when being applied to severity distributions that are of the discrete type or have some anomalies, such as heaping of losses at some round number (e.g., 1,000,000). At any jump in the severity distribution function, a very short interval containing the jump needs to be defined in setting up the points (x_1, x_2, \dots, x_n) .

We save a discussion of simulation for last because it differs greatly from the other methods. For those not familiar with this method, an introduction is provided in Chapter 17. The major advantage is a big one. If you can carefully articulate the model, you should be able to obtain the aggregate distribution by simulation. The programming effort may take a little time but can be done in a straightforward manner. Today's computers will conduct the simulation in a reasonable amount of time. Most of the analytic methods were developed as a response to the excessive computing time that simulations used to require. That is less of a problem now. On the other hand, it is difficult to write a general-purpose simulation program. Instead, it is possibly necessary to write a new routine as each problem occurs. Thus it is probably best to save the simulation approach for those problems that cannot be solved by the other methods. Then, of course, it is worth the effort because there is no alternative.

One other drawback of simulation occurs in extremely low frequency situations (which is where recursion excels). For example, consider an individual excess-of-loss reinsurance in which reinsurance benefits are paid on individual losses above 1,000,000, an event which occurs about 1 time in 100, but when it

does, the tail is extremely long (for example, a Pareto distribution with small α). The simulation will have to discard 99% of the generated losses and then will need a large number of those that exceed the deductible (due to the large variation in losses). It may take a long time to obtain a reliable answer. One possible solution for simulation is to work with the conditional distribution of the loss variable, given that a payment has been made.

No method is clearly superior for all problems. Each method has both advantages and disadvantages when compared with the others. What we really have is an embarrassment of riches. Twenty-five years ago, actuaries wondered if there would ever be effective methods for determining aggregate distributions. Today we can choose from several.

6.11 THE INDIVIDUAL RISK MODEL

6.11.1 Parametric approximation

The **individual risk model** represents the aggregate loss as a fixed sum of independent (but not necessarily identically distributed) random variables:

$$S = X_1 + X_2 + \cdots + X_n.$$

This is usually thought of as the sum of the losses from n insurance contracts, for example, n persons covered under a group insurance policy.

The individual risk model was originally developed for life insurance in which the probability of death within a year is q_j and the fixed benefit paid for the death of the j th person is b_j . In this case, the distribution of the loss to the insurer for the j th policy is

$$f_{X_j}(x) = \begin{cases} 1 - q_j, & x = 0, \\ q_j, & x = b_j. \end{cases}$$

In this case the mean and variance of aggregate losses are

$$E(S) = \sum_{j=1}^n b_j q_j$$

and

$$\text{Var}(S) = \sum_{j=1}^n b_j^2 q_j (1 - q_j)$$

because the X_j s are assumed to be independent. Then, the pgf of aggregate losses is

$$P_S(z) = \prod_{j=1}^n (1 - q_j + q_j z^{b_j}). \quad (6.35)$$

In the special case where all the risks are identical with $q_j = q$ and $b_j = 1$, the pgf reduces to

$$P_S(z) = [1 + q(z - 1)]^n,$$

and in this case S has a binomial distribution.

The individual risk model can be generalized as follows. Let $X_j = I_j B_j$, where $I_1, \dots, I_n, B_1, \dots, B_n$ are independent. The random variable I_j is an indicator variable that takes on the value 1 with probability q_j and the value 0 with probability $1 - q_j$. This variable indicates whether or not the j th policy produced a payment. The random variable B_j can have any distribution and represents the amount of the payment in respect of the j th policy given that a payment was made. In the life insurance case, B_j is degenerate, with all probability on the value b_j . If we let $\mu_j = E(B_j)$ and $\sigma_j^2 = \text{Var}(B_j)$ then

$$E(S) = \sum_{j=1}^n q_j \mu_j \quad (6.36)$$

and

$$\text{Var}(S) = \sum_{j=1}^n [q_j \sigma_j^2 + q_j (1 - q_j) \mu_j^2]. \quad (6.37)$$

You are asked to verify these formulas in Exercise 6.60. The following example is a simple version of this situation.

Example 6.29 Consider a group life insurance contract with an accidental death benefit. Assume that for all members the probability of death in the next year is 0.01 and that 30% of deaths are accidental. For 50 employees the benefit for an ordinary death is 50,000 and for an accidental death it is 100,000. For the remaining 25 employees the benefits are 75,000 and 150,000, respectively. Develop an individual risk model and determine its mean and variance.

For all 75 employees $q_j = 0.01$. For 50 employees, B_j takes on the value 50,000 with probability 0.7 and 100,000 with probability 0.3. For them, $\mu_j = 65,000$ and $\sigma_j^2 = 525,000,000$. For the remaining 25 employees B_j takes on the value 75,000 with probability 0.7 and 150,000 with probability 0.3. For them, $\mu_j = 97,500$ and $\sigma_j^2 = 1,181,250,000$. Then

$$\begin{aligned} E(S) &= 50(0.01)(65,000) + 25(0.01)(97,500) \\ &= 56,875 \end{aligned}$$

and

$$\begin{aligned} \text{Var}(S) &= 50(0.01)(525,000,000) + 50(0.01)(0.99)(65,000)^2 \\ &\quad + 25(0.01)(1,181,250,000) + 25(0.01)(0.99)(97,500)^2 \\ &= 5,001,984,375. \end{aligned}$$

Table 6.18 Employee data for Example 6.30

Employee, j	Age (years)	Sex	Benefit, b_j	Mortality rate, q_j
1	20	M	15,000	0.00149
2	23	M	16,000	0.00142
3	27	M	20,000	0.00128
4	30	M	28,000	0.00122
5	31	M	31,000	0.00123
6	46	M	18,000	0.00353
7	47	M	26,000	0.00394
8	49	M	24,000	0.00484
9	64	M	60,000	0.02182
10	17	F	14,000	0.00050
11	22	F	17,000	0.00050
12	26	F	19,000	0.00054
13	37	F	30,000	0.00103
14	55	F	55,000	0.00479
Total			373,000	

□

When the risks are different, the probabilities defined by pgf (6.35) can be computed exactly or approximately. A normal, gamma, lognormal, or any other distribution can be used to approximate the distribution. This is usually done by matching the first few moments. Because the normal, gamma, and lognormal distributions each have two parameters, the mean and variance are sufficient.

Example 6.30 (Group life insurance) *A small manufacturing business has a group life insurance contract on its 14 permanent employees. The actuary for the insurer has selected a mortality table to represent the mortality of the group. Each employee is insured for the amount of his or her salary rounded up to the next 1,000 dollars. The group's data are given in Table 6.18.*

If the insurer adds a 45% relative loading to the net (pure) premium, what are the chances that it will lose money in a given year? Use the normal and lognormal approximations.

The mean and variance of the aggregate losses for the group are

$$E(S) = \sum_{j=1}^{14} b_j q_j = 2,054.41$$

and

$$\text{Var}(S) = \sum_{j=1}^{14} b_j^2 q_j (1 - q_j) = 1.02534 \times 10^8.$$

The premium being charged is $1.45 \times 2,054.41 = 2,978.89$. For the normal approximation (in units of 1,000), the mean is 2.05441 and the variance is 102.534. Then the probability of a loss is

$$\begin{aligned} \Pr(S > 2.97889) &= \Pr\left[Z > \frac{2.97889 - 2.05441}{(102.534)^{1/2}}\right] \\ &= \Pr(Z > 0.0913) \\ &= 0.46 \text{ or } 46\%. \end{aligned}$$

For the lognormal approximation (as in Example 6.5)

$$\mu + \frac{1}{2}\sigma^2 = \ln 2.05441 = 0.719989$$

and

$$2\mu + 2\sigma^2 = \ln(102.534 + 2.05441^2) = 4.670533.$$

From this $\mu = -0.895289$ and $\sigma^2 = 3.230555$. Then

$$\begin{aligned} \Pr(S > 2.97889) &= 1 - \Phi\left[\frac{\ln 2.97889 + 0.895289}{(3.230555)^{1/2}}\right] \\ &= 1 - \Phi(1.105) \\ &= 0.13 \text{ or } 13\%. \end{aligned}$$

□

In the next subsection we present several ways of obtaining the exact distribution of S for the case where the benefit amounts are fixed.

6.11.2 Exact calculation of the aggregate distribution

6.11.2.1 *Direct calculation* The pf of aggregate losses is given by

$$f_S(x) = f_{X_1} * f_{X_2} * \cdots * f_{X_n}(x), \quad (6.38)$$

where

$$f_{X_j}(x) = \begin{cases} p_j = 1 - q_j, & x = 0, \\ q_j, & x = b_j. \end{cases}$$

The density (6.38) can be calculated recursively over the partial sums $S_j = S_{j-1} + X_j$ for $j = 2, 3, \dots, n$ beginning with $S_1 = X_1$. Then

$$\begin{aligned} f_{S_j}(x) &= \begin{cases} f_{S_{j-1}}(x) f_{X_j}(0), & x < b_j, \\ f_{S_{j-1}}(x) f_{X_j}(0) + f_{S_{j-1}}(x - b_j) f_{X_j}(b_j), & x \geq b_j, \end{cases} \\ &= \begin{cases} p_j f_{S_{j-1}}(x), & x < b_j, \\ p_j f_{S_{j-1}}(x) + q_j f_{S_{j-1}}(x - b_j), & x \geq b_j. \end{cases} \end{aligned}$$

Table 6.19 Cumulative probabilities for Example 6.31

x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$
0	0.95273905	20	0.96157969	40	0.97335098	60	0.99933062
1	0.95273905	21	0.96157969	41	0.97335892	61	0.99933187
2	0.95273905	22	0.96157969	42	0.97338128	62	0.99933191
3	0.95273905	23	0.96157969	43	0.97338740	63	0.99933193
4	0.95273905	24	0.96621337	44	0.97340884	64	0.99933198
5	0.95273905	25	0.96621337	45	0.97341351	65	0.99933202
6	0.95273905	26	0.96998201	46	0.97342561	66	0.99933206
7	0.95273905	27	0.96998201	47	0.97342840	67	0.99933209
8	0.95273905	28	0.97114577	48	0.97343397	68	0.99933217
9	0.95273905	29	0.97114648	49	0.97343866	69	0.99933450
10	0.95273905	30	0.97212950	50	0.97345889	70	0.99934141
11	0.95273905	31	0.97330507	51	0.97346040	71	0.99934796
12	0.95273905	32	0.97330747	52	0.97346606	72	0.99935031
13	0.95273905	33	0.97331344	53	0.97346608	73	0.99936659
14	0.95321566	34	0.97331962	54	0.97347547	74	0.99937973
15	0.95463736	35	0.97332386	55	0.97806678	75	0.99941735
16	0.95599217	36	0.97332585	56	0.97807068	76	0.99944759
17	0.95646878	37	0.97332829	57	0.97807536	77	0.99945823
18	0.95984386	38	0.97333493	58	0.97807660	78	0.99953355
19	0.96035862	39	0.97334251	59	0.97807808	79	0.99956734

If we wish to calculate the distribution of total claims up to some value r , the computer time involved, as measured by the number of multiplications, can be seen to be of order nr . If both r and n are large (e.g., $r = 10,000$ and $n = 10,000$), the number of computations can be prohibitive.

Example 6.31 (Example 6.30 continued) Use the direct method to determine the pf of S as well as the probability required in Example 6.30.

$$\begin{aligned}
 f_{S_1}(0) &= 0.99851, \\
 f_{S_1}(15) &= 0.00149, \\
 f_{S_2}(0) &= p_2 f_{S_1}(0) = 0.99709212, \\
 f_{S_2}(15) &= p_2 f_{S_1}(15) = 0.00148788, \\
 f_{S_2}(16) &= p_2 f_{S_1}(16) + q_2 f_{S_1}(0) = 0.00141788, \\
 f_{S_2}(31) &= p_2 f_{S_1}(31) + q_2 f_{S_1}(15) = 0.0000021158.
 \end{aligned}$$

The final values of the cdf $F_S(x)$ for $x = 0, \dots, 79$ are given in Table 6.19. From Table 6.19, the probability of exceeding 2,978.89 is seen to be 0.047, which shows that both approximations in Example 6.30 are poor. \square

This approach is reasonable when n is not too large, but for larger groups an alternative method is needed.

6.11.2.2 Recursive calculation The following approach allows for the distribution to be calculated recursively based on De Pril [26]. We first divide the portfolio into subportfolios according to policy size and claim probability. Let n_{ij} be the number of policies with benefit i (where $i = 1, 2, \dots, r$)³ and claim probability q_j (where $j = 1, 2, \dots, m$). Then the pgf of total claims may be written as

$$P_S(z) = \prod_{i=1}^r \prod_{j=1}^m (1 - q_j + q_j z^i)^{n_{ij}}.$$

The logarithm of the pgf is

$$\ln P_S(z) = \sum_{i=1}^r \sum_{j=1}^m n_{ij} \ln(1 - q_j + q_j z^i). \quad (6.39)$$

We now differentiate (6.39) to obtain

$$P'_S(z) = P_S(z) \left[\sum_{i=1}^r \sum_{j=1}^m i q_j n_{ij} z^{i-1} (1 - q_j + q_j z^i)^{-1} \right]. \quad (6.40)$$

Setting $z = 1$ in (6.40) yields the mean of the total claims distribution, namely

$$E(S) = P'_S(1) = \sum_{i=1}^r \sum_{j=1}^m i q_j n_{ij}.$$

Now, (6.40) may be rewritten as

$$\begin{aligned}
 z P'_S(z) &= P_S(z) \left[\sum_{i=1}^r \sum_{j=1}^m i n_{ij} \left(\frac{q_j}{1 - q_j} z^i \right) \left(1 + \frac{q_j}{1 - q_j} z^i \right)^{-1} \right] \\
 &= P_S(z) \left[\sum_{i=1}^r \sum_{j=1}^m i n_{ij} \sum_{k=1}^{\infty} (-1)^{k-1} \left(\frac{q_j}{1 - q_j} \right)^k z^{ik} \right] \quad (6.41)
 \end{aligned}$$

for $|z| < \min_{i,j} [q_j^{-1}(1 - q_j)]^{1/i}$. The second term on the right-hand side of (6.41) may be rewritten as

$$\sum_{i=1}^r \sum_{k=1}^{\infty} h(i, k) z^{ik},$$

where

$$h(i, k) = i (-1)^{k-1} \sum_{j=1}^m n_{ij} \left(\frac{q_j}{1 - q_j} \right)^k. \quad (6.42)$$

³As in the discretization of severity distributions, it is necessary that the benefit amounts be in arithmetic progression. However, the monetary unit need not be 1. For example, $i = 1, 2, \dots$ could represent benefit amounts of 5,000, 10,000, \dots .

Thus, (6.41) may be written as

$$zP'_S(z) = P_S(z) \left[\sum_{i=1}^r \sum_{k=1}^{\infty} h(i, k) z^{ik} \right]. \quad (6.43)$$

The coefficient of z^x on the left-hand side of (6.43) is $xf_S(x)$, where $f_S(x)$ is the coefficient of z^x in $P_S(z)$. The right-hand side of (6.43) is a convolution, and the coefficient of z^x is thus given by

$$\sum_{ik \leq x} h(i, k) f_S(x - ik). \quad (6.44)$$

A simpler way of writing (6.44) is

$$\sum_{i=1}^x \sum_{k=1}^{\lfloor x/i \rfloor} h(i, k) f_S(x - ik)$$

where $\lfloor \cdot \rfloor$ denotes the greatest integer function, that is, the largest integer that is less than or equal to the argument. Finally, because $h(i, k) = 0$ if $i > x$, one may equate coefficients of z^x on both sides of (6.43) and divide by x to obtain

$$f_S(x) = \frac{1}{x} \sum_{i=1}^{x \wedge r} \sum_{k=1}^{\lfloor x/i \rfloor} h(i, k) f_S(x - ik), \quad x \geq 1. \quad (6.45)$$

Now,

$$f_S(0) = P_S(0) = \prod_{i=1}^r \prod_{j=1}^m (1 - q_j)^{n_{ij}}, \quad (6.46)$$

and from (6.45),

$$\begin{aligned} f_S(1) &= h(1, 1) f_S(0), \\ f_S(2) &= \frac{1}{2} \{ h(1, 1) f_S(1) + [h(1, 2) + h(2, 1)] f_S(0) \}, \\ &\vdots \end{aligned}$$

The probabilities $\{f_S(x); x = 1, 2, \dots\}$ may be calculated recursively using (6.45), beginning with (6.46).

It can be seen from (6.42) that $h(i, k)$ is a weighted sum of the k th power of $q_j/(1 - q_j)$, $j = 1, 2, \dots, m$. When q_j is close to zero, $[q_j/(1 - q_j)]^k$ is small. Consequently, the magnitude of $h(i, k)$ decreases rapidly as k increases. This suggests that the inner summation in (6.45) can be limited to a small number of terms without significant loss of accuracy in computations while speeding up the computations considerably.

If we limit k to a maximum of K terms, let

$$f_S^{(K)}(x) = \frac{1}{x} \sum_{i=1}^{x \wedge r} \sum_{k=1}^{K \wedge \lfloor x/i \rfloor} h(i, k) f_S^{(K)}(x - ik) \quad (6.47)$$

Table 6.20 Values of n_{ij} for Example 6.32

j	$1000q_j$	i													
		14	15	16	17	18	19	20	24	26	28	30	31	55	60
1	0.50	1	0	0	1	0	0	0	0	0	0	0	0	0	0
2	0.54	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	1.03	0	0	0	0	0	0	0	0	0	0	1	0	0	0
4	1.22	0	0	0	0	0	0	0	0	0	1	0	0	0	0
5	1.23	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6	1.28	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	1.42	0	0	1	0	0	0	0	0	0	0	0	0	0	0
8	1.49	0	1	0	0	0	0	0	0	0	0	0	0	0	0
9	3.53	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10	3.94	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	4.79	0	0	0	0	0	0	0	0	0	0	0	0	1	0
12	4.84	0	0	0	0	0	0	0	1	0	0	0	0	0	0
13	21.82	0	0	0	0	0	0	0	0	0	0	0	0	0	1

denote the approximation using at most K terms in (6.45). In a later paper, De Pril [27] shows that, if $q_j < \frac{1}{2}$, $j = 1, 2, \dots, m$, then

$$\sum_{x=0}^M |f_S(x) - f_S^{(K)}(x)| < e^{\delta(K)} - 1, \quad (6.48)$$

where

$$\delta(K) = \frac{1}{K+1} \sum_{i=1}^r \sum_{j=1}^m n_{ij} \frac{1 - q_j}{1 - 2q_j} \left(\frac{q_j}{1 - q_j} \right)^{K+1} \quad (6.49)$$

and $M = \sum_{i=1}^r \sum_{j=1}^m in_{ij}$ is the maximum possible aggregate claim amount.

The value of $\delta(K)$ is easily calculated for any value of K . Equation (6.48) provides an upper bound on the sum of the absolute errors over the entire distribution of aggregate claims and can be used to guarantee accuracy of results when a limited number of terms is used in (6.47).

Example 6.32 (Example 6.30 continued) *Determine the exact distribution of aggregate losses using the recursive method.*

The values of q_j and the nonzero rows of the matrix of n_{ij} values are given in Table 6.20.

Using (6.49), we find that $\delta(1) = 5.947 \times 10^{-4}$, $\delta(2) = 3.900 \times 10^{-6}$, $\delta(3) = 6.369 \times 10^{-8}$, and $\delta(4) = 1.131 \times 10^{-9}$. Hence, (6.47) with $K = 4$ will give us about 8 decimal place accuracy. The (nonzero) values of $h(i, k)$ computed using (6.42) are given in Table 6.21.

The values of $f_S(x)$ and the associated cdf $F_S(x)$ calculated using (6.47) with $K = 4$ and (6.46) are given in Table 6.22.

Table 6.21 Values of $h(i, k)$ for Example 6.32

i	$h(i, k)$			
	$k = 1$	$k = 2$	$k = 3$	$k = 4$
14	7.0035018×10^{-3}	$-3.5035025 \times 10^{-6}$	1.7526276×10^{-9}	$-8.7675218 \times 10^{-13}$
15	2.2383351×10^{-2}	$-3.3400962 \times 10^{-5}$	4.9841694×10^{-8}	$-7.4374944 \times 10^{-11}$
16	2.2752309×10^{-2}	$-3.2354221 \times 10^{-5}$	4.6008325×10^{-8}	$-6.5424723 \times 10^{-11}$
17	8.5042522×10^{-3}	$-4.2542531 \times 10^{-6}$	2.1281907×10^{-9}	$-1.0646277 \times 10^{-12}$
18	6.3765090×10^{-2}	$-2.2588816 \times 10^{-4}$	8.0020991×10^{-7}	$-2.8347477 \times 10^{-9}$
19	1.0265543×10^{-2}	$-5.5463884 \times 10^{-6}$	2.9966680×10^{-9}	$-1.6190750 \times 10^{-12}$
20	2.5632810×10^{-2}	$-3.2852048 \times 10^{-5}$	4.2104515×10^{-8}	$-5.3962851 \times 10^{-11}$
24	1.1672495×10^{-1}	$-5.6769638 \times 10^{-4}$	2.7610139×10^{-6}	$-1.3428300 \times 10^{-8}$
26	1.0284521×10^{-1}	$-4.0681297 \times 10^{-4}$	1.6091833×10^{-6}	$-6.3652612 \times 10^{-9}$
28	3.4201726×10^{-2}	$-4.1777074 \times 10^{-5}$	5.1030287×10^{-8}	$-6.2332996 \times 10^{-11}$
30	3.0931860×10^{-2}	$-3.1892665 \times 10^{-5}$	3.2883315×10^{-8}	$-3.3904736 \times 10^{-11}$
31	3.8176959×10^{-2}	$-4.7015487 \times 10^{-5}$	5.7900266×10^{-8}	$-7.1305033 \times 10^{-11}$
55	2.6471800×10^{-1}	$-1.2741022 \times 10^{-3}$	6.1323232×10^{-6}	$-2.9515207 \times 10^{-8}$
60	1.3384040	$-2.9855420 \times 10^{-2}$	6.6597689×10^{-4}	$-1.4855768 \times 10^{-5}$

It can be seen that the values in the last column of this table are identical to the corresponding values from Example 6.31 based on the direct method. This method is especially useful when there are a large number of lives in a group insurance contract. \square

Example 6.33 (An expanded group) *For the purpose of illustrating the effect of the size of the portfolio, consider a portfolio consisting of 1,400 independent lives, with exactly 100 lives like each life in the group life portfolio of Example 6.30. Determine the exact distribution of total losses.*

The values of n_{ij} are now either 100 or 0. From (6.42), it can be seen that each $h(i, k)$ is now 100 times larger than that of the previous example. The distribution of total claims can be computed as in the previous example. In Table 6.23 we give some values of the distribution function of total claims (as before, x is measured in units of 1,000).

From Example 6.30, the mean and variance of total claims for this portfolio of 1,400 lives are $\mu_1 = 205,441$ and $\mu_2 = 1.0253356 \times 10^{10}$. Also, the coefficient of skewness is calculated as $\mu_3(\mu_2)^{-3/2} = 0.5267345$, which is exactly one-tenth of the value of 5.267345 for the corresponding group of 14 lives. This indicates that the distribution is much more symmetric. \square

There are numerous papers on the subject of the individual risk model. De Pril [28] develops a generalization of this method to the case where the loss can take on more than one value.

Table 6.22 Aggregate probabilities for Example 6.32

x	$f_S(x)$	$F_S(x)$	x	$f_S(x)$	$F_S(x)$
0	9.5273905×10^{-1}	0.95273905	38	6.6436439×10^{-6}	0.97333493
14	4.7660783×10^{-4}	0.95321566	39	7.5742253×10^{-6}	0.97334251
15	1.4216995×10^{-3}	0.95463736	40	8.4744508×10^{-6}	0.97335098
16	1.3548133×10^{-3}	0.95599217	41	7.9416543×10^{-6}	0.97335892
17	4.7660783×10^{-4}	0.95646878	42	2.2356100×10^{-5}	0.97338128
18	3.3750829×10^{-3}	0.95984386	43	6.1253961×10^{-6}	0.97338740
19	5.1475706×10^{-4}	0.96035862	44	2.1435448×10^{-5}	0.97340884
20	1.2210690×10^{-3}	0.96157969	45	4.6721584×10^{-6}	0.97341351
24	4.6336840×10^{-3}	0.96621337	46	1.2100769×10^{-5}	0.97342561
26	3.7686403×10^{-3}	0.96998201	47	2.7915140×10^{-6}	0.97342840
28	1.1637614×10^{-3}	0.97114577	48	5.5621896×10^{-5}	0.97343397
29	7.1120536×10^{-7}	0.97114649	49	4.6964950×10^{-6}	0.97343866
30	9.8301077×10^{-4}	0.97212950	50	2.0226469×10^{-5}	0.97345889
31	1.1755723×10^{-3}	0.97330507	51	1.5103585×10^{-6}	0.97346040
32	2.3995910×10^{-6}	0.97330747	52	5.6661624×10^{-6}	0.97346607
33	5.9716305×10^{-6}	0.97331344	53	1.3705553×10^{-8}	0.97346608
34	6.1784053×10^{-6}	0.97331962	54	9.3923168×10^{-6}	0.97347547
35	4.2424878×10^{-6}	0.97332386	55	4.5913084×10^{-3}	0.97806678
36	1.9938909×10^{-6}	0.97332585	56	3.9003832×10^{-6}	0.97807068
37	2.4343694×10^{-6}	0.97332829	57	4.6823253×10^{-6}	0.97807536

6.11.3 Compound Poisson approximation

Because of the computational complexity of calculating the distribution of total claims for a portfolio of n risks using the individual risk model, it has been popular to attempt to approximate the distribution by using the compound Poisson distribution. As was seen in Section 6.5, use of the compound Poisson allows calculation of the total claims distribution by using a very simple recursive procedure or by using the Fast Fourier Transform. The pgf (6.35) of aggregate losses is

$$P_S(z) = \prod_{j=1}^n [1 + q_j(z^{b_j} - 1)].$$

By taking logarithms and using a Taylor series expansion of $\ln[1 + q_j(z^{b_j} - 1)]$, we obtain

$$\ln P_S(z) = \sum_{j=1}^n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} [q_j(z^{b_j} - 1)]^k.$$

Table 6.23 Aggregate distribution for Example 6.33

x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$
0	0.00789581	200	0.51793382	400	0.96031865	600	0.99927281
8	0.00789581	208	0.55208736	408	0.96528977	608	0.99939248
16	0.01059183	216	0.57594360	416	0.96999808	616	0.99950270
24	0.01906263	224	0.60583632	424	0.97360199	624	0.99958630
32	0.02561396	232	0.63412023	432	0.97694855	632	0.99965590
40	0.02979597	240	0.66687534	440	0.98031827	640	0.99971784
48	0.03774849	248	0.68632432	448	0.98297484	648	0.99976679
56	0.04770335	256	0.71292641	456	0.98515668	656	0.99980876
64	0.06976756	264	0.73984823	464	0.98728185	664	0.99984205
72	0.07610528	272	0.76061061	472	0.98914069	672	0.99987063
80	0.10013432	280	0.78007667	480	0.99068429	680	0.99989481
88	0.12399672	288	0.80016248	488	0.99194854	688	0.99991371
96	0.13839960	296	0.82073766	496	0.99321771	696	0.99992950
104	0.15945674	304	0.83672920	504	0.99421982	704	0.99994275
112	0.18004988	312	0.85121374	512	0.99504627	712	0.99995364
120	0.22162539	320	0.86849112	520	0.99580992	720	0.99996222
128	0.23569706	328	0.88253629	528	0.99644624	728	0.99996932
136	0.26335063	336	0.89343529	536	0.99701494	736	0.99997545
144	0.30172723	344	0.90530546	544	0.99745896	744	0.99998004
152	0.33041683	352	0.91610001	552	0.99785787	752	0.99998383
160	0.35497038	360	0.92599796	560	0.99821941	760	0.99998707
168	0.38635038	368	0.93340851	568	0.99850204	768	0.99998955
176	0.42177800	376	0.94180939	576	0.99873887	776	0.99999162
184	0.45476409	384	0.94903173	584	0.99895067	784	0.99999326
192	0.47881084	392	0.95482932	592	0.99912971	792	0.99999461

Retaining only the first term in the inner sum yields the approximation

$$\begin{aligned}\ln P_S(z) &\doteq \sum_{j=1}^n q_j(z^{b_j} - 1) \\ &= \lambda \sum_{j=1}^n \frac{\lambda_j}{\lambda} (z^{b_j} - 1),\end{aligned}\quad (6.50)$$

where $\lambda_j = q_j$ and $\lambda = \sum_{j=1}^n \lambda_j$. This results in

$$P_S(z) \doteq \exp\{\lambda[P_X(z) - 1]\},$$

which is the pgf of a compound Poisson distribution with individual loss distribution pgf

$$P_X(z) = \frac{1}{\lambda} \sum_{j=1}^n \lambda_j z^{b_j}. \quad (6.51)$$

Table 6.24 Aggregate distribution for Example 6.34

x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$	x	$F_S(x)$
0	0.9530099	20	0.9618348	40	0.9735771	60	0.9990974
1	0.9530099	21	0.9618348	41	0.9735850	61	0.9990986
2	0.9530099	22	0.9618348	42	0.9736072	62	0.9990994
3	0.9530099	23	0.9618348	43	0.9736133	63	0.9990995
4	0.9530099	24	0.9664473	44	0.9736346	64	0.9990995
5	0.9530099	25	0.9664473	45	0.9736393	65	0.9990996
6	0.9530099	26	0.9702022	46	0.9736513	66	0.9990997
7	0.9530099	27	0.9702022	47	0.9736541	67	0.9990997
8	0.9530099	28	0.9713650	48	0.9736708	68	0.9990998
9	0.9530099	29	0.9713657	49	0.9736755	69	0.9991022
10	0.9530099	30	0.9723490	50	0.9736956	70	0.9991091
11	0.9530099	31	0.9735235	51	0.9736971	71	0.9991156
12	0.9530099	32	0.9735268	52	0.9737101	72	0.9991179
13	0.9530099	33	0.9735328	53	0.9737102	73	0.9991341
14	0.9534864	34	0.9735391	54	0.9737195	74	0.9991470
15	0.9549064	35	0.9735433	55	0.9782901	75	0.9991839
16	0.9562597	36	0.9735512	56	0.9782947	76	0.9992135
17	0.9567362	37	0.9735536	57	0.9782994	77	0.9992239
18	0.9601003	38	0.9735604	58	0.9783006	78	0.9992973
19	0.9606149	39	0.9735679	59	0.9783021	79	0.9993307

This distribution has pf

$$\Pr(X = x) = \frac{1}{\lambda} \sum_{\{j: b_j = x\}} \lambda_j. \quad (6.52)$$

The numerator sums all probabilities associated with amount b_j .

Note that the means of the frequency distribution and the aggregate loss distribution match those of the exact distribution.

Example 6.34 (Example 6.30 continued) *Consider the group life case of Example 6.30. Derive a compound Poisson approximation.*

Using the compound Poisson approximation of this section with Poisson parameter $\lambda = \sum q_j = 0.04813$, the distribution function given in Table 6.24 is obtained.

When these values are compared to those of Example 6.31, it can be seen that the maximum error of 0.0002708 occurs at $x = 0$. \square

Some closely related approximations have been used. One popular one is to let λ_j in (6.50) be set to

$$\lambda_j = -\ln(1 - q_j), \quad j = 1, 2, \dots, n. \quad (6.53)$$

This matches the *no-loss* probability $1 - q_j$ with the no-loss probability of the Poisson distribution, $e^{-\lambda_j}$. This effectively replaces each life in the group by

a Poisson distribution. This approximation is appropriate in the context of a group life insurance contract where a life is "replaced" upon death, leaving the Poisson intensity unchanged by the death. Naturally the expected number of losses is greater than $\sum_{j=1}^n q_j$. An alternative choice was proposed by Kornya [79]. It used $\lambda_j = q_j/(1 - q_j)$ in (6.50). It results in an expected number of losses that exceeds that using (6.53) (see Exercise 6.61).

6.11.3.1 More than one possible loss amount It was noted in the beginning of this section that there may be more than one possible loss amount. Again let B_j be the random variable that measures the amount of the loss given that there was a loss and let $X_j = I_j B_j$. Then

$$P_{X_j}(z) = [1 - q_j + q_j P_{B_j}(z)].$$

The pgf corresponding to (6.35) is

$$P_S(z) = \prod_{j=1}^n [1 - q_j + q_j P_{B_j}(z)].$$

Although it is possible to extend the exact computational methods to this case, it is quite cumbersome. However, the compound Poisson approximation based on matching of moments (6.50) simply requires replacing z^{b_j} by $P_{B_j}(z)$. Then, the pgf of the severity distribution (6.51) becomes

$$P_X(z) = \frac{1}{\lambda} \sum_{j=1}^n \lambda_j P_{B_j}(z),$$

and so

$$f_X(x) = \frac{1}{\lambda} \sum_{j=1}^n \lambda_j f_{B_j}(x), \quad (6.54)$$

which is a weighted average of the n individual severity densities. Extensions to continuous severity distributions also satisfy (6.54) for all values of x .

Example 6.35 (Example 6.29 continued) *Develop compound Poisson approximations using all three methods suggested here. Compute the mean and variance for each approximation and compare it to the exact value.*

Using the method that matches the mean, we have $\lambda = 50(0.01) + 25(0.01) = 0.75$. The severity distribution is

$$\begin{aligned} f_X(50,000) &= \frac{50(0.01)(0.7)}{0.75} = 0.4667, \\ f_X(75,000) &= \frac{25(0.01)(0.7)}{0.75} = 0.2333, \\ f_X(100,000) &= \frac{50(0.01)(0.3)}{0.75} = 0.2000, \\ f_X(150,000) &= \frac{25(0.01)(0.3)}{0.75} = 0.1000. \end{aligned}$$

The mean is $\lambda E(X) = 0.75(75,833.33) = 56,875$, which matches the exact value, and the variance is $\lambda E(X^2) = 0.75(6,729,166,667) = 5,046,875,000$, which exceeds the exact value.

For the method that preserves the probability of no losses, $\lambda = -75 \ln(0.99) = 0.753775$. For this method, the severity distribution turns out to be exactly the same as before (this is due to the fact that all individuals have the same value of q_j). Thus the mean is 57,161 and the variance is 5,072,278,876, both of which exceed the previous approximate values.

Using Kornya's method, $\lambda = 75(0.01)/0.99 = 0.757576$ and again the severity distribution is unchanged. The mean is 57,449 and the variance is 5,097,853,535, which are the largest values of all. \square

6.11.4 Exercises

6.60 Derive (6.36) and (6.37).

6.61 Demonstrate that the compound Poisson model given by $\lambda_j = q_j$ and (6.52) produces a model with the same mean as the exact distribution but with a larger variance. Then show that the one using $\lambda_j = -\ln(1 - q_j)$ must produce a larger mean and even larger variance, and finally show that the one using $\lambda_j = q_j/(1 - q_j)$ must produce the largest mean and variance of all.

6.62 Individual members of an insured group have independent claims. The claim distribution has the statistics given in Table 6.25.

The premium for a group with future claims S is the mean of S plus 2 times the standard deviation of S . If the genders of the members of a group of m members are not known, the number of males is assumed to have a binomial distribution with parameters m and $q = 0.4$. Let A be the premium for a group of 100 for which the genders of the members are not known and let B is the premium for a group of 40 males and 60 females. Determine A/B .

6.63 An insurance company assumes claim probabilities for persons covered by its group life insurance contracts as given in Table 6.26.

A group of mutually independent lives has coverage of 1,000 per life. The company assumes that 20% of the lives are smokers. Based on this assumption, the premium is set equal to 110% of expected claims. If 30% of the lives are smokers, the probability that claims will exceed the premium is less than

Table 6.25 Data for Exercise 6.62

	Mean	Variance
Males	2	4
Females	4	10

Table 6.26 Data for Exercise 6.63

Class	Probability of claim
Smoker	0.02
Nonsmoker	0.01

0.20. Using the normal approximation, determine the minimum number of lives which must be in the group.

6.64 Based on the individual risk model with independent claims, the cumulative distribution function of aggregate claims for a portfolio of life insurance policies is as in Table 6.27.

One policy with face amount 100 and probability of claim 0.20 is increased in face amount to 200. Determine the probability that aggregate claims for the revised portfolio will not exceed 500.

6.65 A group life insurance contract covering independent lives is rated in the three age groupings as given in Table 6.28. The insurer prices the contract

Table 6.27 Distribution for Exercise 6.64

x	$F_S(x)$
0	0.40
100	0.58
200	0.64
300	0.69
400	0.70
500	0.78
600	0.96
700	1.00

Table 6.28 Data for Exercise 6.65

Age group	Number in age group	Probability of claim per life	Mean of the exponential distribution of claim amounts
18-35	400	0.03	5
36-50	300	0.07	3
51-65	200	0.10	2

Table 6.29 Data for Exercise 6.66

Service	Probability of claim	Distribution of annual charges given that a claim occurs	
		Mean	Variance
Office visits	0.7	160	4,900
Surgery	0.2	600	20,000
Other services	0.5	240	8,100

so that the probability that claims will exceed the premium is 0.05. Using the normal approximation, determine the premium that the insurer will charge.

6.66 The probability model for the distribution of annual claims per member in a health plan is shown in Table 6.29. Independence of costs and occurrences among services and members is assumed. Using the normal approximation, determine the minimum number of members that a plan must have such that the probability that actual charges will exceed 115% of the expected charges is less than 0.10.

6.67 An insurer has a portfolio of independent risks as given in Table 6.30. The insurer sets α and k such that aggregate claims have expected value 100,000 and minimum variance. Determine α .

6.68 An insurance company has a portfolio of independent one-year term life policies as given in Table 6.31. The actuary approximates the distribution of claims in the individual model using the compound Poisson model in which the expected number of claims is the same as in the individual model. Determine

Table 6.30 Data for Exercise 6.67

Class	Probability of claim	Benefit	Number of risks
Standard	0.2	k	3,500
Substandard	0.6	αk	2,000

Table 6.31 Data for Exercise 6.68

Class	Number in class	Benefit amount	Probability of a claim
1	500	x	0.01
2	500	$2x$	0.02

Table 6.32 Data for Exercise 6.69

Class	Benefit amount	Probability of death	Number of policies
1	100,000	0.10	500
2	200,000	0.02	500
3	300,000	0.02	500
4	200,000	0.10	300
5	200,000	0.10	500

the maximum value of x such that the variance of the compound Poisson approximation is less than 4,500.

6.69 An insurance company sold one-year term life insurance on a group of 2,300 independent lives as given in Table 6.32.

The insurance company reinsures amounts in excess of 100,000 on each life. The reinsurer wishes to charge a premium that is sufficient to guarantee that it will lose money 5% of the time on such groups. Obtain the appropriate premium by each of the following ways.

- Using a normal approximation to the aggregate claims distribution.
- Using a lognormal approximation.
- Using a gamma approximation.
- Using the compound Poisson approximation which matches the means.
- Carrying out the calculations exactly (using the method developed by De Pril or some other method). This requires a computer.

x 7

Discrete-time ruin models

7.1 INTRODUCTION

The risk assumed with a portfolio of insurance contracts is difficult to assess, but it is nevertheless important to attempt to do so in order to ensure the viability of an insurance operation. The distribution of total claims over a fixed period of time is an obvious input parameter to such a process, and this quantity has been the subject of the previous chapters.

In this chapter we take a multiperiod approach in which the fortunes of the policy, portfolio, or company are followed over time. The most common use of this approach is **ruin theory**, in which the quantity of interest is the amount of surplus, with ruin occurring when the surplus becomes negative. In order to track surplus we must model more than the claim payments. We must include premiums, investment income, and expenses, along with any other item that impacts the cash flow.

The models described in this and the next chapter are quite simple and idealized in order to maintain mathematical simplicity. Consequently the output from the analysis should not be viewed as a representation of absolute reality, but rather as important additional information on the risk associated with the portfolio of business. Such information is useful for long-run financial planning and maintenance of the insurer's solvency.

This chapter is organized into two parts. The first part (Section 7.2) introduces process models. The appropriate definitions are made and the terms of ruin theory defined. The second part (Section 7.3) analyzes discrete-time models. This can be done with the tools presented in the previous chapters. An analysis of continuous-time models is covered in the next chapter. This requires an introduction to stochastic processes. Two processes are analyzed: the compound Poisson process and Brownian motion. The compound Poisson process has been the standard model for ruin analysis in actuarial science, while Brownian motion has found considerable use in modern financial theory and also can be used as an approximation to the compound Poisson process.

7.2 PROCESS MODELS FOR INSURANCE

7.2.1 Processes

The major difference between this chapter and the earlier ones is that we now want to view the evolution of the portfolio over time. With that in mind, we define two kinds of processes. We note that, while processes that involve random events are usually called **stochastic processes**, we will not employ the modifier “stochastic” and instead trust that the context will make it clear which processes are random and which are not.

Definition 7.1 A *continuous-time process* is denoted $\{X_t; t \geq 0\}$. If there are random elements, it is sufficient to specify the joint distribution of $(X_{t_1}, \dots, X_{t_n})$ for all t_1, \dots, t_n and any n .

In general, it is insufficient to describe the process by specifying the distribution of X_t for arbitrary t . Many processes have correlations between the values observed at different times.

Example 7.2 Let $\{S_t; t \geq 0\}$ be the total losses paid from time 0 to time t . Indicate how the collective risk model of Chapter 6 may be used to describe this process.

For the joint distribution of $(S_{t_1}, \dots, S_{t_n})$, suppose $t_1 < \dots < t_n$. Let $W_j = S_{t_j} - S_{t_{j-1}}$ with $S_{t_0} = S_0 = 0$. Let the W_j have independent distributions given by the collective risk model. The individual loss distributions could be identical while the frequency distribution would have a mean that is proportional to the length of the time period, $t_j - t_{j-1}$. An example of a realization of this process (called a **sample path**) is given in Figure 7.1. \square

It is usually easier to describe a process if it does not change much over time. Two specific ways in which this can happen are defined below.

Definition 7.3 A process has *independent increments* if the random variables $X_t - X_s$ and $X_u - X_v$ are independent whenever $s < t \leq v < u$.

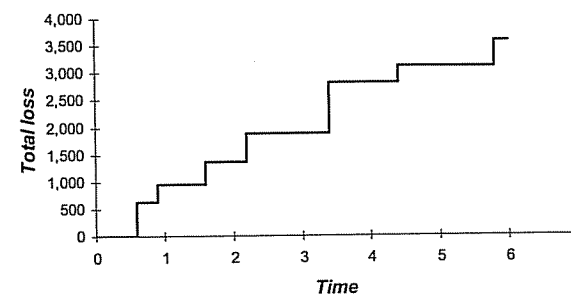


Fig. 7.1 Continuous total loss process, S_t .

This property indicates that the movement in the process in any one period is independent of the movement in a different, nonoverlapping period.

Definition 7.4 A process has *stationary increments* if the distribution of $X_t - X_s$ depends on t and s only through the difference $t - s$.

This property implies that the movement does not depend on the date. In other words, you cannot tell what time it is by looking at increments in the process.

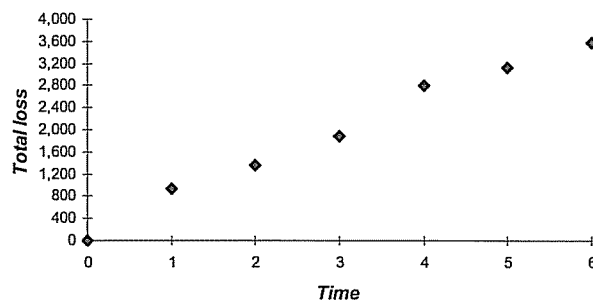
Most business organizations do not continuously monitor their status. Instead, it is checked at regular intervals. This leads to the other kind of process.

Definition 7.5 A *discrete-time process* is denoted by $\{X_t; t = 0, 1, 2, \dots\}$. If there are random elements, it is sufficient to specify the joint distribution of $(X_{t_1}, \dots, X_{t_n})$ for integer t_i and any n .

A discrete-time process can be derived from a continuous-time process by just writing down the values of X_t at integral times. In this chapter, all discrete-time processes will take measurements at the end of each observation period, such as a month, quarter, or year.

Example 7.6 (Example 7.2 continued) Convert the process to a discrete time process with stationary, independent increments.

Let X_1, X_2, \dots be the amount of the total losses in each period where the X_j s are i.i.d. and each X_j has a compound distribution. Then let the total loss process be $S_t = X_1 + \dots + X_t$. The process has stationary increments because $S_t - S_s = X_{s+1} + \dots + X_t$ and its distribution depends only on the number of X_j s, which is $t - s$. The property of independent increments follows directly from the independence of the X_j s. Figure 7.2 is the discrete-time version of Figure 7.1. \square

Fig. 7.2 Discrete total loss process, \tilde{S}_t .

7.2.2 An insurance model

The earlier examples have already illustrated most of the model we will use for the insurance process. We are interested in the **surplus process** $\{U_t; t \geq 0\}$ (or perhaps its discrete-time version, $\{U_t; t = 0, 1, \dots\}$), which measures the surplus of the portfolio at time t . We begin at time zero with $u = U_0$, the **initial surplus**. We think of the surplus in the accounting sense in that it represents excess funds that would not be needed if the portfolio terminated today. For an ongoing concern, a positive value provides protection against adversity. The surplus at time t is

$$U_t = U_0 + P_t - S_t,$$

where $\{P_t; t \geq 0\}$ is the premium process which measures all premiums (net of expenses) collected up to time t , and $\{S_t; t \geq 0\}$ is the loss process, which measures all losses paid up to time t . We make the following observations:

1. P_t may be written or earned premiums, as appropriate;
2. S_t may be paid or incurred losses, again, as appropriate; and
3. P_t may depend on S_u for $u < t$. For example, dividends based on favorable past loss experience may reduce the current premium.

It is possible, though not necessary, to separate the frequency and severity components of S_t . Let $\{N_t; t \geq 0\}$ be the claims process which records the number of claims as of time t . Then let $S_t = X_1 + \dots + X_{N_t}$. The sequence $\{X_1, X_2, \dots\}$ need not consist of i.i.d. variables. However, if it does and the sequence is independent of N_t for all t , then S_t will have a compound distribution.

We now look at two special cases of the surplus process. These are the only ones that will be studied.

7.2.2.1 A discrete-time model Let the increment in the surplus process in year t be defined as

$$W_t = P_t - P_{t-1} - S_t + S_{t-1}, \quad t = 1, 2, \dots$$

Then the progression of surplus is

$$U_t = U_{t-1} + W_t, \quad t = 1, 2, \dots$$

It will be relatively easy to learn about the distribution of $\{U_t; t = 1, 2, \dots\}$ provided that the random variable W_t either is independent of the other W_t s or only depends on the value of U_{t-1} . The dependency of W_t on U_{t-1} allows us to pay a dividend based on the surplus at the end of the previous year (because W_t depends on P_t).

In Section 7.3, two methods of determining the distribution of the U_t s will be presented. These will be computationally intensive, but given enough time and resources, the answers are easy to obtain.

7.2.2.2 A continuous-time model In most cases it is extremely difficult to analyze continuous-time models. This is because the joint distribution must be developed at every time point, not just at a countable set of time points. One model that has been extensively analyzed is the compound Poisson claim process where premiums are collected at a constant continuous nonrandom rate,

$$P_t = (1 + \theta)E(S_1)t,$$

and the total loss process is $S_t = X_1 + \dots + X_{N_t}$, where $\{N_t; t \geq 0\}$ is the Poisson process. This process is discussed in detail in Section 8.1.1. It suffices for now to note that for any period the number of losses has the Poisson distribution with a mean that is proportional to the length of the time period.

Because this model is more difficult to work with, the entire next chapter will be devoted to its development and analysis. We are now ready to define the quantity of interest, the one that measures the portfolio's chance for success.

7.2.3 Ruin

The main purpose for building a process model is to determine if the portfolio will survive over time. The probability of survival can be defined in four different ways.

Definition 7.7 The continuous-time, infinite-horizon survival probability is given by

$$\phi(u) = \Pr(U_t \geq 0 \text{ for all } t \geq 0 | U_0 = u).$$

Here we continuously check the surplus and demand that the portfolio be solvent forever. Both continuous checking and a requirement that the

portfolio survive forever are unrealistic. In practice, it is more likely that surplus is checked at regular intervals. While we would like our portfolio to last forever, it is too much to ask that our model be capable of forecasting infinitely far into the future. A more useful quantity follows.

Definition 7.8 *The discrete-time, finite-horizon survival probability is given by*

$$\tilde{\phi}(u, \tau) = \Pr(U_t \geq 0 \text{ for all } t = 0, 1, \dots, \tau | U_0 = u).$$

Here the portfolio is required to survive τ periods (usually years), and we only check at the end of each period. There are two intermediate cases.

Definition 7.9 *The continuous-time, finite-horizon survival probability is given by*

$$\phi(u, \tau) = \Pr(U_t \geq 0 \text{ for all } 0 \leq t \leq \tau | U_0 = u)$$

and the discrete-time, infinite-horizon survival probability is given by

$$\tilde{\phi}(u) = \Pr(U_t \geq 0 \text{ for all } t = 0, 1, \dots | U_0 = u).$$

It should be clear that the following inequalities hold:

$$\tilde{\phi}(u, \tau) \geq \tilde{\phi}(u) \geq \phi(u)$$

and

$$\tilde{\phi}(u, \tau) \geq \phi(u, \tau) \geq \phi(u).$$

There are also some limits that should be equally obvious. They are

$$\lim_{\tau \rightarrow \infty} \phi(u, \tau) = \phi(u)$$

and

$$\lim_{\tau \rightarrow \infty} \tilde{\phi}(u, \tau) = \tilde{\phi}(u).$$

In many cases, convergence is rapid. This means that the choice of a finite or infinite horizon should depend as much on the ease of calculation as on the appropriateness of the model. We will find that, if the Poisson process holds, infinite-horizon probabilities are easier to obtain. For other cases, the finite-horizon calculation may be easier.

Although we have not defined notation to express them, there is another pair of limits. As the frequency with which surplus is checked increases (that is, the number of times per year), the discrete-time survival probabilities converge to their continuous-time counterparts.

As this subsection refers to ruin, we close by defining the probability of ruin.

Definition 7.10 *The continuous-time, infinite-horizon ruin probability is given by*

$$\psi(u) = 1 - \phi(u)$$

and the other three ruin probabilities are defined and denoted in a similar manner.

7.3 DISCRETE, FINITE-TIME RUIN PROBABILITIES

7.3.1 The discrete-time process

Let P_t be the premium collected in the t th period and let S_t be the losses paid in the t th period. We also add one generalization. Let C_t be any cash flow other than the collection of premiums and the payment of losses. The most significant cash flow is the earning of investment income on the surplus available at the beginning of the period. The surplus at the end of the t th period is then

$$U_t = u + \sum_{j=1}^t (P_j + C_j - S_j) = U_{t-1} + P_t + C_t - S_t.$$

The final assumption is that, given U_{t-1} , the random variable $W_t = P_t + C_t - S_t$ depends only upon U_{t-1} and not upon any other previous experience. This makes $\{U_t; t = 1, 2, \dots\}$ a **Markov process**.

In order to evaluate ruin probabilities, we consider a second process defined as follows. First, define

$$\begin{aligned} W_t^* &= \begin{cases} 0, & U_{t-1}^* < 0, \\ W_t, & U_{t-1}^* \geq 0, \end{cases} \\ U_t^* &= U_{t-1}^* + W_t^*, \end{aligned} \quad (7.1)$$

where the new process starts with $U_0^* = u$. In this case, the finite-horizon survival probability is

$$\tilde{\phi}(u, \tau) = \Pr(U_\tau^* \geq 0).$$

The reason we need only check U_t^* at time τ is that, once ruined, this process is not allowed to become nonnegative. The following example illustrates this distinction and is a preview of the method presented in detail in the next subsection.

Example 7.11 *Consider a process with an initial surplus of 2, a fixed annual premium of 3, and losses of either 0 or 6 with probabilities 0.6 and 0.4, respectively. There are no other cash flows. Determine $\tilde{\phi}(2, 2)$.*

There are only two possible values for U_1 , 5 and -1 , with probabilities 0.6 and 0.4. In each year, W_t takes the values 3 and -3 with probabilities 0.6 and 0.4. For year 2, there are four possible ways for the process to end. They are listed in Table 7.1. Then, $\tilde{\phi}(2, 2) = 0.36 + 0.24 = 0.60$. Note that for U_2 the process would continue for cases 3 and 4, producing values of 2 and -4 . But our process is not allowed to recover from ruin and so case 3 must be forced to remain negative. \square

Table 7.1 Calculations for Example 7.11

Case	U_1	W_2	W_2^*	U_2^*	Probability
1	5	3	3	8	0.36
2	5	-3	-3	2	0.24
3	-1	3	0	-1	0.24
4	-1	-3	0	-1	0.16

7.3.2 Evaluating the probability of ruin

There are three ways to evaluate the ruin probability. One way that is always available is simulation. Just as the aggregate loss distribution can be simulated, the progress of surplus can also be simulated. For extremely complicated models (for example, one encompassing medical benefits, including hospitalization, prescription drugs, and outpatient visits as well as random inflation, interest rates, and utilization rates), this may be the only way to proceed. For more modest settings the other two methods will work well. The first is a brute force method that has few restrictions, and the second is an inversion method that has some restrictions.

7.3.2.1 Evaluation by convolutions For any practical use of this method, the distributions of all the random variables involved should be discrete and have finite support. If they are not, some discrete approximation should be constructed. The calculation is done recursively, using (7.1). For notational purposes, suppose we have obtained the discrete pf of U_{t-1}^* . Then the ruin probability is $\tilde{\psi}(u, t-1) = \Pr(U_{t-1}^* < 0)$ and the distribution of nonnegative surplus is $f_j = \Pr(U_{t-1}^* = u_j)$, $j = 1, 2, \dots, n$, where $u_j \geq 0$ for all j and u_n is the largest possible value of U_{t-1}^* . We have assumed that for each positive value of U_{t-1}^* the distribution of W_t is known. Let $g_{j,k} = \Pr(W_t = w_{j,k} | U_{t-1}^* = u_j)$. We have left open the possibility that even the values W_t may depend on u_j . We then obtain the probabilities of U_t^* by convolution. First,

$$\begin{aligned}
 \tilde{\psi}(u, t) &= \tilde{\psi}(u, t-1) + \Pr(U_{t-1}^* \geq 0 \text{ and } U_{t-1}^* + W_t < 0) \\
 &= \tilde{\psi}(u, t-1) + \sum_{j=1}^n \Pr(U_{t-1}^* + W_t < 0 | U_{t-1}^* = u_j) \Pr(U_{t-1}^* = u_j) \\
 &= \tilde{\psi}(u, t-1) + \sum_{j=1}^n \Pr(u_j + W_t < 0 | U_{t-1}^* = u_j) f_j \\
 &= \tilde{\psi}(u, t-1) + \sum_{j=1}^n \sum_{w_{j,k} < -u_j} g_{j,k} f_j.
 \end{aligned}$$

Then,

$$\begin{aligned}
 \Pr(U_t^* = x) &= \Pr(U_{t-1}^* \geq 0 \text{ and } U_{t-1}^* + W_t = x) \\
 &= \sum_{j=1}^n \Pr(U_{t-1}^* \geq 0 \text{ and } U_{t-1}^* + W_t = x | U_{t-1}^* = u_j) \\
 &\quad \times \Pr(U_{t-1}^* = u_j) \\
 &= \sum_{j=1}^n \Pr(u_j + W_t = x | U_{t-1}^* = u_j) f_j \\
 &= \sum_{j=1}^n \sum_{w_{j,k} + u_j = x} g_{j,k} f_j.
 \end{aligned}$$

Although these formulas look a bit intimidating, they are fairly easy to implement. Consider the following example.

Example 7.12 Suppose that annual losses can assume the values 0, 2, 4, and 6, with probabilities 0.4, 0.3, 0.2, and 0.1, respectively. Further suppose that the initial surplus is 2, and a premium of 2.5 is collected at the beginning of each year. Interest is earned at 10% on any surplus available at the beginning of the year because losses are paid at the end of the year. In addition, a rebate of 0.5 is given in any year in which there are no losses. Determine the survival probability at the end of each of the first two years.

First note that the rebate cannot be such that it is applied to the next year's premium. Doing so would require that we begin the year not only knowing the surplus but also if a rebate was to be provided.

At time zero, $\tilde{\psi}(2, 0) = 0$ and $f_1 = \Pr(U_0^* = 2) = 1$. The possible values of $w_{1,k}$ are given in Table 7.2 along with the probabilities, $g_{1,k}$.

For example, $w_{1,1}$ is based on a premium of 2.5, interest of 0.45 (on the surplus after collection of the premium), a loss payment of 0, and a rebate of 0.5. To evaluate $\tilde{\psi}(2, 1)$, observe that the only value of $w_{1,k}$ which is below $-u_1 = -2$ is $w_{1,4}$ and so $\tilde{\psi}(2, 1) = 0.1$. It is also easy to see that the only values of x that will have positive probability are those that have $2 + w_{1,k} > 0$. This gives the values for the distribution of U_1^* as in Table 7.3.

Table 7.2 w and g for Example 7.12

k	$w_{1,k}$	$g_{1,k}$
1	2.45	0.4
2	0.95	0.3
3	-1.05	0.2
4	-3.05	0.1

Table 7.3 U_1^* for Example 7.12

j	u_j	f_j
1	0.95	0.2
2	2.95	0.3
3	4.45	0.4

Table 7.4 $u + w$ and g for Example 7.12

j	u_j	f_j	k			
			1	2	3	4
1	0.95	0.2	3.295, 0.4	1.795, 0.3	-0.205, 0.2	-2.205, 0.1
2	2.95	0.3	5.495, 0.4	3.995, 0.3	1.995, 0.2	-0.005, 0.1
3	4.45	0.4	7.145, 0.4	5.645, 0.3	3.645, 0.2	1.645, 0.1

The remaining probability is at $\Pr(U_1^* = -1.05) = 0.1$.

One way to visualize year 2 is with a two-way table providing all of the combinations of u_j and $w_{j,k}$. The entries in Table 7.4 are $u_j + w_{j,k}$, $g_{j,k}$. Only the sums need be presented because they are the only interesting quantities.

The joint probability for any cell is the product of f_j from that row and the probability in that cell. The addition to $\tilde{\psi}(2, 1)$ is the probability for all the cells with negative entries, that is,

$$\tilde{\psi}(2, 2) = 0.1 + 0.2(0.2) + 0.2(0.1) + 0.3(0.1) = 0.19.$$

There are no duplicate values for $w_{i,k}$, so the best we can do is to put the values in order and note that they are the new u_j values for the beginning of year 3. They are listed in Table 7.5.

Table 7.5 u for Example 7.12

j	u_j	f_j
1	1.645	0.04
2	1.795	0.06
3	1.995	0.06
4	3.295	0.08
5	3.645	0.08
6	3.995	0.09
7	5.495	0.12
8	5.645	0.12
9	7.145	0.16

Table 7.6 Probabilities for Example 7.13

j	u_j	f_j
1	0	0.0134
2	2	0.189225
3	4	0.258975
4	6	0.2568
5	8	0.0916

The probabilities total 0.81, the complement of $\tilde{\psi}(2, 2)$. By the earlier definition, the remaining 0.19 probability is associated with $U_2^* < 0$. \square

It should be easy to see that the number of possible u values as well as the number of decimal places can increase rapidly. At some point, rounding would seem to be a good idea. A simple way to do this is to demand that at each period the only allowable u values are some multiple of h , a span that may need to increase from period to period. When probability is assigned to some value that is not a multiple of h , it is distributed to the two nearest values in a way that will preserve the mean (spreading to more values could preserve higher moments).

Example 7.13 (Example 7.12 continued) *Distribute the probabilities for the surplus at the end of year 2 using a span of $h = 2$.*

The probability of 0.04 at 1.645 must be distributed to the points 0 and 2. To preserve the mean, $0.355(0.04)/2 = 0.0071$ is placed at zero and the remaining 0.0329 is placed at 2. The expected value is $0.0071(0) + 0.0329(2) = 0.0658$, which matches the original value of $0.04(1.645)$. The value 0.355 is the distance from the point in question (1.645) to the next span point (2), and the denominator is the span. The probability is then placed at the previous span point. The resulting approximate distribution is given in Table 7.6. \square

7.3.2.2 Evaluation by inversion One of the strengths of the inversion method is that the act of computing a convolution is reduced to a few multiplications. This is true, provided that the random variables are independent. In this case that means that W_t is independent of U_{t-1} . We use a different approach with regard to keeping track of ruin (earlier that was accomplished by freezing U_t^* upon ruin). This idea could also be applied to the direct convolution approach. This time, let U_t^{**} be U_t conditioned on $U_t \geq 0$. At the end of each period, all probability associated with ruin is redistributed over the outcomes producing nonnegative surplus. The year-by-year analysis proceeds as follows:

1. Determine $\varphi_{1,t}(z) = E(e^{izU_{t-1}^{**}})$, the characteristic function of U_{t-1}^{**} .

2. Determine $\varphi_{2,t}(z) = E(e^{izW_t})$, the characteristic function of W_t .
3. Then $\varphi_{3,t}(z) = \varphi_{1,t}(z) \cdot \varphi_{2,t}(z)$ is the characteristic function of $U_{t-1}^{**} + W_t$.
4. Use inversion to determine $f_t(u)$, the pf of $U_{t-1}^{**} + W_t$.
5. Let $r_t = \Pr(U_{t-1}^{**} + W_t < 0)$. This is the probability that, given survival to time $t - 1$, the portfolio is ruined at time t .
6. Then $f_t^{**}(u) = f_t(u)/(1 - r_t)$ for $u \geq 0$ is the pf of U_t^{**} .
7. The probability of ruin by time t is then $\tilde{\psi}(u, t) = \tilde{\psi}(u, t - 1) + r_t[1 - \tilde{\psi}(u, t - 1)]$.

The process is initiated by noting that the pf of U_1 can be obtained directly by observing that $U_1 = u + W_1$, so all that needs to be done is to shift the arguments of the pf of W_1 by u .

Example 7.14 Aggregate losses for one year are 0, 2, 4, and 6 with probabilities 0.4, 0.3, 0.2, and 0.1, respectively. Premiums of 2.5 are collected at the beginning of the year and initial surplus is 2. Determine the probability of ruin within the first two years using the Fast Fourier Transform (FFT).

The pf of W_t is the same in all years and is given in Table 7.7. With an initial surplus of 2 it is easy to obtain the distribution of U_1 . It is given in Table 7.8. This immediately gives $\tilde{\psi}(2, 1) = 0.1$ and the distribution of U_1^{**} is given in Table 7.9.

Table 7.7 pf of W_t for Example 7.14

w	$\Pr(W = w)$
-3.5	0.1
-1.5	0.2
0.5	0.3
2.5	0.4

Table 7.8 pf of U_1 for Example 7.14

u	$\Pr(U_1 = u)$
-1.5	0.1
0.5	0.2
2.5	0.3
4.5	0.4

Table 7.9 pf of U_1^{**} for Example 7.14

u	$\Pr(U_1^{**} = u)$
0.5	2/9
2.5	3/9
4.5	4/9

In order for the FFT to work in a simple manner, it is best to have all amounts be positive. This can be accomplished by adding 3.5 to each variable. The shifted distributions are given in the second and third columns of Table 7.10. Anticipating that the shifted $U_1^{**} + W_2$ will take on values from 0 to 14 with a span of 2, we observe that eight values are required. This is already a power of 2, so no extra zeros need be added. In Table 7.10 the fourth and fifth columns provide the FFT of the two input variables. They are followed by the product of the two characteristic functions and then ultimately by the inverse of this characteristic function. The last column is the pf we seek. Of course, in this case it would have been trivial to perform the convolutions, but this way we can also verify that the FFT and its inverse do what they are supposed to.

Table 7.10 Year 2 ruin calculation for Example 7.14

u	$f_1^{**}(u)$	$f_W(u)$	$\varphi_{1,2}/8$	$\varphi_{2,2}/8$
0	0	1/10	0.125	0.125
2	0	2/10	$-0.08502 - 0.05724i$	$-0.00518 - 0.09053i$
4	2/9	3/10	$0.02778 + 0.04167i$	$-0.025 + 0.025i$
6	3/9	4/10	$-0.02609 - 0.00169i$	$0.03018 - 0.01553i$
8	4/9	0	0.04167	-0.025
10	0	0	$-0.02609 + 0.00169i$	$0.03018 + 0.01553i$
12	0	0	$0.02778 - 0.04167i$	$-0.025 - 0.25i$
14	0	0	$-0.08502 + 0.05724i$	$-0.00518 + 0.09053i$
u	$\varphi_{3,2}/64$	$f_2(u)$		
0	0.01563	0		
2	$-0.00474 + 0.00799i$	0		
4	$-0.00174 - 0.00035i$	2/90		
6	$-0.00081 + 0.00035i$	7/90		
8	-0.00104	16/90		
10	$-0.00081 - 0.00035i$	25/90		
12	$-0.00174 + 0.00035i$	24/90		
14	$-0.00474 - 0.00799i$	16/90		

Table 7.11 Distribution of surplus after year 2 for Example 7.14

u	$\Pr(U_1^{**} + W_2 = u)$	$\Pr(U_2^{**} = u)$
-3	2/90	0
-1	7/90	0
1	16/90	16/81
3	25/90	25/81
5	24/90	24/81
7	16/90	16/81

We must note that the probabilities in the last column are shifted by 7. The actual distribution of the sum is given in Table 7.11. We see that $\frac{9}{90} = \frac{1}{10}$ of the probability is associated with negative values and so $\tilde{\psi}(2, 2) = 0.1 + 0.9(0.1) = 0.19$. The conditional distribution U_2^{**} is also given in Table 7.11. \square

7.3.3 Exercises

7.1 The total claims paid in a year can be 0, 5, 10, 15, or 20 with probabilities 0.4, 0.3, 0.15, 0.1, and 0.05, respectively. Annual premiums of 6 are collected at the beginning of each year. Interest of 10% is earned on any funds available at the beginning of the year, and claims are paid at the end of the year.

- Determine $\tilde{\psi}(2, 3)$ exactly.
- Determine $\tilde{\psi}(2, 3)$, but after the premium is added and the interest credited, discretize the resulting distribution with a span of 5.

7.2 Repeat Exercise 7.1 using the FFT and rediscrctizing to maintain a span of 5.

8

Continuous-time ruin models

8.1 INTRODUCTION

In this chapter we turn to models that examine surplus continuously over time. Because these models tend to be difficult to analyze, we begin by restricting attention to models in which the number of claims has a Poisson distribution. In the discrete-time case we found that answers could be obtained by brute force. For the continuous case we find that exact, analytic solutions can be obtained for some situations, and that approximations and an upper bound can be obtained for many situations. In this section we introduce the Poisson process and the continuous-time approach to ruin.

8.1.1 The Poisson process

We consider the basic properties of the Poisson process $\{N_t; t \geq 0\}$ representing the number of claims on a portfolio of business. Thus, N_t is the number of claims in $(0, t]$. A formal definition of a Poisson process is now given.

Definition 8.1 *The number-of-claims process $\{N_t; t \geq 0\}$ is a Poisson process with rate $\lambda > 0$ if the following three conditions hold:*

- $N_0 = 0$.

2. The process has stationary and independent increments.
3. The number of claims in an interval of length t is Poisson distributed with mean λt . That is, for all $s, t > 0$ we have

$$\Pr(N_{t+s} - N_s = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, \dots \quad (8.1)$$

Stationary increments means that the distribution of the number of claims in a fixed interval depends only on the length of the interval and not on when the interval occurs so, for example, there is no trend effect. **Independent increments** means that the number of claims in an interval is statistically independent of the number of claims in any previous interval (not overlapping the present interval). Together, **stationary and independent increments** imply that the process can be thought of intuitively as starting over at any point in time. Actually, the assumption of stationarity in Condition 2 in the definition is not necessary because it is implied by Condition 3, but it is stated for clarity.

An important property of the Poisson process is that the times between claims are independent and identically exponentially distributed, each with mean $1/\lambda$. To see this, let W_j be the time between the $(j-1)$ th and j th claims for $j = 1, 2, \dots$, where W_1 is the time of the first claim. Then,

$$\Pr(W_1 > t) = \Pr(N_t = 0) = e^{-\lambda t},$$

and so W_1 is exponential with mean $1/\lambda$. Also,

$$\begin{aligned} \Pr(W_2 > t | W_1 = s) &= \Pr(W_1 + W_2 > s + t | W_1 = s) \\ &= \Pr(N_{t+s} = 1 | N_s = 1) \\ &= \Pr(N_{t+s} - N_s = 0 | N_s = 1) \\ &= \Pr(N_{t+s} - N_s = 0) \end{aligned}$$

because the increments are independent. From Condition 3, we then have

$$\Pr(W_2 > t | W_1 = s) = e^{-\lambda t}.$$

Because this is true for all s , $\Pr(W_2 > t) = e^{-\lambda t}$ and W_2 is independent of W_1 . Similarly, W_3, W_4, W_5, \dots are independent and exponentially distributed, each with mean $1/\lambda$.

Finally, we remark that, from a fixed point in time $t_0 \geq 0$ the time until the next claim occurs is also exponentially distributed with mean $1/\lambda$, due to the memoryless property of the exponential distribution. That is, if the n th claim occurred s time units before t_0 , the probability that the next claim occurs at least t time units after t_0 is $\Pr(W_{n+1} > t + s | W_{n+1} > s) = e^{-\lambda t}$, which is the same exponential survival function no matter what s and n happen to be. \square

8.1.2 The continuous-time problem

The model for claims payments will be the compound Poisson process. A formal definition follows.

Definition 8.2 Let the number of claims process $\{N_t; t \geq 0\}$ be a Poisson process with rate λ . Let the individual losses $\{X_1, X_2, \dots\}$ be independent and identically distributed positive random variables, independent of N_t , each with cumulative distribution function $F(x)$ and mean $\mu < \infty$. Thus X_j is the amount of the j th loss. Let S_t be the total loss in $(0, t]$. It is given by $S_t = 0$ if $N_t = 0$ and $S_t = \sum_{j=1}^{N_t} X_j$ if $N_t > 0$. Then, for fixed t , S_t has a compound Poisson distribution. The process $\{S_t; t \geq 0\}$ is said to be a **compound Poisson process**. Because $\{N_t; t \geq 0\}$ has stationary and independent increments, so does $\{S_t; t \geq 0\}$. Also,

$$E(S_t) = E(N_t)E(X_j) = (\lambda t)(\mu) = \lambda \mu t.$$

We assume that premiums are payable continuously at constant rate c per unit time. That is, the total net premium in $(0, t]$ is ct and we ignore interest for mathematical simplicity. We further assume that net premiums have a positive loading, that is, $ct > E(S_t)$, which implies that $c > \lambda \mu$. Thus let

$$c = (1 + \theta)\lambda \mu, \quad (8.2)$$

where $\theta > 0$ is called the **relative security loading** or **premium loading factor**.

For our model, we have now specified the loss and premium processes. The surplus process is thus

$$U_t = u + ct - S_t, \quad t \geq 0,$$

where $u = U_0$ is the initial surplus. We say that ruin occurs if U_t ever becomes negative, and survival occurs otherwise. Thus, the infinite-time survival probability is defined as

$$\phi(u) = \Pr(U_t \geq 0 \text{ for all } t \geq 0 | U_0 = u),$$

and the infinite-time ruin probability is

$$\psi(u) = 1 - \phi(u).$$

Our goal is to analyze $\phi(u)$ and/or $\psi(u)$.

8.2 THE ADJUSTMENT COEFFICIENT AND LUNDBERG'S INEQUALITY

In this section we determine a special quantity and then show that it can be used to obtain a bound on the value of $\psi(u)$. While it is only a bound, it is easy to obtain, and as an upper bound it provides a conservative estimate.

8.2.1 The adjustment coefficient

It is difficult to motivate the definition of the adjustment coefficient from a physical standpoint, so we just state it. We adopt the notational convention that X is an arbitrary claim size random variable in what follows.

Definition 8.3 Let $t = \kappa$ be the smallest positive solution to the equation

$$1 + (1 + \theta)\mu t = M_X(t), \quad (8.3)$$

where $M_X(t) = E(e^{tX})$ is the moment generating function of the claim severity random variable X . If such a value exists, it is called the **adjustment coefficient**.

To see that there may be a solution, consider the two lines in the (t, y) plane given by $y_1(t) = 1 + (1 + \theta)\mu t$ and $y_2(t) = M_X(t) = E(e^{tX})$. Now, $y_1(t)$ is a straight line with positive slope $(1 + \theta)\mu$. The mgf may not exist at all or may exist only for some values of t . Assume for this discussion that the mgf exists for all nonnegative t . Then $y_2'(t) = E(Xe^{tX}) > 0$ and $y_2''(t) = E(X^2 e^{tX}) > 0$. Because $y_1(0) = y_2(0) = 1$, the two curves intersect when $t = 0$. But $y_2'(0) = E(X) = \mu < (1 + \theta)\mu = y_1'(0)$. Thus, as t increases from 0 the curve $y_2(t)$ initially falls below $y_1(t)$, but because $y_2'(t) > 0$ and $y_2''(t) > 0$, eventually $y_2(t)$ will cross $y_1(t)$ at a point $\kappa > 0$. The point κ is the adjustment coefficient.

We remark that there may not be a positive solution to (8.3), for example, if the single claim amount distribution has no moment generating function (e.g., Pareto, lognormal).

Example 8.4 (Exponential claim amounts) If X has an exponential distribution with mean μ , determine the adjustment coefficient.

We have $F(x) = 1 - e^{-x/\mu}$, $x \geq 0$. Then, $M_X(t) = (1 - \mu t)^{-1}$, $t < \mu^{-1}$. Thus, from (8.3), κ satisfies

$$1 + (1 + \theta)\mu\kappa = (1 - \mu\kappa)^{-1}. \quad (8.4)$$

As noted earlier, $\kappa = 0$ is one solution and the positive solution is $\kappa = \theta/[\mu(1 + \theta)]$. The graph in Figure 8.1 displays plots of the left- and right-hand sides of (8.4) for the case $\theta = 0.2$ and $\mu = 1$. They intersect at 0 and at the adjustment coefficient, $\kappa = 0.2/1.2 = 0.1667$. \square

Example 8.5 (A gamma distribution) Suppose that the relative security loading is $\theta = 2$ and the gamma distribution has $\alpha = 2$. To avoid confusion, let β be the scale parameter for the gamma distribution. Determine the adjustment coefficient.

The single claim size density is

$$f(x) = \beta^{-2} x e^{-x/\beta}, \quad x > 0.$$

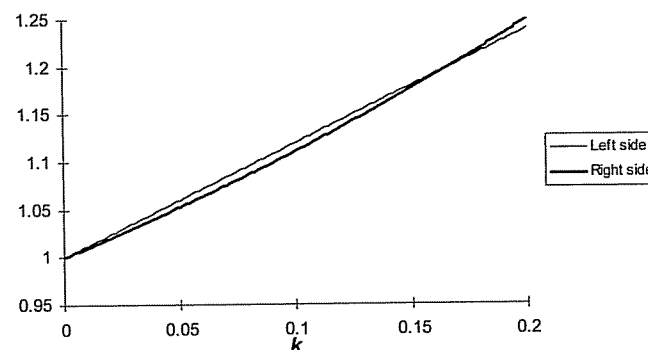


Fig. 8.1 Left and right sides of the adjustment coefficient equation.

For the gamma distribution $\mu = 2\beta$ and

$$M_X(t) = \int_0^\infty e^{tx} f(x) dx = (1 - \beta t)^{-2}, \quad t < \frac{1}{\beta}.$$

Then from (8.3) we obtain

$$1 + 6\kappa\beta = (1 - \beta\kappa)^{-2},$$

which may be rearranged as

$$6\beta^3 \kappa^3 - 11\beta^2 \kappa^2 + 4\beta \kappa = 0.$$

This is easily factored as

$$\kappa\beta(2\kappa\beta - 1)(3\kappa\beta - 4) = 0.$$

The adjustment coefficient is the only root that solves the original equation,¹ namely $\kappa = 1/(2\beta)$. \square

For general claim amount distributions, it is not possible to explicitly solve for κ as was done in the above two examples. Normally, one must resort to numerical methods, many of which require an initial guess as to the value of

¹Of the two roots, the larger one, $4/(3\beta)$, is not a legitimate argument for the mgf. The mgf exists only for values less than $1/\beta$. When solving such equations, the adjustment coefficient will always be the smallest positive solution.

κ . To find such a value, note that for (8.3) we may write

$$\begin{aligned} 1 + (1 + \theta)\mu\kappa &= E(e^{\kappa X}) \\ &= E(1 + \kappa X + \tfrac{1}{2}\kappa^2 X^2 + \cdots) \\ &> E(1 + \kappa X + \tfrac{1}{2}\kappa^2 X^2) \\ &= 1 + \kappa\mu + \tfrac{1}{2}\kappa^2 E(X^2). \end{aligned}$$

Then subtraction of $1 + \kappa\mu$ from both sides of the inequality and division by κ results in

$$\kappa < \frac{2\theta\mu}{E(X^2)}. \quad (8.5)$$

The right-hand side of (8.5) is usually a satisfactory initial value of κ . Other inequalities for κ are given in the exercises.

Example 8.6 *The aggregate loss random variable has variance equal to three times the mean. Determine a bound on the adjustment coefficient.*

For the compound Poisson distribution, $E(S_t) = \lambda\mu t$, $\text{Var}(S_t) = \lambda t E(X^2)$, and so $E(X^2) = 3\mu$. Hence, from (8.5), $\kappa < 2\theta/3$. \square

Define

$$H(t) = 1 + (1 + \theta)\mu t - M_X(t). \quad (8.6)$$

Then the adjustment coefficient $\kappa > 0$ satisfies $H(\kappa) = 0$. To solve this equation, use the Newton-Raphson formula,

$$\kappa_{j+1} = \kappa_j - \frac{H(\kappa_j)}{H'(\kappa_j)},$$

where

$$H'(t) = (1 + \theta)\mu - M'_X(t)$$

beginning with an initial value κ_0 . Because $H(0) = 0$, care must be taken so as not to converge to the value 0.

Example 8.7 *Suppose the Poisson parameter is $\lambda = 4$ and the premium rate is $c = 7$. Further suppose the individual loss amount distribution is given by*

$$\Pr(X = 1) = 0.6, \quad \Pr(X = 2) = 0.4.$$

Determine the adjustment coefficient.

We have

$$\mu = E(X) = (1)(0.6) + (2)(0.4) = 1.4$$

and

$$E(X^2) = (1)^2(0.6) + (2)^2(0.4) = 2.2.$$

Then $\theta = c(\lambda\mu)^{-1} - 1 = 7(5.6)^{-1} - 1 = 0.25$. From (8.5), we know that κ must be less than $\kappa_0 = 2(0.25)(1.4)/2.2 = 0.3182$. Now,

$$M_X(t) = 0.6e^t + 0.4e^{2t}$$

and so from (8.6)

$$H(t) = 1 + 1.75t - 0.6e^t - 0.4e^{2t}.$$

We also have

$$M'_X(t) = (1e^t)(0.6) + (2e^{2t})(0.4)$$

and so

$$H'(t) = 1.75 - 0.6e^t - 0.8e^{2t}.$$

Our initial guess is $\kappa_0 = 0.3182$. Then $H(\kappa_0) = -0.02381$ and $H'(\kappa_0) = -0.5865$. Thus, an updated estimate of κ is

$$\kappa_1 = 0.3182 - \frac{-0.02381}{-0.5865} = 0.2776.$$

Then $H(0.2776) = -0.003091$, $H'(0.2776) = -0.4358$, and

$$\kappa_2 = 0.2776 - \frac{-0.003091}{-0.4358} = 0.2705.$$

Continuing in this fashion, we get $\kappa_3 = 0.2703$, $\kappa_4 = 0.2703$, and so the adjustment coefficient $\kappa = 0.2703$ to four decimal places of accuracy. \square

There is another form for the adjustment coefficient equation (8.3) which is often useful. In particular, the following is an alternative definition of κ :

$$1 + \theta = \int_0^\infty e^{\kappa x} f_e(x) dx, \quad (8.7)$$

where

$$f_e(x) = \frac{1 - F(x)}{\mu}, \quad x > 0, \quad (8.8)$$

is the equilibrium probability density function discussed in Section 4.3.3.

To see that (8.7) is equivalent to (8.3), note that

$$\int_0^\infty e^{\kappa x} f_e(x) dx = \frac{M_X(\kappa) - 1}{\mu\kappa}$$

from Exercise 4.15. Thus replacement of $M_X(\kappa)$ by $1 + (1 + \theta)\mu\kappa$ in this expression yields (8.7).

8.2.2 Lundberg's inequality

The first main use of the adjustment coefficient lies in the following result.

Theorem 8.8 Suppose $\kappa > 0$ satisfies (8.3). Then the probability of ruin $\psi(u)$ satisfies

$$\psi(u) \leq e^{-\kappa u}, \quad u \geq 0. \quad (8.9)$$

Proof: Let $\psi_n(u)$ be the probability that ruin occurs on or before the n th claim for $n = 0, 1, 2, \dots$. We will prove by induction on n that $\psi_n(u) \leq e^{-\kappa u}$. Obviously, $\psi_0(u) = 0 \leq e^{-\kappa u}$. Now assume that $\psi_n(u) \leq e^{-\kappa u}$ and we wish to show that $\psi_{n+1}(u) \leq e^{-\kappa u}$. Let us consider what happens on the first claim. The time until the first claim occurs is exponential with probability density function $\lambda e^{-\lambda t}$. If the claim occurs at time $t > 0$, the surplus available to pay the claim at time t is $u + ct$. Thus, ruin occurs on the first claim if the amount of the claim exceeds $u + ct$. The probability that this happens is $1 - F(u + ct)$. If the amount of the claim is x , where $0 \leq x \leq u + ct$, ruin does not occur on the first claim. After payment of the claim, there is still a surplus of $u + ct - x$ remaining. Ruin can still occur on the next n claims. Because the surplus process has stationary and independent increments, this is the same probability as if we had started at the time of the first claim with initial reserve $u + ct - x$ and been ruined in the first n claims. Thus, by the law of total probability, we have the recursive equation²

$$\psi_{n+1}(u) = \int_0^\infty \left[1 - F(u + ct) + \int_0^{u+ct} \psi_n(u + ct - x) dF(x) \right] \lambda e^{-\lambda t} dt.$$

Thus, using the inductive hypothesis,

$$\begin{aligned} \psi_{n+1}(u) &= \int_0^\infty \left[\int_{u+ct}^\infty dF(x) + \int_0^{u+ct} \psi_n(u + ct - x) dF(x) \right] \lambda e^{-\lambda t} dt \\ &\leq \int_0^\infty \left[\int_{u+ct}^\infty e^{-\kappa(u+ct-x)} dF(x) \right. \\ &\quad \left. + \int_0^{u+ct} e^{-\kappa(u+ct-x)} dF(x) \right] \lambda e^{-\lambda t} dt, \end{aligned}$$

²The Stieltjes integral notation " $dF(x)$ " may be viewed as a notational convention to cover the situation where X is discrete, continuous, or mixed. In the continuous case, replace $dF(x)$ by $f(x)dx$ and proceed in the usual Riemann integral fashion. In the discrete case, the integral is simply notation for the usual sum involving the probability mass function.

where we have also used the fact that $-\kappa(u + ct - x) > 0$ when $x > u + ct$. Combining the two inner integrals gives

$$\begin{aligned} \psi_{n+1}(u) &\leq \int_0^\infty \left[\int_0^\infty e^{-\kappa(u+ct-x)} dF(x) \right] \lambda e^{-\lambda t} dt \\ &= \lambda e^{-\kappa u} \int_0^\infty e^{-\kappa ct} \left[\int_0^\infty e^{\kappa x} dF(x) \right] e^{-\lambda t} dt \\ &= \lambda e^{-\kappa u} \int_0^\infty e^{-(\lambda+\kappa c)t} [M_X(\kappa)] dt \\ &= \lambda M_X(\kappa) e^{-\kappa u} \int_0^\infty e^{-(\lambda+\kappa c)t} dt \\ &= \frac{\lambda M_X(\kappa)}{\lambda + \kappa c} e^{-\kappa u}. \end{aligned}$$

But from (8.3) and (8.2)

$$\lambda M_X(\kappa) = \lambda[1 + (1 + \theta)\kappa\mu] = \lambda + \kappa(1 + \theta)\lambda\mu = \lambda + \kappa c$$

and so $\psi_{n+1}(u) \leq e^{-\kappa u}$. Therefore, $\psi_n(u) \leq e^{-\kappa u}$ for all n and so $\psi(u) = \lim_{n \rightarrow \infty} \psi_n(u) \leq e^{-\kappa u}$. \square

This result is important because it may be used to examine the interplay between the level of surplus u and the premium loading θ , both parameters which are under the control of the insurer. Suppose one is willing to tolerate a probability of ruin of α (e.g., $\alpha = 0.01$) and a surplus of u is available. Then a loading of

$$\theta = \frac{u \left\{ E \left[\exp \left(-\frac{\ln \alpha}{u} X \right) \right] - 1 \right\}}{-\mu \ln \alpha} - 1$$

ensures that (8.3) is satisfied by $\kappa = (-\ln \alpha)/u$. Then, by Theorem 8.8, $\psi(u) \leq e^{-\kappa u} = e^{\ln \alpha} = \alpha$. On the other hand, if a specified loading of θ is desired, the surplus u required to ensure a ruin probability of no more than α is given by

$$u = \frac{-\ln \alpha}{\kappa}$$

because $\psi(u) \leq e^{-\kappa u} = e^{\ln \alpha} = \alpha$ as before.

Also, (8.9) allows us to show that

$$\psi(\infty) = \lim_{u \rightarrow \infty} \psi(u) = 0. \quad (8.10)$$

Because the ruin probability is also nonnegative, we have

$$0 \leq \psi(u) \leq e^{-\kappa u} \quad (8.11)$$

and thus

$$0 \leq \lim_{u \rightarrow \infty} \psi(u) \leq \lim_{u \rightarrow \infty} e^{-\kappa u} = 0,$$

which establishes (8.10). We then have that for the survival probability

$$\phi(\infty) = 1. \quad (8.12)$$

8.2.3 Exercises

8.1 Calculate the adjustment coefficient if $\theta = 0.32$ and the claim size distribution is the same as that of Example 8.5.

8.2 Calculate the adjustment coefficient if the individual loss size density is $f(x) = \sqrt{\beta/(\pi x)} e^{-\beta x}$, $x > 0$.

8.3 Calculate the adjustment coefficient if $c = 3$, $\lambda = 4$, and the individual loss size density is $f(x) = e^{-2x} + \frac{3}{2}e^{-3x}$, $x > 0$. Do not use an iterative numerical procedure.

8.4 If $c = 2.99$, $\lambda = 1$, and the individual loss size distribution is given by $\Pr(X = 1) = 0.2$, $\Pr(X = 2) = 0.3$, and $\Pr(X = 3) = 0.5$, use the Newton-Raphson procedure to numerically obtain the adjustment coefficient.

8.5 Repeat Exercise 8.3 using the Newton-Raphson procedure beginning with an initial estimate based on (8.5).

8.6 Suppose that $E(X^3)$ is known where X is a generic individual loss amount random variable. Prove that the adjustment coefficient κ satisfies

$$\kappa < \frac{-3E(X^2) + \sqrt{9[E(X^2)]^2 + 24\theta\mu E(X^3)}}{2E(X^3)}.$$

Also prove that the right-hand side of this inequality is strictly less than the bound given in (8.5), namely $2\theta\mu/E(X^2)$.

8.7 Recall that, if $g''(x) \geq 0$, Jensen's inequality implies $E[g(Y)] \geq g[E(Y)]$. Also, from Section 4.3.3,

$$\int_0^\infty x f_e(x) dx = \frac{E(X^2)}{2\mu},$$

where $f_e(x)$ is defined by (8.8).

(a) Use (8.7) and the above results to show that

$$\kappa \leq \frac{2\mu \ln(1 + \theta)}{E(X^2)}.$$

(b) Show that $\ln(1 + \theta) < \theta$ for $\theta > 0$ and thus the inequality in (a) is tighter than that from (8.5). *Hint:* Consider $h(\theta) = \theta - \ln(1 + \theta)$, $\theta > 0$.

(c) If there is a maximum claim size of m , show that (8.7) becomes

$$1 + \theta = \int_0^m e^{\kappa x} f_e(x) dx.$$

Show that the right-hand side of this equality satisfies

$$\int_0^m e^{\kappa x} f_e(x) dx \leq e^{\kappa m}$$

and hence that

$$\kappa \geq \frac{1}{m} \ln(1 + \theta).$$

8.8 In Section 4.3.3 it was shown that, if $F(x)$ has an increasing mean residual life [which is implied if $F(x)$ has a decreasing hazard rate], then

$$\int_x^\infty f_e(y) dy \geq 1 - F(x), \quad x \geq 0.$$

(a) Let Y have probability density function $f_e(y)$, $y \geq 0$, and let X have cumulative distribution function $F(x)$. Show that

$$\Pr(Y > y) \geq \Pr(X > y), \quad y \geq 0$$

and hence that

$$\Pr(e^{\kappa Y} > t) \geq \Pr(e^{\kappa X} > t), \quad t \geq 1.$$

(b) Use (a) to show that $E(e^{\kappa Y}) \geq E(e^{\kappa X})$.

(c) Use (b) to show that $\kappa \leq \theta/[\mu(1 + \theta)]$.

(d) Prove that, if the above inequality is reversed,

$$\kappa \geq \frac{\theta}{\mu(1 + \theta)}.$$

8.9 Suppose that $\kappa > 0$ satisfies (8.3) and also that

$$S(x) \leq \rho e^{-\kappa x} \int_x^\infty e^{\kappa y} dF(y) \quad (8.13)$$

for $0 < \rho \leq 1$, where $S(x) = 1 - F(x)$. Prove that $\psi(u) \leq \rho e^{-\kappa u}$, $u \geq 0$. *Hint:* Use the method of Theorem 8.8.

8.10 Continue the previous exercise. Use integration by parts to show that

$$\int_x^\infty e^{\kappa y} dF(y) = e^{\kappa x} S(x) + \kappa \int_x^\infty e^{\kappa y} S(y) dy, \quad x \geq 0.$$

8.11 Suppose $F(x)$ has a decreasing hazard rate (Section 4.3.3). Prove that $S(y) \geq S(x)S(y-x)$, $x \geq 0$, $y \geq x$. Then use Exercise 8.10 to show that (8.13) is satisfied with $\rho^{-1} = E(e^{\kappa X})$. Use (8.3) to conclude that

$$\psi(x) \leq [1 + (1 + \theta)\kappa\mu]^{-1} e^{-\kappa x}, \quad x \geq 0.$$

8.12 Suppose $F(x)$ has a hazard rate $\mu(x) = -(d/dx) \ln S(x)$ which satisfies $\mu(x) \leq m < \infty$, $x \geq 0$. Use the result in Exercise 8.10 to show that (8.13) is satisfied with $\rho = 1 - \kappa/m$ and thus

$$\psi(x) \leq (1 - \kappa/m) e^{-\kappa x}, \quad x \geq 0.$$

Hint: Show that, for $y > x$, $S(y) \geq S(x)e^{-(y-x)m}$.

8.3 AN INTEGRODIFFERENTIAL EQUATION

We now consider the problem of finding an explicit formula for the ruin probability $\psi(u)$ or (equivalently) the survival probability $\phi(u)$. It will be useful in what follows to consider a slightly more general function.

Definition 8.9 $G(u, y) = \Pr(\text{ruin occurs with initial reserve } u \text{ and deficit immediately after ruin occurs is at most } y)$, $u \geq 0, y \geq 0$.

For the event described, the surplus immediately after ruin is between 0 and $-y$. We then have

$$\psi(u) = \lim_{y \rightarrow \infty} G(u, y), \quad u \geq 0. \quad (8.14)$$

We have the following result.

Theorem 8.10 *The function $G(u, y)$ satisfies the equation*

$$\frac{\partial}{\partial u} G(u, y) = \frac{\lambda}{c} G(u, y) - \frac{\lambda}{c} \int_0^u G(u-x, y) dF(x) - \frac{\lambda}{c} [F(u+y) - F(u)], \quad u \geq 0. \quad (8.15)$$

Proof: Let us again consider what happens with the first claim. The time of the first claim has the exponential probability density function $\lambda e^{-\lambda t}$, and the surplus available to pay the first claim at time t is $u + ct$. If the amount of the claim is x , where $0 \leq x \leq u + ct$, then the first claim does not cause ruin but reduces the surplus to $u + ct - x$. By the stationary and independent increments

property, ruin with a deficit of at most y would then occur thereafter with probability $G(u + ct - x, y)$. The only other possibility for ruin to occur with a deficit of at most y is that the first claim does cause ruin, that is, it occurs for an amount x where $x > u + ct$ but $x \leq u + ct + y$ because, if $x > u + ct + y$, the deficit would then exceed y . The probability that the claim amount x satisfies $u + ct < x \leq u + ct + y$ is $F(u + ct + y) - F(u + ct)$. Consequently, by the law of total probability, we have

$$G(u, y) = \int_0^\infty \left[\int_0^{u+ct} G(u+ct-x, y) dF(x) + F(u+ct+y) - F(u+ct) \right] \lambda e^{-\lambda t} dt.$$

We wish to differentiate this expression with respect to u , and in order to do so, it is convenient to change the variable of integration from t to $z = u + ct$. Thus, $t = (z - u)/c$ and $dt = dz/c$. Then with this change of variable we have

$$G(u, y) = \frac{\lambda}{c} e^{(\lambda/c)u} \int_u^\infty e^{-(\lambda/c)z} \left[\int_0^z G(z-x, y) dF(x) + F(z+y) - F(z) \right] dz.$$

Recall from the fundamental theorem of calculus that, if k is a function, then $\frac{d}{du} \int_u^\infty k(z) dz = -k(u)$, and we may differentiate with the help of the product rule to obtain

$$\frac{\partial}{\partial u} G(u, y) = \frac{\lambda}{c} G(u, y) + \frac{\lambda}{c} e^{(\lambda/c)u} \left\{ -e^{-(\lambda/c)u} \left[\int_0^u G(u-x, y) dF(x) + F(u+y) - F(u) \right] \right\},$$

from which the result follows. \square

We now determine an explicit formula for $G(0, y)$.

Theorem 8.11 *The function $G(0, y)$ is given by*

$$G(0, y) = \frac{\lambda}{c} \int_0^y [1 - F(x)] dx, \quad y \geq 0. \quad (8.16)$$

Proof: First note that

$$0 \leq G(u, y) \leq \psi(u) \leq e^{-\kappa u}$$

and thus

$$0 \leq G(\infty, y) = \lim_{u \rightarrow \infty} G(u, y) \leq \lim_{u \rightarrow \infty} e^{-\kappa u} = 0,$$

and therefore $G(\infty, y) = 0$. Also,

$$\int_0^\infty G(u, y) du \leq \int_0^\infty e^{-\kappa u} du = \kappa^{-1} < \infty.$$

Thus let $\tau(y) = \int_0^\infty G(u, y) du$ and we know that $0 < \tau(y) < \infty$. Then, integrate (8.15) with respect to u from 0 to ∞ to get, using the above,

$$-G(0, y) = \frac{\lambda}{c} \tau(y) - \frac{\lambda}{c} \int_0^\infty \int_0^u G(u-x, y) dF(x) du - \frac{\lambda}{c} \int_0^\infty [F(u+y) - F(u)] du.$$

Interchanging the order of integration in the double integral yields

$$\begin{aligned} G(0, y) &= -\frac{\lambda}{c} \tau(y) + \frac{\lambda}{c} \int_0^\infty \int_x^\infty G(u-x, y) du dF(x) \\ &\quad + \frac{\lambda}{c} \int_0^\infty [F(u+y) - F(u)] du \end{aligned}$$

and changing the variable of integration from u to $v = u - x$ in the inner integral of the double integral results in

$$\begin{aligned} G(0, y) &= -\frac{\lambda}{c} \tau(y) + \frac{\lambda}{c} \int_0^\infty \int_0^\infty G(v, y) dv dF(x) \\ &\quad + \frac{\lambda}{c} \int_0^\infty [F(u+y) - F(u)] du \\ &= -\frac{\lambda}{c} \tau(y) + \frac{\lambda}{c} \int_0^\infty \tau(y) dF(x) + \frac{\lambda}{c} \int_0^\infty [F(u+y) - F(u)] du. \end{aligned}$$

Because $\int_0^\infty dF(x) = 1$, the first two terms on the right-hand side cancel, and so

$$\begin{aligned} G(0, y) &= \frac{\lambda}{c} \int_0^\infty [F(u+y) - F(u)] du \\ &= \frac{\lambda}{c} \int_0^\infty [1 - F(u)] du - \frac{\lambda}{c} \int_0^\infty [1 - F(u+y)] du. \end{aligned}$$

Then change the variable from u to $x = u$ in the first integral and from u to $x = u + y$ in the second integral. The result is

$$G(0, y) = \frac{\lambda}{c} \int_0^\infty [1 - F(x)] dx - \frac{\lambda}{c} \int_y^\infty [1 - F(x)] dx = \frac{\lambda}{c} \int_0^y [1 - F(x)] dx. \quad \square$$

We remark that (8.16) holds even if there is no adjustment coefficient. The function $G(0, y)$ is itself of considerable interest in its own right, but for now we shall return to the analysis of $\phi(u)$.

Theorem 8.12 *The survival probability with no initial reserve satisfies*

$$\phi(0) = \frac{\theta}{1 + \theta}. \quad (8.17)$$

Proof: Recall that $\mu = \int_0^\infty [1 - F(x)] dx$ and note that from (8.16)

$$\psi(0) = \lim_{y \rightarrow \infty} G(0, y) = \frac{\lambda}{c} \int_0^\infty [1 - F(x)] dx = \frac{\lambda\mu}{c} = \frac{1}{1 + \theta}.$$

Thus, $\phi(0) = 1 - \psi(0) = \theta/(1 + \theta)$. \square

The general solution to $\phi(u)$ may be obtained from the following integro-differential equation subject to the initial condition (8.17).

Theorem 8.13 *The probability of ultimate survival $\phi(u)$ satisfies*

$$\phi'(u) = \frac{\lambda}{c} \phi(u) - \frac{\lambda}{c} \int_0^u \phi(u-x) dF(x), \quad u \geq 0. \quad (8.18)$$

Proof: From (8.15) with $y \rightarrow \infty$ and (8.14),

$$\psi'(u) = \frac{\lambda}{c} \psi(u) - \frac{\lambda}{c} \int_0^u \psi(u-x) dF(x) - \frac{\lambda}{c} [1 - F(u)], \quad u \geq 0. \quad (8.19)$$

In terms of the survival probability $\phi(u) = 1 - \psi(u)$, (8.19) may be expressed as

$$\begin{aligned} -\phi'(u) &= \frac{\lambda}{c} [1 - \phi(u)] - \frac{\lambda}{c} \int_0^u [1 - \phi(u-x)] dF(x) - \frac{\lambda}{c} [1 - F(u)] \\ &= -\frac{\lambda}{c} \phi(u) - \frac{\lambda}{c} \int_0^u dF(x) + \frac{\lambda}{c} \int_0^u \phi(u-x) dF(x) + \frac{\lambda}{c} F(u) \\ &= -\frac{\lambda}{c} \phi(u) + \frac{\lambda}{c} \int_0^u \phi(u-x) dF(x) \end{aligned}$$

because $F(u) = \int_0^u dF(x)$. The result then follows. \square

It is largely a matter of taste whether one uses (8.18) or (8.19). We shall often use (8.18) because it is slightly simpler algebraically. Unfortunately, the solution for general $F(x)$ is rather complicated and we shall defer this general solution to Section 8.4. At this point we shall obtain the solution for one special choice of $F(x)$.

Example 8.14 (The exponential distribution) *Suppose, as in Example 8.4, that $F(x) = 1 - e^{-x/\mu}$, $x \geq 0$. Determine $\phi(u)$.*

In this case (8.18) becomes

$$\phi'(u) = \frac{\lambda}{c} \phi(u) - \frac{\lambda}{\mu c} \int_0^u \phi(u-x) e^{-x/\mu} dx.$$

Change variables in the integral from x to $y = u - x$ to obtain

$$\phi'(u) = \frac{\lambda}{c}\phi(u) - \frac{\lambda}{\mu c}e^{-u/\mu} \int_0^u \phi(y)e^{y/\mu} dy. \quad (8.20)$$

We wish to eliminate the integral term in (8.20) so we differentiate with respect to u . This gives

$$\phi''(u) = \frac{\lambda}{c}\phi'(u) + \frac{\lambda}{\mu^2 c}e^{-u/\mu} \int_0^u \phi(y)e^{y/\mu} dy - \frac{\lambda}{\mu c}\phi(u).$$

The integral term can be eliminated using (8.20) to produce

$$\phi''(u) = \frac{\lambda}{c}\phi'(u) - \frac{\lambda}{\mu c}\phi(u) + \frac{1}{\mu} \left[\frac{\lambda}{c}\phi(u) - \phi'(u) \right],$$

which simplifies to

$$\phi''(u) = \left(\frac{\lambda}{c} - \frac{1}{\mu} \right) \phi'(u) = -\frac{\theta}{\mu(1+\theta)} \phi'(u).$$

After multiplication by the integrating factor $e^{\theta u/[\mu(1+\theta)]}$, this may be rewritten as

$$\frac{d}{du} \left[e^{\theta u/[\mu(1+\theta)]} \phi'(u) \right] = 0.$$

Integrating with respect to u gives

$$e^{\theta u/[\mu(1+\theta)]} \phi'(u) = K_1.$$

From (8.20) with $u = 0$ and using (8.17), we thus have

$$K_1 = \phi'(0) = \frac{\lambda}{c} \frac{\theta}{1+\theta} = \frac{\lambda}{\lambda\mu(1+\theta)} \frac{\theta}{1+\theta} = \frac{\theta}{\mu(1+\theta)^2}.$$

Thus,

$$\phi'(u) = \frac{\theta}{\mu(1+\theta)^2} \exp \left[-\frac{\theta u}{\mu(1+\theta)} \right],$$

which may be integrated again to give

$$\phi(u) = -\frac{1}{1+\theta} \exp \left[-\frac{\theta u}{\mu(1+\theta)} \right] + K_2.$$

Now (8.17) gives $\phi(0) = \theta/(1+\theta)$, and so with $u = 0$ we have $K_2 = 1$. Thus

$$\phi(u) = 1 - \frac{1}{1+\theta} \exp \left[-\frac{\theta u}{\mu(1+\theta)} \right]$$

is the required probability. \square

8.3.1 Exercises

8.13 Suppose that the claim size distribution is exponential with $F(x) = 1 - e^{-x/\mu}$ as in Example 8.14.

- Prove, using (8.15), that $G(u, y) = \psi(u)F(y)$ in this case.
- Prove that the distribution of the deficit immediately after ruin occurs, given that ruin does occur, has the same exponential distribution given above.

8.14 This exercise involves the derivation of integral equations called defective renewal equations for $G(u, y)$ and $\psi(u)$. These may be used to derive various properties of these functions.

- Integrate (8.15) over u from 0 to t and use (8.16) to show that

$$\begin{aligned} G(t, y) &= \frac{\lambda}{c}\Lambda(t, y) - \frac{\lambda}{c} \int_0^t \Lambda(t-x, y) dF(x) \\ &\quad + \frac{\lambda}{c} \int_0^y [1-F(x)] dx - \frac{\lambda}{c} \int_0^t [1-F(u)] du \\ &\quad + \frac{\lambda}{c} \int_0^t [1-F(u+y)] du, \end{aligned}$$

where $\Lambda(x, y) = \int_0^x G(v, y) dv$.

- Use integration by parts on the integral $\int_0^t \Lambda(t-x, y) dF(x)$ to show from (a) that

$$\begin{aligned} G(t, y) &= \frac{\lambda}{c}\Lambda(t, y) - \frac{\lambda}{c} \int_0^t G(t-x, y) F(x) dx \\ &\quad + \frac{\lambda}{c} \int_0^{y+t} [1-F(x)] dx - \frac{\lambda}{c} \int_0^t [1-F(u)] du. \end{aligned}$$

- Prove using (b) that

$$G(u, y) = \frac{\lambda}{c} \int_0^u G(u-x, y) [1-F(x)] dx + \frac{\lambda}{c} \int_u^{y+u} [1-F(x)] dx.$$

- Prove that

$$\psi(u) = \frac{\lambda}{c} \int_0^u \psi(u-x) [1-F(x)] dx + \frac{\lambda}{c} \int_u^\infty [1-F(x)] dx.$$

8.4 THE MAXIMUM AGGREGATE LOSS

We now derive the general solution to the integrodifferential equation (8.18) subject to the boundary conditions (8.12) and (8.17).

Beginning with an initial reserve u , the probability that the surplus will ever fall below the initial level u is $\psi(0)$ because the surplus process has stationary and independent increments. Thus the probability of dropping below the initial level u is the same for all u , but we know that when $u = 0$ it is $\psi(0)$.

The key result is that, given that there is a drop below the initial level u , the random variable Y which represents the amount of this initial drop has the equilibrium probability density function $f_e(y)$, where $f_e(y)$ is given by (8.8).

Theorem 8.15 *Given that there is a drop below the initial level u , the random variable Y which represents the amount of this initial drop has probability density function $f_e(y) = [1 - F(y)]/\mu$.*

Proof: Recall the function $G(u, y)$ from Definition 8.9. Because the surplus process has stationary and independent increments, $G(0, y)$ also represents the probability that the surplus drops below its initial level, and the amount of this drop is at most y . Thus, using Theorem 8.11, the amount of the drop, given that there is a drop, has cumulative distribution function

$$\begin{aligned} \Pr(Y \leq y) &= \frac{G(0, y)}{\psi(0)} \\ &= \frac{\lambda}{c\psi(0)} \int_0^y [1 - F(u)] du \\ &= \frac{1}{\mu} \int_0^y [1 - F(u)] du \end{aligned}$$

and the result follows by differentiation. \square

If there is a drop of y , the surplus immediately after the drop is $u - y$, and because the surplus process has stationary and independent increments, ruin occurs thereafter with probability $\psi(u - y)$, provided $u - y$ is nonnegative; otherwise ruin would have already occurred. The probability of a second drop is $\psi(0)$, and the amount of the second drop also has density $f_e(y)$ and is independent of the first drop. Due to the memoryless property of the Poisson process, the process “starts over” after each drop. Therefore, the total number of drops K is geometrically distributed, that is, $\Pr(K = 0) = 1 - \psi(0)$, $\Pr(K = 1) = [1 - \psi(0)]\psi(0)$, and more generally,

$$\Pr(K = k) = [1 - \psi(0)][\psi(0)]^k = \frac{\theta}{1 + \theta} \left(\frac{1}{1 + \theta} \right)^k, \quad k = 0, 1, 2, \dots,$$

because $\psi(0) = 1/(1 + \theta)$. The usual geometric parameter β (in Appendix B) is thus $1/\theta$ in this case.

After a drop, the surplus immediately begins to increase again. Thus, the lowest level of the surplus is $u - L$, where L , called the **maximum aggregate**

loss, is the total of all the drop amounts. Let Y_j be the amount of the j th drop, and because the surplus process has stationary and independent increments, $\{Y_1, Y_2, \dots\}$ is a sequence of independent and identically distributed random variables [each with density $f_e(y)$]. Because the number of drops is K , it follows that

$$L = Y_1 + Y_2 + \dots + Y_K$$

with $L = 0$ if $K = 0$. Thus, L is a compound geometric random variable with “claim size density” $f_e(y)$.

Clearly, ultimate survival beginning with initial reserve u occurs if the maximum aggregate loss L does not exceed u , that is,

$$\phi(u) = \Pr(L \leq u), \quad u \geq 0.$$

Let $F_e^{*0}(y) = 0$ if $y < 0$ and 1 if $y \geq 0$. Also $F_e^{*k}(y) = \Pr\{Y_1 + Y_2 + \dots + Y_k \leq y\}$ is the cumulative distribution function of the k -fold convolution of the distribution of Y with itself. We then have the general solution, namely,

$$\phi(u) = \sum_{k=0}^{\infty} \frac{\theta}{1 + \theta} \left(\frac{1}{1 + \theta} \right)^k F_e^{*k}(u), \quad u \geq 0.$$

In terms of the ruin probability, this general solution may be expressed as

$$\psi(u) = \sum_{k=1}^{\infty} \frac{\theta}{1 + \theta} \left(\frac{1}{1 + \theta} \right)^k S_e^{*k}(u), \quad u \geq 0,$$

where $S_e^{*k}(y) = 1 - F_e^{*k}(y)$. Evidently, $\psi(u)$ is the survival function associated with the compound geometric random variable L , and analytic solutions may be obtained in a similar manner as to those obtained in Section 6.4. An analytic solution for the important Erlang mixture claim severity pdf³

$$f(x) = \sum_{k=1}^r q_k \frac{\beta^{-k} x^{k-1} e^{-x/\beta}}{(k-1)!},$$

where the q_k are positive weights that sum to 1, may be found in Exercise 8.17, and for some other claim severity distributions in the next section.

We may also compute ruin probabilities numerically by computing the cumulative distribution function of a compound geometric distribution using any of the techniques described in Chapter 6.

Example 8.16 *Suppose the individual loss distribution is Pareto with $\alpha = 3$ and a mean of 500. Let the security loading be $\theta = 0.2$. Determine $\phi(u)$ for $u = 100, 200, 300, \dots$*

³Any continuous positive probability density function may be approximated arbitrarily accurately by an Erlang mixture pdf, as shown by Tijms [129], p. 163. An Erlang distribution is a gamma distribution for which the shape parameter α must be an integer.

Table 8.1 Survival probabilities, Pareto losses

u	$\phi(u)$	u	$\phi(u)$
100	0.193	5,000	0.687
200	0.216	7,500	0.787
300	0.238	10,000	0.852
500	0.276	15,000	0.923
1,000	0.355	20,000	0.958
2,000	0.473	25,000	0.975
3,000	0.561		

We first require the cdf, $F_e(u)$. It can be found from its pdf

$$\begin{aligned}
 f_e(u) &= \frac{1 - F(u)}{\mu} = \frac{1 - \left[1 - \left(\frac{1,000}{1,000 + u} \right)^3 \right]}{500} \\
 &= \frac{1}{500} \left(\frac{1,000}{1,000 + u} \right)^3,
 \end{aligned}$$

which happens to be the density function of a Pareto distribution with $\alpha = 2$ and a mean of 1,000. This new Pareto distribution is the severity distribution for a compound geometric distribution where the parameter is $\beta = 1/\theta = 5$. The compound geometric distribution can be evaluated using any of the techniques in Chapter 6. We used the recursive formula with a discretization which preserves the mean and a span of $h = 5$. The cumulative probabilities are then obtained by summing the discrete probabilities generated by the recursive formula. The values appear in Table 8.1. \square

8.4.1 Exercises

8.15 Suppose the number of claims follows the Poisson process and the amount of an individual claim is exponentially distributed with mean 100. The relative security loading is $\theta = 0.1$. Determine $\psi(1,000)$ by using the method of this section. Use the method of rounding with a span of 50 to discretize the exponential distribution. Compare your answer to the exact ruin probability (see Example 8.14).

8.16 Consider the problem of Example 8.5 with $\beta = 50$. Use the method of this section (with discretization by rounding and a span of 1) to approximate $\psi(200)$. Compare your answer to the exact ruin probability which may be found in Example 8.19.

8.17 Suppose that the claim severity pdf is given by

$$f(x) = \sum_{k=1}^r q_k \frac{\beta^{-k} x^{k-1} e^{-x/\beta}}{(k-1)!}, \quad x > 0,$$

where $\sum_{k=1}^r q_k = 1$. Note that this is a mixture of gamma densities.

(a) Show that

$$f_e(x) = \sum_{k=1}^r q_k^* \frac{\beta^{-k} x^{k-1} e^{-x/\beta}}{(k-1)!}, \quad x > 0,$$

where

$$q_k^* = \frac{\sum_{j=k}^r q_j}{\sum_{j=1}^r j q_j}, \quad k = 1, 2, \dots, r,$$

and also show that $\sum_{k=1}^r q_k^* = 1$.

(b) Define

$$Q^*(z) = \sum_{k=1}^r q_k^* z^k$$

and use the results of Exercise 6.35 to show that

$$\psi(u) = \sum_{n=1}^{\infty} c_n \sum_{j=0}^{n-1} \frac{(u/\beta)^j e^{-u/\beta}}{j!}, \quad u \geq 0,$$

where

$$C(z) = \left\{ 1 - \frac{1}{\theta} [Q^*(z) - 1] \right\}^{-1}$$

is a compound geometric pgf, with probabilities which may be computed recursively by $c_0 = \theta(1 + \theta)^{-1}$ and

$$c_k = \frac{1}{1 + \theta} \sum_{j=1}^k q_j^* c_{k-j}, \quad k = 1, 2, \dots,$$

where $q_j^* = 0$ if $j \neq 1, 2, \dots, r$.

(c) Use (b) to show that

$$\psi(u) = e^{-u/\beta} \sum_{j=0}^{\infty} \bar{C}_j \frac{(u/\beta)^j}{j!}, \quad u \geq 0,$$

where $\bar{C}_j = \sum_{k=j+1}^{\infty} c_k$, $j = 0, 1, \dots$. Then use (b) to show that the \bar{C}_n s may be computed recursively from

$$\bar{C}_n = \frac{1}{1+\theta} \sum_{k=1}^n q_k^* \bar{C}_{n-k} + \frac{1}{1+\theta} \sum_{k=n+1}^{\infty} q_k^*, \quad n = 1, 2, \dots,$$

beginning with $\bar{C}_0 = (1+\theta)^{-1}$.

8.18 (a) Using Exercise 8.14(c) prove that

$$G(u, y) = \frac{1}{1+\theta} \int_0^u G(u-x, y) f_e(x) dx + \frac{1}{1+\theta} \int_u^{y+u} f_e(x) dx,$$

where $G(u, y)$ is defined in Section 8.3, and using Exercise 8.14(d) that

$$\psi(u) = \frac{1}{1+\theta} \int_0^u \psi(u-x) f_e(x) dx + \frac{1}{1+\theta} \int_u^{\infty} f_e(x) dx,$$

where $f_e(x)$ is given by (8.8).

(b) Prove the results in (a) directly by using probabilistic arguments.

Hint: Condition on the amount of the first drop in surplus and use the law of total probability.

8.5 CRAMER'S ASYMPTOTIC RUIN FORMULA AND TIJMS' APPROXIMATION

There is another very useful piece of information regarding the ruin probability which involves the adjustment coefficient κ . The following theorem gives a result known as Cramér's asymptotic ruin formula. The notation $a(x) \sim b(x)$, $x \rightarrow \infty$, means $\lim_{x \rightarrow \infty} a(x)/b(x) = 1$.

Theorem 8.17 Suppose $\kappa > 0$ satisfies (8.3). Then the ruin probability satisfies

$$\psi(u) \sim C e^{-\kappa u}, \quad u \rightarrow \infty, \quad (8.21)$$

where

$$C = \frac{\mu\theta}{M'_X(\kappa) - \mu(1+\theta)}, \quad (8.22)$$

and $M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} dF(x)$ is the moment generating function of the claim severity random variable X .

Proof: The proof of this result is complicated and utilizes the key renewal theorem together with a defective renewal equation for $\psi(u)$ which may be found in Exercise 8.14(d), or equivalently in Exercise 8.18(a). The interested reader should see Rolski et. al. [113], Sec. 5.4.2 for details. \square

Thus, in addition to Lundberg's inequality given by Theorem 8.8, the ruin probability behaves like an exponential function for large u . Note that, for

Lundberg's inequality (8.9) to hold, it must be the case that C given by (8.22) must satisfy $C \leq 1$. Also, although (8.21) is an asymptotic approximation, it is known to be quite accurate even for u which is not too large (particularly if the relative security loading θ is itself not too large). Before continuing, let us consider an important special case.

Example 8.18 (The exponential distribution) If $F(x) = 1 - e^{-x/\mu}$, $x \geq 0$, determine the asymptotic ruin formula.

We found in Example 8.4 that the adjustment coefficient was given by $\kappa = \theta/[\mu(1+\theta)]$ and $M_X(t) = (1 - \mu t)^{-1}$. Thus,

$$M'_X(t) = \frac{d}{dt}(1 - \mu t)^{-1} = \mu(1 - \mu t)^{-2}.$$

Also,

$$M'_X(\kappa) = \mu(1 - \mu\kappa)^{-2} = \mu[1 - \theta(1+\theta)^{-1}]^{-2} = \mu(1+\theta)^2.$$

Thus, from (8.22),

$$C = \frac{\mu\theta}{\mu(1+\theta)^2 - \mu(1+\theta)} = \frac{\theta}{(1+\theta)(1+\theta-1)} = \frac{1}{1+\theta}.$$

The asymptotic formula (8.21) becomes

$$\psi(u) \sim \frac{1}{1+\theta} \exp\left[-\frac{\theta u}{\mu(1+\theta)}\right], \quad u \rightarrow \infty.$$

This is the exact ruin probability as was demonstrated in Example 8.14. \square

In cases other than when $F(x)$ is the exponential distribution, the exact solution for $\psi(u)$ is more complicated (including in particular the general compound geometric solution given in Section 8.4). A simple analytic approximation was suggested by Tijms [129], pp. 271-272 to take advantage of the accuracy for large u of Cramér's asymptotic ruin formula given in Theorem 8.17. The idea is to add an exponential term to (8.21) to improve the accuracy for small u as well. Thus, the Tijms approximation is defined as

$$\psi_T(u) = \left(\frac{1}{1+\theta} - C\right) e^{-u/\alpha} + C e^{-\kappa u}, \quad u \geq 0, \quad (8.23)$$

where α is chosen so that the approximation also matches the compound geometric mean of the maximum aggregate loss. As shown in Section 4.3.3, the mean of the amount of the drop in surplus (in the terminology of Section 8.4) is $E(Y) = \int_0^{\infty} y f_e(y) dy = E(X^2)/(2\mu)$, where $\mu = E(X)$ and X is a generic claim severity random variable. Similarly, the number of drops in surplus K is geometrically distributed with parameter $1/\theta$, so from Appendix

But we have $E(K) = 1/\theta$. Because the maximum aggregate loss is the compound geometric random variable L , it follows from (6.6) that its mean is

$$E(L) = E(K)E(Y) = \frac{E(X^2)}{2\mu\theta}.$$

But $\psi(u) = \Pr(L > u)$, and from (3.9) on page 32 with $k = 1$ and $u = \infty$ we have $E(L) = \int_0^\infty \psi(u)du$. Therefore, in order that the Tijms approximation match the mean, we need to replace $\psi(u)$ by $\psi_T(u)$ in the integral. Thus from (8.23)

$$\begin{aligned} \int_0^\infty \psi_T(u)du &= \left(\frac{1}{1+\theta} - C \right) \int_0^\infty e^{-u/\alpha} du + C \int_0^\infty e^{-\kappa u} du \\ &= \alpha \left(\frac{1}{1+\theta} - C \right) + \frac{C}{\kappa}, \end{aligned}$$

and equating this to $E(L)$ yields

$$\alpha \left(\frac{1}{1+\theta} - C \right) + \frac{C}{\kappa} = \frac{E(X^2)}{2\mu\theta},$$

which may be solved for α to give

$$\alpha = \frac{E(X^2)/(2\mu\theta) - C/\kappa}{1/(1+\theta) - C}. \quad (8.24)$$

To summarize, Tijms' approximation to the ruin probability is given by (8.23), with α given by (8.24).

In addition to providing a simple analytic approximation which is of good quality, Tijms' approximation $\psi_T(u)$ has the added benefit of exactly reproducing the true value of $\psi(u)$ in some cases. (Some insight into this phenomenon is provided by Exercise 8.21.) In particular, it can be shown that $\psi_T(u) = \psi(u)$ if the claim size pdf is of the form $f(x) = p(\beta^{-1}e^{-x/\beta}) + (1-p)(\beta^{-2}xe^{-x/\beta})$, $x \geq 0$, with $0 \leq p < 1$ (of course, if $p = 1$, this is the exponential density for which Tijms' approximation is not used). We have the following example.

Example 8.19 (A gamma distribution with a shape parameter⁴ of 2) As in Example 8.5, suppose that $\theta = 2$, and the single claim size density is $f(x) = \beta^{-2}xe^{-x/\beta}$, $x \geq 0$. Determine the Tijms approximation to the ruin probability.

The moment generating function is $M_X(t) = (1 - \beta t)^{-2}$, $t < 1/\beta$, from which one finds that $M'_X(t) = 2\beta(1 - \beta t)^{-3}$, and $\mu = M'_X(0) = 2\beta$. As shown

in Example 8.5, the adjustment coefficient $\kappa > 0$ satisfies $1 + (1 + \theta)\kappa\mu = M_X(\kappa)$, which in this example becomes $1 + 6\beta\kappa = (1 - \beta\kappa)^{-2}$ and is given by $\kappa = 1/(2\beta)$. We will first compute Cramér's asymptotic ruin formula. We have $M'_X(\kappa) = M'_X[1/(2\beta)] = 2\beta(1 - \frac{1}{2})^{-3} = 16\beta$. Thus, (8.22) yields

$$C = \frac{(2\beta)(2)}{16\beta - (2\beta)(1+2)} = \frac{2}{5},$$

and from (8.21), $\psi(u) \sim \frac{2}{5}e^{-u/(2\beta)}$, $u \rightarrow \infty$. We next turn to Tijms' approximation given by (8.23), which becomes in this case

$$\psi_T(u) = \left(\frac{1}{1+2} - \frac{2}{5} \right) e^{-u/\alpha} + \frac{2}{5} e^{-u/(2\beta)} = \frac{2}{5} e^{-u/(2\beta)} - \frac{1}{15} e^{-u/\alpha}.$$

It remains to compute α . We have $M''_X(t) = 6\beta^2(1 - \beta t)^{-4}$, from which it follows that $E(X^2) = M''_X(0) = 6\beta^2$. The amount of the drop in surplus has mean $E(Y) = E(X^2)/(2\mu) = 6\beta^2/(4\beta) = 3\beta/2$. Because the number of drops has mean $E(K) = 1/\theta = \frac{1}{2}$, the maximum aggregate loss has mean $E(L) = E(K)E(Y) = 3\beta/4$, and α must satisfy $E(L) = \int_0^\infty \psi_T(u)du$, or equivalently (8.24). That is, α is given by⁵

$$\alpha = \frac{\frac{3\beta}{4} - \frac{2}{5}(2\beta)}{\frac{1}{1+2} - \frac{2}{5}} = \frac{3\beta}{4}.$$

Tijms' approximation thus becomes

$$\psi_T(u) = \frac{2}{5}e^{-u/(2\beta)} - \frac{1}{15}e^{-4u/(3\beta)}, \quad u \geq 0.$$

As mentioned above, $\psi(u) = \psi_T(u)$ in this case. □

Another class of claim severity distributions for which the Tijms approximation is exactly equal to the true ruin probability is that, with probability density function of the form $f(x) = p(\beta_1 e^{-\beta_1 x}) + (1-p)(\beta_2 e^{-\beta_2 x})$, $x \geq 0$. If $0 < p < 1$, then this distribution is a mixture of two exponentials, whereas if, $p = \beta_2/(\beta_2 - \beta_1)$ then the distribution, referred to as a combination of two exponentials, is that of the sum of two independent exponential random variables with means β_1 and β_2 , where $\beta_1 \neq \beta_2$. The next example illustrates these ideas.

⁵It is actually not a coincidence that $1/\alpha$ is the other root of the adjustment coefficient equation, as may be seen from Example 8.5. It is instructive to compute α in this manner, however, because this approach is applicable in general for arbitrary claim size distributions, including those in which Tijms' approximation does not exactly reproduce the true ruin probability.

⁴For a gamma distribution, the shape parameter is the one denoted by α in Appendix A and is not to be confused with the value of α in the Tijms approximation.

Example 8.20 (A mixture of exponential distributions) Suppose that $\theta = \frac{4}{11}$ and the single claim size density is $f(x) = e^{-3x} + 10e^{-5x}/3$, $x \geq 0$. Determine the Tijms approximation to the ruin probability.

First we note that the moment generating function is

$$M_X(t) = \int_0^\infty e^{tx} f(x) dx = (3-t)^{-1} + \frac{10}{3}(5-t)^{-1}.$$

Thus, $M'_X(t) = (3-t)^{-2} + \frac{10}{3}(5-t)^{-2}$, from which it follows that $\mu = M'_X(0) = \frac{1}{9} + \frac{10}{75} = \frac{11}{45}$. Equation (8.3) then implies that the adjustment coefficient $\kappa > 0$ satisfies $1 + \frac{1}{3}\kappa = (3-\kappa)^{-1} + \frac{10}{3}(5-\kappa)^{-1}$. Multiplication by $3(3-\kappa)(5-\kappa)$ yields

$$3(\kappa-3)(\kappa-5) + \kappa(\kappa-3)(\kappa-5) = 3(5-\kappa) + 10(3-\kappa).$$

That is,

$$3(\kappa^2 - 8\kappa + 15) + \kappa^3 - 8\kappa^2 + 15\kappa = 45 - 13\kappa.$$

Rearrangement yields

$$0 = \kappa^3 - 5\kappa^2 + 4\kappa = \kappa(\kappa-1)(\kappa-4),$$

and $\kappa = 1$ because it is the smallest positive root.

Next, we determine Cramér's asymptotic formula. Equation (8.22) becomes, with $M'_X(\kappa) = M'_X(1) = \frac{1}{4} + \frac{10}{3} \frac{1}{16} = \frac{11}{24}$,

$$C = \frac{\left(\frac{11}{45}\right)\left(\frac{4}{11}\right)}{\frac{11}{24} - \left(\frac{11}{45}\right)\left(\frac{15}{11}\right)} = \frac{32}{45},$$

and thus Cramér's asymptotic formula is $\psi(u) \sim \frac{32}{45}e^{-u}$, $u \rightarrow \infty$.

Equation (8.23) then becomes

$$\psi_T(u) = \left(\frac{1}{1+\frac{4}{11}} - \frac{32}{45}\right)e^{-u/\alpha} + \frac{32}{45}e^{-u} = \frac{1}{45}e^{-u/\alpha} + \frac{32}{45}e^{-u}.$$

In order to compute α , we note that $M''_X(t) = 2(3-t)^{-3} + \frac{20}{3}(5-t)^{-3}$, and thus $E(X^2) = M''_X(0) = \frac{2}{27} + \frac{20}{3}\left(\frac{1}{125}\right) = \frac{86}{675}$. The mean of the maximum aggregate loss is therefore

$$E(L) = \frac{E(X^2)}{2\mu\theta} = \frac{\frac{86}{675}}{2\left(\frac{11}{45}\right)\left(\frac{4}{11}\right)} = \frac{43}{60}.$$

Equation (8.24) then yields

$$\alpha = \frac{\frac{43}{60} - \frac{32}{45}}{\frac{1}{1+\frac{4}{11}} - \frac{32}{45}} = \frac{1}{4},$$

and so Tijms' approximation becomes $\psi_T(u) = \frac{1}{45}e^{-4u} + \frac{32}{45}e^{-u}$. As mentioned above, $\psi(u) = \psi_T(u)$ in this case also. \square

It is not hard to see from (8.23) that $\psi_T(u) \sim Ce^{-\kappa u}$, $u \rightarrow \infty$, if $\kappa < 1/\alpha$. In this situation, $\psi_T(u)$ will equal $\psi(u)$ when $u = 0$ and when $u \rightarrow \infty$ as well as matching the compound geometric mean. It can be shown that a sufficient condition for the asymptotic agreement between $\psi_T(u)$ and $\psi(u)$ to hold as $u \rightarrow \infty$ is that the nonexponential claim size cumulative distribution function $F(x)$ has either a nondecreasing or nonincreasing mean residual life function [which is implied if $F(x)$ has a nonincreasing or nondecreasing hazard rate, as discussed in Section 4.3.3]. It is also interesting to note that $\psi_T(x) > Ce^{-\kappa x}$ in the former case and $\psi_T(x) < Ce^{-\kappa x}$ in the latter case. See Willmot [137] for proofs of these facts.

The following example illustrates the accuracy of Cramér's asymptotic formula and Tijms' approximation.

Example 8.21 (A gamma distribution with a shape parameter of 3) Suppose the claim severity distribution is a gamma distribution with a mean of 1 and density given by $f(x) = 27x^2e^{-3x}/2$, $x \geq 0$. Determine the exact ruin probability, Cramér's asymptotic ruin formula, and Tijms' approximation when the relative security loading θ in each is 0.25, 1, and 4, and the initial surplus u is 0.10, 0.25, 0.50, 0.75, and 1.

The moment generating function is $M_X(t) = (1-t/3)^{-3}$.

The exact values of $\psi(u)$ may be obtained using the algorithm presented in Exercise 8.17. That is, $\psi(u) = e^{-3u} \sum_{j=0}^{\infty} \bar{C}_j (3u)^j / j!$, $u \geq 0$, where the \bar{C}_j s may be computed recursively using

$$\bar{C}_j = \frac{1}{1+\theta} \sum_{k=1}^j q_k^* \bar{C}_{j-k} + \frac{1}{1+\theta} \sum_{k=j+1}^{\infty} q_k^*, \quad j = 1, 2, 3, \dots,$$

with $\bar{C}_0 = 1/(1+\theta)$, $q_1^* = q_2^* = q_3^* = \frac{1}{3}$, and $q_k^* = 0$ otherwise. The required values are listed in Table 8.2 under the heading titled Exact.

Cramér's asymptotic ruin probabilities are given by the approximation (8.21), with κ obtained from (8.3) numerically for each value of θ using the Newton-Raphson approach described in Section 8.2.1. The coefficient C is then obtained from (8.22). The required values are listed in Table 8.2 under the heading titled Cramér.

Tijms' approximation is obtained using (8.23) with α satisfying (8.24), and the values are listed in Table 8.2 under the heading titled 'Tijms.'

The values in the table, which may also be found in Tijms [129], p. 272 and Willmot [137], demonstrate that Tijms' approximation is an accurate approximation to the true value in this situation, particularly for small θ . Cramér's asymptotic formula is also remarkably accurate for small θ and u . Because this gamma distribution has an increasing hazard rate (as discussed

Table 8.2 Ruin probabilities with gamma losses

θ	u	Exact	Cramér	Tijms
0.25	0.10	0.7834	0.8076	0.7844
	0.25	0.7562	0.7708	0.7571
	0.50	0.7074	0.7131	0.7074
	0.75	0.6577	0.6597	0.6573
	1.00	0.6097	0.6103	0.6093
1.00	0.10	0.4744	0.5332	0.4764
	0.25	0.4342	0.4700	0.4361
	0.50	0.3664	0.3809	0.3665
	0.75	0.3033	0.3088	0.3026
	1.00	0.2484	0.2502	0.2476
4.00	0.10	0.1839	0.2654	0.1859
	0.25	0.1594	0.2106	0.1615
	0.50	0.1209	0.1432	0.1212
	0.75	0.0882	0.0974	0.0875
	1.00	0.0626	0.0663	0.0618

in Example 4.17), Tijms' approximate ruin probabilities are guaranteed to be smaller than Cramér's asymptotic ruin probabilities, and this may be seen to be true from the table. It also follows that the exact values, Cramér's asymptotic values, and Tijms' approximate values all must converge as $u \rightarrow \infty$, but the agreement can be seen to be fairly close even for $u = 1$. \square

8.5.1 Exercises

8.19 Show that (8.22) may be reexpressed as

$$C = \frac{\theta}{\kappa E(Y e^{\kappa Y})},$$

where Y has pdf $f_e(y)$. Hence prove for the problem of Exercise 8.17 that

$$\psi(u) \sim \frac{\theta}{\kappa \beta \sum_{j=1}^r j q_j^* (1 - \beta \kappa)^{-j-1}} e^{-\kappa u}, \quad u \rightarrow \infty,$$

where $\kappa > 0$ satisfies

$$1 + \theta = Q^*[(1 - \beta \kappa)^{-1}] = \sum_{j=1}^r q_j^* (1 - \beta \kappa)^{-j}.$$

8.20 Recall the function $G(u, y)$ defined in Section 8.3. It can be shown using the result of Exercise 8.14(c) that Cramér's asymptotic ruin formula may be generalized to

$$G(u, y) \sim C(y) e^{-\kappa u}, \quad u \rightarrow \infty,$$

where

$$C(y) = \frac{\mu \kappa \int_0^\infty e^{\kappa t} \int_y^{t+y} f_e(x) dx dt}{M_X'(\kappa) - \mu(1 + \theta)}.$$

- Demonstrate that Cramér's asymptotic ruin formula is recovered as $y \rightarrow \infty$.
- Demonstrate using Exercise 8.13 that the above asymptotic formula for $G(u, y)$ is an equality for all u in the exponential claims case with $F(x) = 1 - e^{-x/\mu}$.

8.21 Suppose that the following formula for the ruin probability is known to hold:

$$\psi(u) = C_1 e^{-r_1 u} + C_2 e^{-r_2 u}, \quad u \geq 0,$$

where $C_1 \neq 0, C_2 \neq 0$, and (without loss of generality) $0 < r_1 < r_2$.

- Determine the relative security loading θ .
- Determine the adjustment coefficient κ .
- Prove that $0 < C_1 \leq 1$.
- Determine Cramér's asymptotic ruin formula.
- Prove that $\psi_T(u) = \psi(u)$, where $\psi_T(u)$ is Tijms' approximation to the ruin probability.

8.22 Suppose that $\theta = \frac{4}{5}$ and the claim size density is given by $f(x) = (1 + 6x)e^{-3x}$, $x \geq 0$.

- Determine Cramér's asymptotic ruin formula.
- Determine the ruin probability $\psi(u)$.

8.23 Suppose that $\theta = \frac{3}{11}$ and the claim size density is given by $f(x) = 2e^{-4x} + \frac{7}{2}e^{-7x}$, $x \geq 0$.

- Determine Cramér's asymptotic ruin formula.
- Determine the ruin probability $\psi(u)$.

8.24 Suppose that $\theta = \frac{3}{5}$ and the claim size density is given by $f(x) = 3e^{-4x} + \frac{1}{2}e^{-2x}$, $x \geq 0$.

- Determine Cramér's asymptotic ruin formula.
- Determine the ruin probability $\psi(u)$.

8.25 Suppose that $\theta = \frac{7}{5}$ and the claim size density is the convolution of two exponential distributions given by $f(x) = \int_0^x 3e^{-3(x-y)} 2e^{-2y} dy$, $x \geq 0$.

- Determine Cramér's asymptotic ruin formula.
- Determine the ruin probability $\psi(u)$.

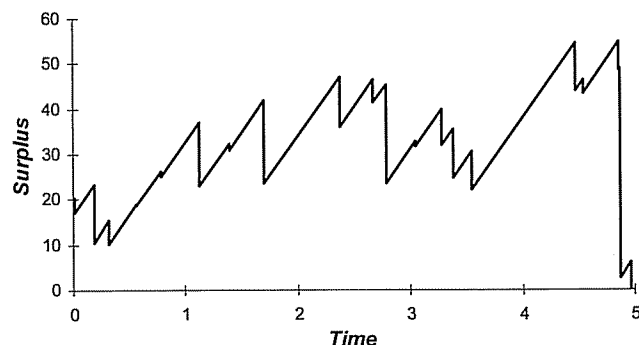


Fig. 8.2 Sample path for a Poisson surplus process.

8.6 THE BROWNIAN MOTION RISK PROCESS

In this section, we study the relationship between Brownian motion (the Wiener process) and the surplus process $\{U_t; t \geq 0\}$, where

$$U_t = u + ct - S_t, \quad t \geq 0 \quad (8.25)$$

and $\{S_t; t \geq 0\}$ is the total loss process defined by

$$S_t = X_1 + X_2 + \cdots + X_{N_t}, \quad t \geq 0,$$

where $\{N_t; t \geq 0\}$ is a Poisson process with rate λ and $S_t = 0$ when $N_t = 0$. As earlier in this chapter we assume that the individual losses $\{X_1, X_2, \dots\}$ are independently and identically distributed positive random variables whose moment generating function exists. The surplus process $\{U_t; t \geq 0\}$ increases continuously with slope c , the premium rate per unit time, and has successive downward jumps of $\{X_1, X_2, \dots\}$ at random jump times $\{T_1, T_2, \dots\}$, as illustrated by Figure 8.2. In that figure, $u = 20$, $c = 35$, $\lambda = 3$, and X has an exponential distribution with mean 10.

Let

$$Z_t = U_t - u = ct - S_t, \quad t \geq 0. \quad (8.26)$$

Then $Z_0 = 0$. Because S_t has a compound distribution, the process $\{Z_t; t \geq 0\}$ has mean

$$\begin{aligned} E(Z_t) &= ct - E(S_t) \\ &= ct - \lambda t E(X) \end{aligned}$$

and variance

$$\text{Var}(Z_t) = \lambda t E(X^2).$$

We now introduce the corresponding stochastic process based on Brownian motion.

Definition 8.22 A continuous-time stochastic process $\{W_t; t \geq 0\}$ is a *Brownian motion process* if

1. $W_0 = 0$;
2. $\{W_t, t \geq 0\}$ has stationary and independent increments; and
3. for every $t > 0$, W_t is normally distributed with mean 0 and variance $\sigma^2 t$.

The Brownian motion process, also called the **Wiener process** or **white noise**, has been used extensively in describing various physical phenomena. When $\sigma^2 = 1$, it is called **standard Brownian motion**. The English botanist Robert Brown discovered the process in 1827 and used it to describe the continuous irregular motion of a particle immersed in a liquid or gas. In 1905 Albert Einstein explained this motion by postulating perpetual collision of the particle with the surrounding medium. Norbert Wiener provided the analytical description of the process in a series of papers beginning in 1918. Since then it has been used in many areas of application from quantum mechanics to describing price levels on the stock market. It has become the key model underpinning modern financial theory.

Definition 8.23 A continuous-time stochastic process $\{W_t; t \geq 0\}$ is called a *Brownian motion with drift process* if it satisfies the properties of a Brownian motion process except that W_t has mean μt rather than 0, for some $\mu > 0$.

A Brownian motion with drift is illustrated in Figure 8.3. This process has $u = 20$, $\mu = 5$, and $\sigma^2 = 600$. The illustrated process has an initial surplus of 20, so the mean of W_t is $20 + 5t$. Technically, $W_t - 20$ is a Brownian motion with drift process.

We now show how the surplus process (8.26) based on the compound Poisson risk process is related to the Brownian motion with drift process. We will take a limit of the process (8.26) as the expected number of downward jumps becomes large and simultaneously the size of the jumps becomes small. Because the Brownian motion with drift process is characterized by the infinitesimal mean μ and infinitesimal variance σ^2 , we force the mean and variance functions to be the same for the two processes. In this way, the Brownian motion with drift can be thought of as an approximation to the compound Poisson-based surplus process. Similarly, the compound Poisson process can be used as an approximation for Brownian motion.

Let

$$\mu = c - \lambda E[X]$$

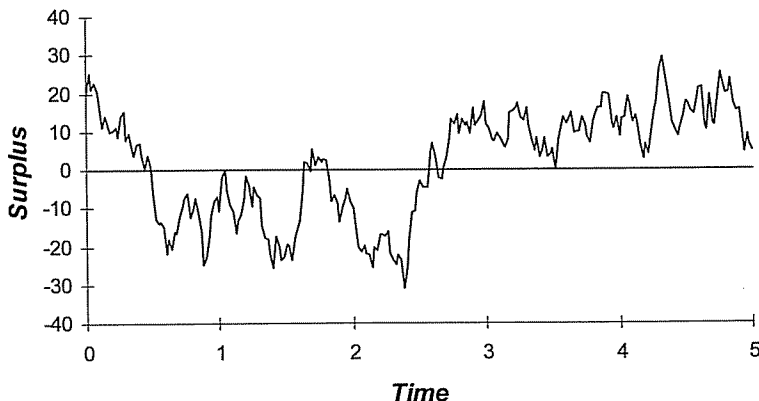


Fig. 8.3 Sample path for a Brownian motion with drift.

and

$$\sigma^2 = \lambda E[X^2]$$

denote the infinitesimal mean and variance of the Brownian motion with drift process. Then

$$\lambda = \frac{\sigma^2}{E[X^2]} \quad (8.27)$$

and

$$c = \mu + \sigma^2 \frac{E[X]}{E[X^2]}. \quad (8.28)$$

Now, in order to take limits, we can treat the jump size X as a scaled version of some other random variable Y , so that $X = \alpha Y$, where Y has fixed mean and variance. Then

$$\lambda = \frac{\sigma^2}{E[Y^2]} \cdot \frac{1}{\alpha^2}$$

and

$$c = \mu + \sigma^2 \frac{E[Y]}{E[Y^2]} \cdot \frac{1}{\alpha}.$$

Then, in order for $\lambda \rightarrow \infty$, we let $\alpha \rightarrow 0$.

Because the process $\{S_t; t \geq 0\}$ is a continuous-time process with stationary and independent increments, so are the processes $\{U_t; t \geq 0\}$ and $\{Z_t; t \geq 0\}$. This will then also be the case for the limiting process. Because $Z_0 = 0$, we only need to establish that for every t , in the limit, Z_t is normally distributed with mean μt and variance $\sigma^2 t$ according to Definitions 8.22 and

8.23. We do this by looking at the moment generating function of Z_t .

$$\begin{aligned} M_{Z_t}(r) &= M_{ct-S_t}(r) \\ &= E\{\exp[r(ct - S_t)]\} \\ &= \exp(t\{rc + \lambda[M_X(-r) - 1]\}). \end{aligned}$$

Then

$$\begin{aligned} \frac{\ln M_{Z_t}(r)}{t} &= rc + \lambda[M_X(-r) - 1] \\ &= r[\mu + \lambda E(X)] \\ &\quad + \lambda \left[1 - rE(X) + \frac{r^2}{2!}E(X^2) - \frac{r^3}{3!}E(X^3) + \cdots - 1 \right] \\ &= r\mu + \frac{r^2}{2}\lambda E(X^2) - \lambda \left[\frac{r^3}{3!}E(X^3) - \frac{r^4}{4!}E(X^4) + \cdots \right] \\ &= r\mu + \frac{r^2}{2}\sigma^2 - \lambda\alpha^2 \left[\alpha \frac{r^3}{3!}E(Y^3) - \alpha^2 \frac{r^4}{4!}E(Y^4) + \cdots \right] \\ &= r\mu + \frac{r^2}{2}\sigma^2 - \sigma^2 \left[\alpha \frac{r^3}{3!} \frac{E(Y^3)}{E(Y^2)} - \alpha^2 \frac{r^4}{4!} \frac{E(Y^4)}{E(Y^2)} + \cdots \right]. \end{aligned}$$

Because all terms except α are fixed, as $\alpha \rightarrow 0$, we have

$$\lim_{\alpha \rightarrow 0} M_{Z_t}(r) = \exp\left(r\mu t + \frac{r^2}{2}\sigma^2 t\right),$$

which is the mgf of the normal distribution with mean μt and $\sigma^2 t$. This establishes that the limiting process is Brownian motion with drift.

From Figure 8.2, it is clear that the process U_t is differentiable everywhere except at jump points. As the number of jump points increases indefinitely, the process becomes **nowhere differentiable**. Another property of a Brownian motion process is that **its paths are continuous functions of t with probability 1**. Intuitively, this occurs because the jump sizes become small as $\alpha \rightarrow 0$.

Finally, the total distance traveled in $(0, t]$ by the process U_t is

$$\begin{aligned} D &= ct + S_t \\ &= ct + X_1 + \cdots + X_{N_t}, \end{aligned}$$

which has expected value

$$\begin{aligned} E[D] &= ct + \lambda t E[X] \\ &= t \left[\mu + \sigma^2 \frac{E(Y)}{E(Y^2)} \frac{1}{\alpha} + \sigma^2 \frac{E(Y)}{E(Y^2)} \frac{1}{\alpha} \right] \\ &= t \left[\mu + 2\sigma^2 \frac{E(Y)}{E(Y^2)} \frac{1}{\alpha} \right]. \end{aligned}$$

This quantity becomes indefinitely large as $\alpha \rightarrow 0$. Hence, we have

$$\lim_{\alpha \rightarrow 0} E[D] = \infty.$$

This means that the expected distance traveled in a finite time interval is infinitely large! For a more rigorous discussion of the properties of the Brownian motion process, the text by Karlin and Taylor [71], Ch. 7 is recommended.

Because $Z_t = U_t - u$, we can just add u to the Brownian motion with drift process and then use (8.27) and (8.28) to develop an approximation for the process (8.26). Of course, the larger the value of λ and the smaller the jumps, the better will be the approximation. For a very large block of insurance policies (for example, for an entire company), this may be appropriate. In this case, the probability of ultimate ruin and the distribution of time until ruin are easily obtained from the approximating Brownian motion with drift process. This is done in the next section. Similarly, if a process is known to be Brownian motion with drift, a compound Poisson surplus process can be used as an approximation.

8.7 BROWNIAN MOTION AND THE PROBABILITY OF RUIN

Let $\{W_t; t \geq 0\}$ denote the Brownian motion with drift process with mean function μt and variance function $\sigma^2 t$. Let $U_t = u + W_t$ denote the Brownian motion with drift process with initial surplus $U_0 = u$.

We consider the probability of ruin in a finite time interval $(0, \tau)$ as well as the distribution of time until ruin if ruin occurs. Let $T = \min_{t > 0} \{t : U_t < 0\}$ be the time at which ruin occurs (with $T = \infty$ if ruin does not occur). Letting $\tau \rightarrow \infty$ will give ultimate ruin probabilities.

The probability of ruin before time τ can be expressed as

$$\begin{aligned} \psi(u, \tau) &= 1 - \phi(u, \tau) \\ &= \Pr\{T < \tau\} \\ &= \Pr\left\{\min_{0 < t < \tau} U_t < 0\right\} \\ &= \Pr\left\{\min_{0 < t < \tau} W_t < -U_0\right\} \\ &= \Pr\left\{\min_{0 < t < \tau} W_t < -u\right\}. \end{aligned}$$

Theorem 8.24 For the process U_t described above, the ruin probability is given by

$$\psi(u, \tau) = \Phi\left(-\frac{u + \mu\tau}{\sqrt{\sigma^2\tau}}\right) + \exp\left(-\frac{2\mu}{\sigma^2}u\right) \Phi\left(-\frac{u - \mu\tau}{\sqrt{\sigma^2\tau}}\right), \quad (8.29)$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

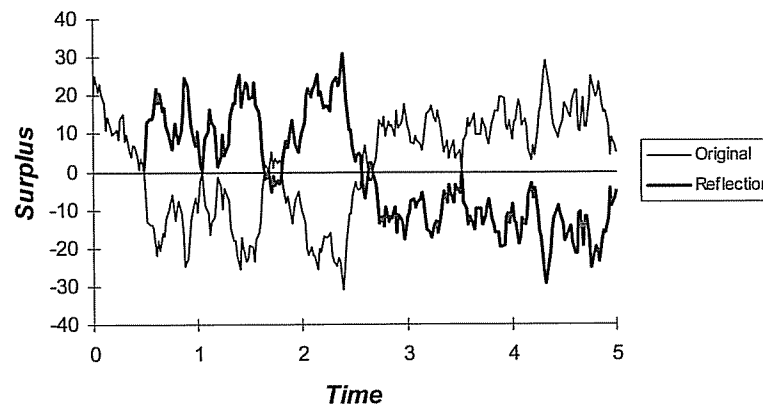


Fig. 8.4 Type B sample path and its reflection.

Proof: Any sample path of U_t with a final level $U_\tau < 0$ must have first crossed the barrier $U_t = 0$ at some time $T < \tau$. For any such path U_t , we define a new path U_t^* which is the same as the original sample path for all $t < T$ but is the reflection about the barrier $U_t = 0$ of the original sample path for all $t > T$. Then

$$U_t^* = \begin{cases} U_t, & t \leq T, \\ -U_t, & t > T. \end{cases} \quad (8.30)$$

The reflected path U_t^* has final value $U_\tau^* = -U_\tau$. These are illustrated in Figure 8.4, which is based on the sample path in Figure 8.3.

Now consider any path that crosses the barrier $U_t = 0$ in the time interval $(0, \tau)$. Any such path is one of two possible types:

Type A: One that has a final value of $U_\tau < 0$.

Type B: One that has a final value of $U_\tau > 0$.

Any path of Type B is a reflection of some other path of Type A. Hence, sample paths can be considered in reflecting pairs. The probability of ruin at some time in $(0, \tau)$ is the total probability of all such pairs:

$$\psi(u, \tau) = \Pr\{T < \tau\} = \Pr\left\{\min_{0 < t < \tau} U_t < 0\right\},$$

where it is understood that all probabilities are conditional on $U_0 = u$. This probability is obtained by considering all original paths of Type A with final values $U_\tau = x < 0$. By adding all the corresponding reflecting paths U_t^* as well, all possible paths that cross the ruin barrier are considered. Note that

the case $U_\tau = 0$ has been left out. The probability of this happening is zero and so this event can be ignored. In Figure 8.4 the original sample path ended at a positive surplus value and so is of Type B. The reflection is of Type A.

Let A_x and B_x denote the sets of all possible paths of Types A and B respectively, which end at $U_\tau = x$ for Type A and $U_\tau = -x$ for Type B. Further let $\Pr\{A_x\}$ and $\Pr\{B_x\}$ denote the total probability associated with the paths in the sets.⁶ Hence the probability of ruin is

$$\Pr\{T < \tau\} = \int_{-\infty}^0 \Pr\{U_\tau = x\} \frac{\Pr\{A_x\} + \Pr\{B_x\}}{\Pr\{U_\tau = x\}} dx. \quad (8.31)$$

Because A_x is the set of all possible paths ending at x ,

$$\Pr\{A_x\} = \Pr\{U_\tau = x\},$$

where the right-hand side is the pdf of U_τ . Then

$$\Pr\{T < \tau\} = \int_{-\infty}^0 \Pr\{U_\tau = x\} \left[1 + \frac{\Pr\{B_x\}}{\Pr\{A_x\}} \right] dx. \quad (8.32)$$

Because $U_t - u$ is a Brownian motion with drift process, U_τ is normally distributed with mean $u + \mu\tau$ and variance $\sigma^2\tau$, and so

$$\Pr\{U_\tau = x\} = (2\pi\sigma^2\tau)^{-1/2} \exp\left[-\frac{(x - u - \mu\tau)^2}{2\sigma^2\tau}\right].$$

To obtain $\Pr\{B_x\}/\Pr\{A_x\}$, we condition on all possible ruin times T . Then

$$\begin{aligned} \frac{\Pr\{B_x\}}{\Pr\{A_x\}} &= \frac{\int_0^\tau \Pr\{B_x|T=t\} \Pr\{T=t\} dt}{\int_0^\tau \Pr\{A_x|T=t\} \Pr\{T=t\} dt} \\ &= \frac{\int_0^\tau \Pr\{U_\tau = -x|T=t\} \Pr\{T=t\} dt}{\int_0^\tau \Pr\{U_\tau = x|T=t\} \Pr\{T=t\} dt}. \end{aligned}$$

The conditional pdf of $U_\tau|T=t$ is the same as the pdf of $U_\tau - U_t$ because $T=t$ implies $U_t = 0$. The process U_t has independent increments and so

⁶We are abusing probability notation here. The actual probability of these events is zero. What is being called a probability is really a probability density which can be integrated to produce probabilities for sets which have positive probability.

$U_\tau - U_t$ has a normal distribution. Then,

$$\begin{aligned} \Pr\{U_\tau = x|T=t\} &= \Pr\{U_\tau - U_t = x\} \\ &= \frac{\exp\left\{-\frac{[x - \mu(\tau-t)]^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}} \\ &= \frac{\exp\left\{-\frac{x^2 - 2x\mu(\tau-t) + \mu^2(\tau-t)^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}} \\ &= \exp\left(\frac{x\mu}{\sigma^2}\right) \frac{\exp\left\{-\frac{x^2 + \mu^2(\tau-t)^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}}. \end{aligned}$$

Similarly, by replacing x with $-x$,

$$\Pr\{U_\tau = -x|T=t\} = \frac{\exp\left(-\frac{x\mu}{\sigma^2}\right) \exp\left\{-\frac{x^2 + \mu^2(\tau-t)^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}}.$$

We then have

$$\begin{aligned} \frac{\Pr\{B_x\}}{\Pr\{A_x\}} &= \frac{\int_0^\tau \frac{\exp\left(-\frac{x\mu}{\sigma^2}\right) \exp\left\{-\frac{x^2 + \mu^2(\tau-t)^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}} \Pr\{T=t\} dt}{\int_0^\tau \frac{\exp\left(\frac{x\mu}{\sigma^2}\right) \exp\left\{-\frac{x^2 + \mu^2(\tau-t)^2}{2\sigma^2(\tau-t)}\right\}}{\sqrt{2\pi\sigma^2(\tau-t)}} \Pr\{T=t\} dt} \\ &= \exp\left(-\frac{2\mu x}{\sigma^2}\right). \end{aligned}$$

Then from (8.31)

$$\begin{aligned}
 \psi(u, \tau) &= \Pr\{T < \tau\} \\
 &= \int_{-\infty}^0 \Pr\{U_\tau = x\} \left[1 + \frac{\Pr\{B_x\}}{\Pr\{A_x\}}\right] dx \\
 &= \Phi\left(\frac{-u - \mu\tau}{\sqrt{\sigma^2\tau}}\right) \\
 &\quad + \int_{-\infty}^0 (2\pi\sigma^2\tau)^{-1/2} \exp\left[-\frac{(x - u - \mu\tau)^2}{2\sigma^2\tau} - \frac{2\mu x}{\sigma^2}\right] dx \\
 &= \Phi\left(-\frac{u + \mu\tau}{\sqrt{\sigma^2\tau}}\right) \\
 &\quad + \int_{-\infty}^0 (2\pi\sigma^2\tau)^{-1/2} \exp\left[-\frac{(x - u + \mu\tau)^2}{2\sigma^2\tau} + 4\mu u\tau\right] dx \\
 &= \Phi\left(-\frac{u + \mu\tau}{\sqrt{\sigma^2\tau}}\right) + \exp\left(-\frac{2\mu}{\sigma^2}u\right) \Phi\left(-\frac{u - \mu\tau}{\sqrt{\sigma^2\tau}}\right). \quad \square
 \end{aligned}$$

Corollary 8.25 *The probability of ultimate ruin is given by*

$$\psi(u) = 1 - \phi(u) = \Pr\{T < \infty\} = \exp\left(-\frac{2\mu}{\sigma^2}u\right). \quad (8.33)$$

By letting $\tau \rightarrow \infty$, this result follows from Theorem 8.24. It should be noted that the distribution (8.29) is a **defective distribution** because the cdf does not approach 1 as $\tau \rightarrow \infty$. The corresponding proper distribution is obtained by conditioning on ultimate ruin.

Corollary 8.26 *The distribution of the time until ruin given that it occurs is given by*

$$\begin{aligned}
 \frac{\psi(u, \tau)}{\psi(u)} &= \Pr\{T < \tau | T < \infty\} \\
 &= \exp\left(\frac{2\mu}{\sigma^2}u\right) \Phi\left(-\frac{u + \mu\tau}{\sqrt{\sigma^2\tau}}\right) + \Phi\left(-\frac{u - \mu\tau}{\sqrt{\sigma^2\tau}}\right), \tau > 0. \quad (8.34)
 \end{aligned}$$

Corollary 8.27 *The probability density function of the time until ruin is given by*

$$f_T(\tau) = \frac{u}{\sqrt{2\pi\sigma^2}} \tau^{-3/2} \exp\left[-\frac{(u - \mu\tau)^2}{2\sigma^2\tau}\right], \quad \tau > 0. \quad (8.35)$$

This is obtained by differentiation of (8.34) with respect to τ . It is not hard to see, using Appendix A, that if $\mu > 0$ then (8.35) is the pdf of an inverse Gaussian distribution with mean u/μ and variance $u\sigma^2/\mu^3$. With $\mu = 0$, ruin is certain and the time until ruin (8.35) has pdf

$$f_T(\tau) = \frac{u}{\sqrt{2\pi\sigma^2}} \tau^{-3/2} \exp\left(-\frac{u^2}{2\sigma^2\tau}\right), \quad \tau > 0,$$

and cdf [from (8.34)]

$$F_T(\tau) = 2\Phi\left(-\frac{u}{\sigma\tau^{1/2}}\right), \quad \tau > 0.$$

This distribution is the **one-sided stable law with index 1/2**.

These results can be used as approximations for the original surplus process (8.26) based on the compound Poisson model. In this situation $c = (1 + \theta)\lambda E(X)$ where θ is the relative premium loading. We do this by substituting back using (8.27) and (8.28).

Then, for the process (8.26) from (8.29), (8.33), and (8.35), we have

$$\begin{aligned}
 \psi(u, \tau) &\doteq \Phi\left[-\frac{u + \theta\lambda\tau E(X)}{\sqrt{\lambda\tau E(X^2)}}\right] \\
 &\quad + \exp\left[-\frac{2E(X)}{E(X^2)}\theta u\right] \Phi\left[-\frac{u - \theta\lambda\tau E(X)}{\sqrt{\lambda\tau E(X^2)}}\right], \quad u > 0, \tau > 0, \\
 \psi(u) &\doteq \exp\left[-\frac{2E(X)}{E(X^2)}\theta u\right], \quad u > 0,
 \end{aligned}$$

and

$$f_T(\tau) \doteq \frac{u}{\sqrt{2\pi\lambda E(X^2)}} \tau^{-3/2} \exp\left\{-\frac{[u - \theta\lambda\tau E(X)]^2}{2\lambda\tau E(X^2)}\right\}, \quad \tau > 0.$$

Then, for any compound Poisson-based process, it is easy to get simple numerical approximations. For example, the expected time until ruin, given that it occurs, is

$$E(T) = \frac{u}{\mu} = \frac{u}{\theta\lambda E(X)}. \quad (8.36)$$

Naturally, the accuracy of this approximation depends on the relative sizes of the quantities involved.

It should be noted from (8.36) that the expected time of ruin, given that it occurs, depends (as we might expect) on the four key quantities which describe the surplus process. A higher initial surplus (u) increases the time to ruin, while increasing any of the other components decreases the expected time. This may appear surprising at first, but, for example, increasing the loading increases the rate of expected growth in surplus, making ruin difficult. Therefore, if ruin should occur, it will have to happen soon, before the high loading leads to large gains. If λ is large, the company is essentially much larger and events happen more quickly. Therefore, ruin, if it happens, will occur sooner. Finally, a large value of $E(X)$ makes it easier for an early claim to wipe out the initial surplus.

All these are completely intuitive. However, formula (8.36) shows how each factor can have an influence on the expected ruin time.

On a final note, it is also possible to use the compound Poisson-based risk process Z_t as an approximation for a Brownian motion process. For known

drift and variance parameters μ and σ^2 , one can use (8.27) and (8.28) to obtain

$$\mu = \theta \lambda E(X), \quad (8.37)$$

$$\sigma^2 = \lambda E(X^2). \quad (8.38)$$

It is convenient to fix the jump sizes so that $E(X) = k$, say, and $E(X^2) = k^2$.

Then we have

$$\lambda = \frac{\sigma^2}{k^2}, \quad (8.39)$$

$$\theta = \frac{\mu}{\lambda k} = \frac{\mu}{\sigma^2} k. \quad (8.40)$$

When μ and σ^2 are fixed, choosing a value of k fixes λ and θ . Hence, the Poisson-based process can be used to approximate the Brownian motion with accuracy determined only by the parameter k . The smaller the value of k , the smaller the jump sizes and the larger the number of jumps per unit time.

Hence, simulation of the Brownian motion process can be done using the Poisson-based process. To simulate the Poisson-based process, the waiting times between successive events are generated first because these are exponentially distributed with mean $1/\lambda$. As k becomes small, λ becomes large and the mean waiting time becomes small.

Part III

Construction of empirical models

9

Review of mathematical statistics

9.1 INTRODUCTION

Before studying empirical models and then parametric models, we review some concepts from mathematical statistics. Mathematical statistics is a broad subject that includes many topics not covered in this chapter. For those that are covered, it is assumed that the reader has had some prior exposure. The topics of greatest importance for constructing actuarial models are estimation and hypothesis testing. Because the Bayesian approach to statistical inference is often either ignored or treated lightly in introductory mathematical statistics texts and courses, it receives more in-depth coverage in this text in Section 12.4. Bayesian methodology also provides the basis for the credibility methods covered in Chapter 16.

To see the need for methods of statistical inference, consider the case where your boss needs a model for basic dental payments. One option is to simply announce the model. You announce that it is the lognormal distribution with $\mu = 5.1239$ and $\sigma = 1.0345$ (the many decimal places are designed to give your announcement an aura of precision). When your boss, or a regulator, or an attorney who has put you on the witness stand asks you how you know that to be so, it will likely not be sufficient to answer that “I just know these

Loss Models: From Data to Decisions, Second Edition.

By Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot

ISBN 0-471-21577-5 Copyright © 2004 John Wiley & Sons, Inc.

things." It may not even be sufficient to announce that your friend at Gamma Dental uses that model.

An alternative is to collect some data and use it to formulate a model. Most distributional models have two components. The first is a name, such as "Pareto." The second is the set of values of parameters that complete the specification. Matters would be simpler if modeling could be done in that order. Most of the time we need to fix the parameters that go with a named model before we can decide if we want to use that type of model.

Because the parameter estimates are based on a sample from the population and not the entire population, the results will not be the true values. It is important to have an idea of the potential error. One way to express this is with an interval estimate. That is, rather than announcing a particular value, a range of plausible values is presented.

When named parametric distributions are used, the parameterizations used are those from Appendices A and B.

9.2 POINT ESTIMATION

9.2.1 Introduction

Regardless of how a model is estimated, it is extremely unlikely that the estimated model will exactly match the true distribution. Ideally, we would like to be able to measure the error we will be making when using the estimated model. But this is clearly impossible! If we knew the amount of error we had made, we could adjust our estimate by that amount and then have no error at all. The best we can do is discover how much error is inherent in repeated use of the *procedure*, as opposed to how much error we made with our current estimate. Therefore, this section is about the quality of the ensemble of answers produced from the procedure, not about the quality of a particular answer.

This is an important point with regard to actuarial practice. What is important is that an appropriate procedure be used, with everyone understanding that even the best procedure can lead to a poor result once the random future outcome has been revealed. This point is stated nicely in a Society of Actuaries principles draft ([124], pp. 779–780) regarding the level of adequacy of a provision for a block of life risk obligations (that is, the probability that the company will have enough money to meet its contractual obligations):

The indicated level of adequacy is prospective, but the actuarial model is generally validated against past experience. It is incorrect to conclude on the basis of subsequent experience that the actuarial assumptions were inappropriate or that the indicated level of adequacy was overstated or understated.

When constructing models, there are a number of types of error. Several will not be covered here. Among them are model error (choosing the wrong

model) and sampling frame error (trying to draw inferences about a population that differs from the one sampled). An example of model error is selecting a Pareto distribution when the true distribution is Weibull. An example of sampling frame error is sampling claims from policies sold by independent agents in order to price policies sold over the Internet.

The type of error we can measure is that due to using a sample from the population to make inferences about the entire population. Errors occur when the items sampled do not represent the population. As noted earlier, we cannot know if the particular items sampled today do or do not represent the population. We can, however, estimate the extent to which estimators are affected by the possibility of a nonrepresentative sample.

The approach taken in this section is to consider all the samples that might be taken from the population. Each such sample leads to an estimated quantity (for example, a probability, a parameter value, or a moment). We do not expect the estimated quantities to always match the true value. For a sensible estimation procedure we do expect that for some samples the quantity will match the true value, for many it will be close, and for only a few it will be quite different. If we can construct a measure of how well the set of potential estimates matches the true value, we have a handle on the quality of our estimation procedure. The approach outlined here is often called the *classical* or *frequentist* approach to estimation.

Finally, we need a word about the difference between *estimate* and *estimator*. The former refers to the specific value obtained when applying an estimation procedure to a set of numbers. The latter refers to a rule or formula that produces the estimate. An estimate is a number or function while an estimator is a random variable or a random function. Usually, both the words and the context will make clear which is being referred to.

9.2.2 Measures of quality

9.2.2.1 Introduction There are a number of ways to measure the quality of an estimator. Three of them are discussed here. Two examples will be used throughout to illustrate them.

Example 9.1 A population contains the values 1, 3, 5, and 9. We want to estimate the population mean by taking a sample of size 2 with replacement.

Example 9.2 A population has the exponential distribution with a mean of θ . We want to estimate the population mean by taking a sample of size 3 with replacement.

Both examples are clearly artificial in that we know the answers prior to sampling (4.5 and θ). However, that knowledge will make apparent the error in the procedure we select. For practical applications, we will need to be able to estimate the error when we do not know the true value of the quantity being estimated.

9.2.2.2 Unbiasedness When constructing an estimator, it would be good if, on average, the errors we make cancel each other out. More formally, let θ be the quantity we want to estimate. Let $\hat{\theta}$ be the random variable that represents the estimator and let $E(\hat{\theta}|\theta)$ be the expected value of the estimator $\hat{\theta}$ when θ is the true parameter value.

Definition 9.3 An estimator, $\hat{\theta}$, is *unbiased* if $E(\hat{\theta}|\theta) = \theta$ for all θ . The *bias* is $\text{bias}_{\hat{\theta}}(\theta) = E(\hat{\theta}|\theta) - \theta$.

The bias depends on the estimator being used and may also depend on the particular value of θ .

Example 9.4 For Example 9.1 determine the bias of the sample mean as an estimator of the population mean.

The population mean is $\theta = 4.5$. The sample mean is the average of the two observations. It is also the estimator we would use employing the empirical approach. In all cases, we assume that sampling is random. In other words, every sample of size n has the same chance of being drawn. Such sampling also implies that any member of the population has the same chance of being observed as any other member. For this example, there are 16 equally likely ways the sample could have turned out. They are listed below.

1,1	1,3	1,5	1,9	3,1	3,3	3,5	3,9
5,1	5,3	5,5	5,9	9,1	9,3	9,5	9,9

This leads to the following 16 equally likely values for the sample mean:

1	2	3	5	2	3	4	6
3	4	5	7	5	6	7	9

Combining the common values, the sample mean, usually denoted \bar{X} , has the following probability distribution:

x	1	2	3	4	5	6	7	9
$p_{\bar{X}}(x)$	1/16	2/16	3/16	2/16	3/16	2/16	2/16	1/16

The expected value of the estimator is

$$E(\bar{X}) = [1(1) + 2(2) + 3(3) + 4(2) + 5(3) + 6(2) + 7(2) + 9(1)]/16 = 4.5$$

and so the sample mean is an unbiased estimator of the population mean for this example. \square

Example 9.5 For Example 9.2 determine the bias of the sample mean and the sample median as estimators of the population mean.

The sample mean is $\bar{X} = (X_1 + X_2 + X_3)/3$, where each X_j represents one of the observations from the exponential population. Its expected value is

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + X_2 + X_3}{3}\right) = \frac{1}{3}[E(X_1) + E(X_2) + E(X_3)] \\ &= \frac{1}{3}(\theta + \theta + \theta) = \theta \end{aligned}$$

and therefore the sample mean is an unbiased estimator of the population mean.

Investigating the sample median is a bit more difficult. The distribution function of the middle of three observations can be found as follows, using Y as the random variable of interest and X as the random variable for an observation from the population:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(X_1, X_2, X_3 \leq y) + \Pr(X_1, X_2 \leq y, X_3 > y) \\ &\quad + \Pr(X_1, X_3 \leq y, X_2 > y) + \Pr(X_2, X_3 \leq y, X_1 > y) \\ &= F_X(y)^3 + 3F_X(y)^2[1 - F_X(y)] \\ &= [1 - e^{-y/\theta}]^3 + 3[1 - e^{-y/\theta}]^2 e^{-y/\theta}. \end{aligned}$$

The density function is

$$f_Y(y) = F'_Y(y) = \frac{6}{\theta} (e^{-2y/\theta} - e^{-3y/\theta}).$$

The expected value of this estimator is

$$\begin{aligned} E(Y|\theta) &= \int_0^\infty y \frac{6}{\theta} (e^{-2y/\theta} - e^{-3y/\theta}) dy \\ &= \frac{5\theta}{6}. \end{aligned}$$

This estimator is clearly biased,¹ with $\text{bias}_Y(\theta) = 5\theta/6 - \theta = -\theta/6$. On average, this estimator underestimates the true value. It is also easy to see that the sample median can be turned into an unbiased estimator by multiplying it by 1.2. \square

For Example 9.2 we have two estimators (the sample mean and 1.2 times the sample median) that are both unbiased. We will need additional criteria to decide which one we prefer.

¹The sample median is not likely to be a good estimator of the population mean. This example studies it for comparison purposes. Because the population median is $\theta \ln 2$, the sample median is biased for the population median.

Some estimators exhibit a small amount of bias, which vanishes as the sample size goes to infinity.

Definition 9.6 Let $\hat{\theta}_n$ be an estimator of θ based on a sample size of n . The estimator is *asymptotically unbiased* if

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n | \theta) = \theta$$

for all θ .

Example 9.7 Suppose a random variable has the uniform distribution on the interval $(0, \theta)$. Consider the estimator $\hat{\theta}_n = \max(X_1, \dots, X_n)$. Show that this estimator is asymptotically unbiased.

Let Y_n be the maximum from a sample of size n . Then

$$\begin{aligned} F_{Y_n}(y) &= \Pr(Y_n \leq y) = \Pr(X_1 \leq y, \dots, X_n \leq y) \\ &= [F_X(y)]^n \\ &= (y/\theta)^n \\ f_{Y_n}(y) &= \frac{ny^{n-1}}{\theta^n}, \quad 0 < y < \theta. \end{aligned}$$

The expected value is

$$E(Y_n | \theta) = \int_0^\theta ny^n \theta^{-n} dy = \frac{n}{n+1} y^{n+1} \theta^{-n} \Big|_0^\theta = \frac{n\theta}{n+1}.$$

As $n \rightarrow \infty$, the limit is θ , making this estimator asymptotically unbiased. \square

9.2.2.3 Consistency A second desirable property of an estimator is that it works well for extremely large samples. Slightly more formally, as the sample size goes to infinity, the probability that the estimator is in error by more than a small amount goes to zero. A formal definition follows.

Definition 9.8 An estimator is *consistent* (often called, in this context, *weakly consistent*) if, for all $\delta > 0$ and any θ ,

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\theta}_n - \theta| > \delta) = 0.$$

A sufficient (although not necessary) condition for weak consistency is that the estimator be asymptotically unbiased and $\text{Var}(\hat{\theta}_n) \rightarrow 0$.

Example 9.9 Prove that, if the variance of a random variable is finite, the sample mean is a consistent estimator of the population mean.

From Exercise 9.2, the sample mean is unbiased. In addition,

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) \\ &= \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) \\ &= \frac{\text{Var}(X)}{n} \rightarrow 0. \end{aligned}$$

The second step follows from assuming that the observations are independent. \square

Example 9.10 Show that the maximum observation from a uniform distribution on the interval $(0, \theta)$ is a consistent estimator of θ .

From Example 9.7, the maximum is asymptotically unbiased. The second moment is

$$E(Y_n^2) = \int_0^\theta ny^{n+1} \theta^{-n} dy = \frac{n}{n+2} y^{n+2} \theta^{-n} \Big|_0^\theta = \frac{n\theta^2}{n+2}$$

and then

$$\text{Var}(Y_n) = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+2)(n+1)^2} \rightarrow 0. \quad \square$$

9.2.2.4 Mean-squared error While consistency is nice, most estimators have this property. What would be truly impressive is an estimator that is not only correct on average but comes very close most of the time and, in particular, comes closer than rival estimators. One measure for a finite sample is motivated by the definition of consistency. The quality of an estimator could be measured by the probability that it gets within δ of the true value—that is, by measuring $\Pr(|\hat{\theta}_n - \theta| < \delta)$. But the choice of δ is arbitrary and we prefer measures that cannot be altered to suit the investigator's whim. Then we might consider $E(|\hat{\theta}_n - \theta|)$, the average absolute error. But we know that working with absolute values often presents unpleasant mathematical challenges, and so the following has become widely accepted as a measure of accuracy.

Definition 9.11 The *mean-squared error (MSE)* of an estimator is

$$\text{MSE}_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2 | \theta].$$

Note that the MSE is a function of the true value of the parameter. An estimator may perform extremely well for some values of the parameter but poorly for others.

Example 9.12 Consider the estimator $\hat{\theta} = 5$ of an unknown parameter θ . The MSE is $(5 - \theta)^2$, which is very small when θ is near 5 but becomes poor for other values. Of course this estimate is both biased and inconsistent unless θ is exactly equal to 5.

A result that follows directly from the various definitions is

$$\text{MSE}_{\hat{\theta}}(\theta) = E\{[\hat{\theta} - E(\hat{\theta}|\theta) + E(\hat{\theta}|\theta) - \theta]^2|\theta\} = \text{Var}(\hat{\theta}|\theta) + [\text{bias}_{\hat{\theta}}(\theta)]^2. \quad (9.1)$$

If we restrict attention to only unbiased estimators, the best such could be defined as follows.

Definition 9.13 An estimator, $\hat{\theta}$, is called a **uniformly minimum variance unbiased estimator (UMVUE)** if it is unbiased and for any true value of θ there is no other unbiased estimator that has a smaller variance.

Because we are looking only at unbiased estimators, it would have been equally effective to make the definition in terms of MSE. We could also generalize the definition by looking for estimators that are uniformly best with regard to MSE, but the previous example indicates why that is not feasible. There are a few theorems that can assist with the determination of UMVUEs. However, such estimators are difficult to determine. On the other hand, MSE is still a useful criterion for comparing two alternative estimators.

Example 9.14 For Example 9.2 compare the MSEs of the sample mean and 1.2 times the sample median.

The sample mean has variance

$$\frac{\text{Var}(X)}{3} = \frac{\theta^2}{3}.$$

When multiplied by 1.2, the sample median has second moment

$$\begin{aligned} E[(1.2Y)^2] &= 1.44 \int_0^\infty y^2 \frac{6}{\theta} \left(e^{-2y/\theta} - e^{-3y/\theta} \right) dy \\ &= 1.44 \frac{6}{\theta} \left[y^2 \left(\frac{-\theta}{2} e^{-2y/\theta} + \frac{\theta}{3} e^{-3y/\theta} \right) \right. \\ &\quad \left. - 2y \left(\frac{\theta^2}{4} e^{-2y/\theta} - \frac{\theta^2}{9} e^{-3y/\theta} \right) \right. \\ &\quad \left. + 2 \left(\frac{-\theta^3}{8} e^{-2y/\theta} + \frac{\theta^3}{27} e^{-3y/\theta} \right) \right] \Big|_0^\infty \\ &= \frac{8.64}{\theta} \left(\frac{2\theta^3}{8} - \frac{2\theta^3}{27} \right) = \frac{38\theta^2}{25} \end{aligned}$$

for a variance of

$$\frac{38\theta^2}{25} - \theta^2 = \frac{13\theta^2}{25} > \frac{\theta^2}{3}.$$

The sample mean has the smaller MSE regardless of the true value of θ . Therefore, for this problem, it is a superior estimator of θ . \square

Example 9.15 For the uniform distribution on the interval $(0, \theta)$ compare the MSE of the estimators $2\bar{X}$ and $[(n+1)/n] \max(X_1, \dots, X_n)$. Also evaluate the MSE of $\max(X_1, \dots, X_n)$.

The first two estimators are unbiased, so it is sufficient to compare their variances. For twice the sample mean,

$$\text{Var}(2\bar{X}) = \frac{4}{n} \text{Var}(X) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

For the adjusted maximum, the second moment is

$$E \left[\left(\frac{n+1}{n} Y_n \right)^2 \right] = \frac{(n+1)^2}{n^2} \frac{n\theta^2}{n+2} = \frac{(n+1)^2 \theta^2}{(n+2)n}$$

for a variance of

$$\frac{(n+1)^2 \theta^2}{(n+2)n} - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

Except for the case $n = 1$ (and then the two estimators are identical), the one based on the maximum has the smaller MSE. The third estimator is biased. For it, the MSE is

$$\frac{n\theta^2}{(n+2)(n+1)^2} + \left(\frac{n\theta}{n+1} - \theta \right)^2 = \frac{2\theta^2}{(n+1)(n+2)},$$

which is also larger than that for the adjusted maximum. \square

9.2.3 Exercises

9.1 For Example 9.1, show that the mean of three observations drawn without replacement is an unbiased estimator of the population mean while the median of three observations drawn without replacement is a biased estimator of the population mean.

9.2 Prove that for random samples the sample mean is always an unbiased estimator of the population mean.

9.3 Let X have the uniform distribution over the range $(\theta - 2, \theta + 2)$. That is, $f_X(x) = 0.25$, $\theta - 2 < x < \theta + 2$. Show that the median from a sample of size 3 is an unbiased estimator of θ .

9.4 Explain why the sample mean may not be a consistent estimator of the population mean for a Pareto distribution.

9.5 For the sample of size 3 in Exercise 9.3, compare the MSE of the sample mean and median as estimates of θ .

9.6 (*) You are given two independent estimators of an unknown quantity θ . For estimator A , $E(\hat{\theta}_A) = 1,000$ and $\text{Var}(\hat{\theta}_A) = 160,000$, while for estimator B , $E(\hat{\theta}_B) = 1,200$ and $\text{Var}(\hat{\theta}_B) = 40,000$. Estimator C is a weighted average, $\hat{\theta}_C = w\hat{\theta}_A + (1-w)\hat{\theta}_B$. Determine the value of w that minimizes $\text{Var}(\hat{\theta}_C)$.

9.7 (*) A population of losses has the Pareto distribution (see Appendix A) with $\theta = 6,000$ and α unknown. Simulation of the results from maximum likelihood estimation based on samples of size 10 has indicated that $E(\hat{\alpha}) = 2.2$ and $\text{MSE}(\hat{\alpha}) = 1$. Determine $\text{Var}(\hat{\alpha})$ if it is known that $\alpha = 2$.

9.8 (*) Two instruments are available for measuring a particular nonzero distance. The random variable X represents a measurement with the first instrument, and the random variable Y with the second instrument. Assume X and Y are independent with $E(X) = 0.8m$, $E(Y) = m$, $\text{Var}(X) = m^2$, and $\text{Var}(Y) = 1.5m^2$, where m is the true distance. Consider estimators of m that are of the form $Z = \alpha X + \beta Y$. Determine the values of α and β that make Z a UMVUE within the class of estimators of this form.

9.9 A population contains six members, with values 1, 1, 2, 3, 5, and 10. A random sample of size 3 is drawn without replacement. In each case the objective is to estimate the population mean. *Note:* A spreadsheet with an optimization routine may be the best way to solve this problem.

- Determine the bias, variance, and MSE of the sample mean.
- Determine the bias, variance, and MSE of the sample median.
- Determine the bias, variance, and MSE of the sample midrange (the average of the largest and smallest observations).
- Consider an arbitrary estimator of the form $aX_{(1)} + bX_{(2)} + cX_{(3)}$, where $X_{(1)} \leq X_{(2)} \leq X_{(3)}$, are the sample order statistics.
 - Determine a restriction on the values of a , b , and c that will assure that the estimator is unbiased.
 - Determine the values of a , b , and c that will produce the unbiased estimator with the smallest variance.
 - Determine the values of a , b , and c that will produce the (possibly biased) estimator with the smallest MSE.

9.10 (*) Two different estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, are being considered. To test their performance, 75 trials have been simulated, each with the true value set at $\theta = 2$. The following totals were obtained:

$$\sum_{j=1}^{75} \hat{\theta}_{1j} = 165, \quad \sum_{j=1}^{75} \hat{\theta}_{1j}^2 = 375, \quad \sum_{j=1}^{75} \hat{\theta}_{2j} = 147, \quad \sum_{j=1}^{75} \hat{\theta}_{2j}^2 = 312,$$

where $\hat{\theta}_{ij}$ is the estimate based on the j th simulation using estimator $\hat{\theta}_i$. Estimate the MSE for each estimator and determine the **relative efficiency** (the ratio of the MSEs).

9.3 INTERVAL ESTIMATION

All of the estimators discussed to this point have been **point estimators**. That is, the estimation process produces a single value that represents our best attempt to determine the value of the unknown population quantity. While that value may be a good one, we do not expect it to exactly match the true value. A more useful statement is often provided by an **interval estimator**. Instead of a single value, the result of the estimation process is a range of possible numbers, any of which is likely to be the true value. A specific type of interval estimator is the confidence interval.

Definition 9.16 A $100(1-\alpha)\%$ **confidence interval** for a parameter θ is a pair of values L and U computed from a random sample such that $\Pr(L \leq \theta \leq U) \geq 1-\alpha$ for all θ .

Note that this definition does not uniquely specify the interval. Because the definition is a probability statement and must hold for all θ , it says nothing about whether or not a particular interval encloses the true value of θ from a particular population. Instead, the **level of confidence**, $1-\alpha$, is a property of the method used to obtain L and U and not of the particular values obtained. The proper interpretation is that, if we use a particular interval estimator over and over on a variety of samples, at least $100(1-\alpha)\%$ of the time our interval will enclose the true value.

Constructing confidence intervals is usually very difficult. For example, we know that, if a population has a normal distribution with unknown mean and variance, a $100(1-\alpha)\%$ confidence interval for the mean uses

$$L = \bar{X} - t_{\alpha/2, n-1} s / \sqrt{n}, \quad U = \bar{X} + t_{\alpha/2, n-1} s / \sqrt{n}, \quad (9.2)$$

where $s = \sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 / (n-1)}$ and $t_{\alpha/2, b}$ is the $100(1-\alpha/2)$ th percentile of the t distribution with b degrees of freedom. But it takes a great deal of effort to verify that this is correct (see, for example, [58], p. 214).

However, there is a method for constructing approximate confidence intervals that is often accessible. Suppose we have a point estimator $\hat{\theta}$ of parameter θ such that $E(\hat{\theta}) = \theta$, $\text{Var}(\hat{\theta}) = v(\theta)$, and $\hat{\theta}$ has approximately a normal distribution. Theorem 12.13 shows that this is often the case. With all these approximations, we have that approximately

$$1-\alpha \doteq \Pr \left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{v(\theta)}} \leq z_{\alpha/2} \right), \quad (9.3)$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)$ th percentile of the standard normal distribution. Solving for θ produces the desired interval. Sometimes this is difficult to do (due to the appearance of θ in the denominator) and so, if necessary, replace $v(\theta)$ in (9.3) with $v(\hat{\theta})$ to obtain a further approximation,

$$1 - \alpha \doteq \Pr \left(\hat{\theta} - z_{\alpha/2} \sqrt{v(\hat{\theta})} \leq \theta \leq \hat{\theta} + z_{\alpha/2} \sqrt{v(\hat{\theta})} \right). \quad (9.4)$$

Example 9.17 Use (9.4) to construct an approximate 95% confidence interval for the mean of a normal population with unknown variance.

Use $\hat{\theta} = \bar{X}$ and then note that $E(\hat{\theta}) = \theta$, $\text{Var}(\hat{\theta}) = \sigma^2/n$, and $\hat{\theta}$ does have a normal distribution. The confidence interval is then $\bar{X} \pm 1.96s/\sqrt{n}$. Because $t_{0.025, n-1} > 1.96$, this approximate interval must be narrower than the exact interval given by (9.2). That means that our level of confidence is something less than 95%. \square

Example 9.18 Use (9.3) and (9.4) to construct approximate 95% confidence intervals for the mean of a Poisson distribution. Obtain intervals for the particular case where $n = 25$ and $\bar{x} = 0.12$.

Let $\hat{\theta} = \bar{X}$, the sample mean. For the Poisson distribution, $E(\hat{\theta}) = E(X) = \theta$ and $v(\theta) = \text{Var}(\bar{X}) = \text{Var}(X)/n = \theta/n$. For the first interval

$$0.95 \doteq \Pr \left(-1.96 \leq \frac{\bar{X} - \theta}{\sqrt{\theta/n}} \leq 1.96 \right)$$

is true if and only if

$$|\bar{X} - \theta| \leq 1.96 \sqrt{\frac{\theta}{n}},$$

which is equivalent to

$$(\bar{X} - \theta)^2 \leq \frac{3.8416\theta}{n}$$

or

$$\theta^2 - \theta \left(2\bar{X} + \frac{3.8416}{n} \right) + \bar{X}^2 \leq 0.$$

Solving the quadratic produces the interval

$$\bar{X} + \frac{1.9208}{n} \pm \frac{1}{2} \sqrt{\frac{15.3664\bar{X} + 3.8416^2/n}{n}}$$

and for this problem the interval is 0.197 ± 0.156 .

For the second approximation the interval is $\bar{X} \pm 1.96\sqrt{\bar{X}/n}$ and for the example it is 0.12 ± 0.136 . This interval extends below zero (which is not possible for the true value of θ). This is because (9.4) is too crude an approximation in this case. \square

9.3.1 Exercise

9.11 Let x_1, \dots, x_n be a random sample from a population with pdf $f(x) = \theta^{-1}e^{-x/\theta}$, $x > 0$. This exponential distribution has a mean of θ and a variance of θ^2 . Consider the sample mean, \bar{X} , as an estimator of θ . It turns out that \bar{X}/θ has a gamma distribution with $\alpha = n$ and $\theta = 1/n$, where in the second expression the “ θ ” on the left is the parameter of the gamma distribution. For a sample of size 50 and a sample mean of 275, develop 95% confidence intervals by each of the following methods. In each case, if the formula requires the true value of θ , substitute the estimated value.

- Use the gamma distribution to determine an exact interval.
- Use a normal approximation, estimating the variance prior to solving the inequalities.
- Use a normal approximation, estimating θ after solving the inequalities.

9.4 TESTS OF HYPOTHESES

Hypothesis testing is covered in detail in most mathematical statistics texts. This review will be fairly straightforward and will not address philosophical issues or consider alternative approaches. A hypothesis test begins with two hypotheses, one called the null and one called the alternative. The traditional notation is H_0 for the null hypothesis and H_1 for the alternative hypothesis. The two hypotheses are not treated symmetrically. Reversing them may alter the results. To illustrate this process, a simple example will be used.

Example 9.19 Your company has been basing its premiums on an assumption that the average claim is 1,200. You want to raise the premium and a regulator has insisted that you provide evidence that the average now exceeds 1,200. To provide such evidence, the following numbers have been obtained. What are the hypotheses for this problem?

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1,193	1,340	1,884	2,558	15,743

Let μ be the population mean. One hypothesis (the one you claim is true) is that $\mu > 1,200$. Because hypothesis tests must present an either/or situation, the other hypothesis must be $\mu \leq 1,200$. The only remaining task is to decide which of them is the null hypothesis. Whenever the universe of continuous possibilities is divided in two there is likely to be a boundary that needs to be assigned to one hypothesis or the other. The hypothesis that includes

the boundary must be the null hypothesis. Therefore, the problem can be succinctly stated as:

$$H_0 : \mu \leq 1,200$$

$$H_1 : \mu > 1,200. \quad \square$$

The decision is made by calculating a quantity called a **test statistic**. It is a function of the observations and is treated as a random variable. That is, in designing the test procedure we are concerned with the samples that might have been obtained and not with the particular sample that was obtained. The test specification is completed by constructing a **rejection region**. It is a subset of the possible values of the test statistic. If the value of the test statistic for the observed sample is in the rejection region, the null hypothesis is rejected and the alternative hypothesis is announced as the result that is supported by the data. Otherwise, the null hypothesis is not rejected (more on this later). The boundaries of the rejection region (other than plus or minus infinity) are called the **critical values**.

Example 9.20 (Example 9.19 continued) *Complete the test using the test statistic and rejection region that is promoted in most statistics books. Assume that the population has a normal distribution with standard deviation 3,435.*

The traditional test statistic for this problem is

$$z = \frac{\bar{x} - 1,200}{3,435/\sqrt{20}} = 0.292$$

and the null hypothesis is rejected if $z > 1.645$. Because 0.292 is less than 1.645, the null hypothesis is not rejected. The data do not support the assertion that the average claim exceeds 1,200. \square

The test in the previous example was constructed to meet certain objectives. The first objective is to control what is called the Type I error. It is the error made when the test rejects the null hypothesis in a situation where it happens to be true. In the example, the null hypothesis can be true in more than one way. This leads to the most common measure of the propensity of a test to make a Type I error.

Definition 9.21 *The significance level of a hypothesis test is the probability of making a Type I error given that the null hypothesis is true. If it can be true in more than one way, the level of significance is the maximum of such probabilities. The significance level is usually denoted by the letter α .*

This is a conservative definition in that it looks at the worst case. It is typically a case that is on the boundary between the two hypotheses.

Example 9.22 *Determine the level of significance for the test in Example 9.20.*

Begin by computing the probability of making a Type I error when the null hypothesis is true with $\mu = 1,200$. Then,

$$\Pr(Z > 1.645 | \mu = 1,200) = 0.05.$$

That is because the assumptions imply that Z has a standard normal distribution.

Now suppose μ has a value that is below 1,200. Then

$$\begin{aligned} & \Pr\left(\frac{\bar{X} - 1,200}{3,435/\sqrt{20}} > 1.645\right) \\ &= \Pr\left(\frac{\bar{X} - \mu + \mu - 1,200}{3,435/\sqrt{20}} > 1.645\right) \\ &= \Pr\left(\frac{\bar{X} - \mu}{3,435/\sqrt{20}} > 1.645 - \frac{\mu - 1,200}{3,435/\sqrt{20}}\right). \end{aligned}$$

Because μ is known to be less than 1,200, the right-hand side is always greater than 1.645. The left-hand side has a standard normal distribution and therefore the probability is less than 0.05. Therefore the significance level is 0.05. \square

The significance level is usually set in advance and is often between 1 and 10%. The second objective is to keep the Type II error (not rejecting the null hypothesis when the alternative is true) probability small. Generally, attempts to reduce the probability of one type of error increase the probability of the other. The best we can do once the significance level has been set is to make the Type II error as small as possible, though there is no assurance that the probability will be a small number. The best test is one that meets the following requirement.

Definition 9.23 *A hypothesis test is uniformly most powerful if no other test exists that has the same or lower significance level and for a particular value within the alternative hypothesis has a smaller probability of making a Type II error.*

Example 9.24 (Example 9.22 continued) *Determine the probability of making a Type II error when the alternative hypothesis is true with $\mu = 2,000$.*

$$\begin{aligned} & \Pr\left(\frac{\bar{X} - 1,200}{3,435/\sqrt{20}} < 1.645 | \mu = 2,000\right) \\ &= \Pr(\bar{X} - 1,200 < 1,263.51 | \mu = 2,000) \\ &= \Pr(\bar{X} < 2,463.51 | \mu = 2,000) \\ &= \Pr\left(\frac{\bar{X} - 2,000}{3,435/\sqrt{20}} < \frac{2,463.51 - 2,000}{3,435/\sqrt{20}} = 0.6035\right) = 0.7269. \end{aligned}$$

For this value of μ , the test is not very powerful, having over a 70% chance of making a Type II error. Nevertheless (though this is not easy to prove), the test used is the most powerful test for this problem. \square

Because the Type II error probability can be high, it is customary to not make a strong statement when the null hypothesis is not rejected. Rather than say we choose or accept the null hypothesis, we say that we fail to reject it. That is, there was not enough evidence in the sample to make a strong argument in favor of the alternative hypothesis, so we take no stand at all.

A common criticism of this approach to hypothesis testing is that the choice of the significance level is arbitrary. In fact, by changing the significance level, any result can be obtained.

Example 9.25 (Example 9.24 continued) *Complete the test using a significance level of $\alpha = 0.45$. Then determine the range of significance levels for which the null hypothesis is rejected and for which it is not rejected.*

Because $\Pr(Z > 0.1257) = 0.45$, the null hypothesis is rejected when

$$\frac{\bar{X} - 1,200}{3,435/\sqrt{20}} > 0.1257.$$

In this example, the test statistic is 0.292, which is in the rejection region, and thus the null hypothesis is rejected. Of course, few people would place confidence in the results of a test that was designed to make errors 45% of the time. Because $\Pr(Z > 0.292) = 0.3851$, the null hypothesis is rejected for those who select a significance level that is greater than 38.51% and is not rejected by those who use a significance level that is less than 38.51%. \square

Few people are willing to make errors 38.51% of the time. Announcing this figure is more persuasive than the earlier conclusion based on a 5% significance level. When a significance level is used, readers are left to wonder what the outcome would have been with other significance levels. The value of 38.51% is called a p -value. A working definition is:

Definition 9.26 *For a hypothesis test, the p -value is the probability that the test statistic takes on a value that is less in agreement with the null hypothesis than the value obtained from the sample. Tests conducted at a significance level that is greater than the p -value will lead to a rejection of the null hypothesis, while tests conducted at a significance level that is smaller than the p -value will lead to a failure to reject the null hypothesis.*

Also, because the p -value must be between 0 and 1, it is on a scale that carries some meaning. The closer to zero the value is, the more support the data give to the alternative hypothesis. Common practice is that values above

10% indicate that the data provide no evidence in support of the alternative hypothesis, while values below 1% indicate strong support for the alternative hypothesis. Values in between indicate uncertainty as to the appropriate conclusion and may call for more data or a more careful look at the data or the experiment that produced it.

9.4.1 Exercise

9.12 (*Exercise 9.11 continued*) Test $H_0 : \theta \geq 325$ vs $H_1 : \theta < 325$ using a significance level of 5% and the sample mean as the test statistic. Also, compute the p -value. Do this using the exact distribution of the test statistic and a normal approximation.

10

Estimation for complete data

10.1 INTRODUCTION

The material in this and the next chapter has been traditionally presented under the heading of “survival models” with the accompanying notion that the techniques are useful only when studying lifetime distributions. Standard texts on the subject such as Klein and Moeschberger [74] and Lawless [81] contain examples that are exclusively oriented in that direction. However, as will be seen in these two chapters, the same problems that occur when modeling lifetime occur when modeling payment amount. The examples we present will be of both types. To emphasize that point, some of the starred exercises were taken from the former Society of Actuaries Course 160 exam, but the setting was changed to a payment environment. Only a handful of references are presented, most of the results being well developed in the survival models literature. Readers wanting more detail and proofs should consult a text dedicated to the subject, such as the ones mentioned above.

In this chapter it is assumed that the type of model is known but not the full description of the model. In Chapter 4, models were divided into two types—data-dependent and parametric. The definitions are repeated below.

Loss Models: From Data to Decisions, Second Edition.
By Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot
ISBN 0-471-21577-5 Copyright © 2004 John Wiley & Sons, Inc.

Definition 10.1 A *data-dependent distribution* is at least as complex as the data or knowledge that produced it and the number of "parameters" increases as the number of data points or amount of knowledge increases.

Definition 10.2 A *parametric distribution* is a set of distribution functions, each member of which is determined by specifying one or more values called *parameters*. The number of parameters is fixed and finite.

Here, only two data-dependent distributions will be considered. They depend on the data in similar ways. The simplest definitions for the two types considered appear below.

Definition 10.3 The *empirical distribution* is obtained by assigning probability $1/n$ to each data point.

Definition 10.4 A *kernel smoothed distribution* is obtained by replacing each data point with a continuous random variable and then assigning probability $1/n$ to each such random variable. The random variables used must be identical except for a location or scale change that is related to its associated data point.

Note that the empirical distribution is a special type of kernel smoothed distribution in which the random variable assigns probability 1 to the data point. An alternative to the empirical distribution that is similar in spirit but produces different numbers will also be presented. In Chapter 11 it will be shown how the definition can be modified to account for data that have been altered through censoring and truncation. With regard to kernel smoothing, there are several distributions that could be used, a few of which are introduced in Section 11.3.

Throughout this part, four examples will be used repeatedly. Because they are simply data sets, they will be referred to as Data Sets A, B, C and D.

Data Set A This data set is well-known in the casualty actuarial literature. It was first analyzed in the paper [30] by Dropkin in 1959. He collected data from 1956–1958 on the number of accidents by one driver in one year. The results for 94,935 drivers are in Table 10.1.

Data Set B These numbers are artificial. They represent the amounts paid on workers compensation medical benefits but are not related to any particular policy or set of policyholders. These payments are the full amount of the loss. A random sample of 20 payments is given in Table 10.2.

Data Set C These observations represent payments on 227 claims from a general liability insurance policy. The data are in Table 10.3.

Data Set D These numbers are artificial. They represent the time at which a five-year term insurance policy terminates. For some policyholders, termination is by death, for some it is by surrender (the cancellation of the insurance

Table 10.1 Data Set A

Number of accidents	Number of drivers
0	81,714
1	11,306
2	1,618
3	250
4	40
5 or more	7

Table 10.2 Data Set B

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1,193	1,340	1,884	2,558	15,743

Table 10.3 Data Set C

Payment range	Number of payments
0–7,500	99
7,500–17,500	42
17,500–32,500	29
32,500–67,500	28
67,500–125,000	17
125,000–300,000	9
Over 300,000	3

contract), and for the remainder it is expiration of the five-year period. Two separate versions are presented. For Data Set D1 (Table 10.4) there were 30 policies observed from issue. For each, both the time of death and time of surrender are presented, provided they were before the expiration of the five-year period. Of course, normally we do not know the time of death of policyholders who surrender and we do not know when policyholders who died would have surrendered had they not died. Note that the final 12 policyholders survived both death and surrender to the end of the five-year period.

For Data Set D2 (Table 10.5), only the time of the first event is observed. In addition, there are 10 more policyholders who were first observed at some time after the policy was issued. The table presents the results for all 40 policies. The column headed "First observed" gives the duration at which the policy was first observed; the column headed "Last observed" gives the duration at which the policy was last observed; and the column headed "Event" is coded "s" for surrender, "d" for death, and "e" for expiration of the five-year period.

Table 10.4 Data Set D1

Policyholder	Time of death	Time of surrender
1	—	0.1
2	4.8	0.5
3	—	0.8
4	0.8	3.9
5	3.1	1.8
6	—	1.8
7	—	2.1
8	—	2.5
9	—	2.8
10	2.9	4.6
11	2.9	4.6
12	—	3.9
13	4.0	—
14	—	4.0
15	—	4.1
16	4.8	—
17	—	4.8
18	—	4.8
19–30	—	—

When observations are collected from a probability distribution, the ideal situation is to have the (essentially) exact¹ value of each observation. This case is referred to as “complete, individual data.” This is the case in Data Sets B and D1. There are two reasons why exact data may not be available. One is grouping, in which all that is recorded is the range of values in which the observation belongs. This is the case for Data Set C and for Data Set A for those with five or more accidents.

A second reason that exact values may not be available is the presence of censoring or truncation. When data are censored from below, observations below a given value are known to be below that value but the exact value is unknown. When data are censored from above, observations above a given value are known to be above that value but the exact value is unknown. Note that censoring effectively creates grouped data. When the data are grouped in the first place, censoring has no effect. For example, the data in Data Set C may have been censored from above at 300,000, but we cannot know for sure

¹Some measurements are never exact. Ages may be rounded to the nearest whole number, monetary amounts to the nearest dollar, car mileage to the nearest tenth of a mile, and so on. This text is not concerned with such rounding errors. Rounded values will be treated as if they are exact.

Table 10.5 Data Set D2

Policy	First observed	Last observed	Event	Policy	First observed	Last observed	Event
1	0	0.1	s	16	0	4.8	d
2	0	0.5	s	17	0	4.8	s
3	0	0.8	s	18	0	4.8	s
4	0	0.8	d	19–30	0	5.0	e
5	0	1.8	s	31	0.3	5.0	e
6	0	1.8	s	32	0.7	5.0	e
7	0	2.1	s	33	1.0	4.1	d
8	0	2.5	s	34	1.8	3.1	d
9	0	2.8	s	35	2.1	3.9	s
10	0	2.9	d	36	2.9	5.0	e
11	0	2.9	d	37	2.9	4.8	s
12	0	3.9	s	38	3.2	4.0	d
13	0	4.0	d	39	3.4	5.0	e
14	0	4.0	s	40	3.9	5.0	e
15	0	4.1	s				

from the data set and that knowledge has no effect on how we treat the data. On the other hand, were Data Set B to be censored at 1,000, we would have 15 individual observations and then 5 grouped observations in the interval from 1,000 to infinity.

In insurance settings, censoring from above is fairly common. For example, if a policy pays no more than 100,000 for an accident any time the loss is above 100,000 the actual amount will be unknown but we will know that it happened. In Data Set D2 we have random censoring. Consider the fifth policy in the table. When the “other information” is not available, all that is known about the time of death is that it will be after 1.8 years. All of the policies are censored at 5 years by the nature of the policy itself. Also, note that Data Set A has been censored from above at 5. This is more common language than to say that Data Set A has some individual data and some grouped data.

When data are truncated from below, observations below a given value are not recorded. Truncation from above implies that observations above a given value are not recorded. In insurance settings, truncation from below is fairly common. If an automobile physical damage policy has a per claim deductible of 250, any losses below 250 will not come to the attention of the insurance company and so will not appear in any data sets. Data Set D2 has observations 31–40 truncated from below at varying values. Data sets may have truncation forced on them. For example, if Data Set B were to be truncated from below at 250, the first 7 observations would disappear and the remaining 13 would be unchanged.

10.2 THE EMPIRICAL DISTRIBUTION FOR COMPLETE, INDIVIDUAL DATA

As noted in Definition 10.3, the empirical distribution assigns probability $1/n$ to each data point. That works well when the value of each data point is recorded. An alternative definition follows:

Definition 10.5 The empirical distribution function is

$$F_n(x) = \frac{\text{number of observations} \leq x}{n},$$

where n is the total number of observations.

Example 10.6 Provide the empirical probability functions for the data in Data Sets A and B. For Data Set A also provide the empirical distribution function. For Data Set A assume all seven drivers who had five or more accidents had exactly five accidents.

For notation, a subscript of the sample size (or of n if the sample size is not known) will be used to indicate an empirical function. Without the subscript, the function represents the true function for the underlying random variable. For Data Set A, the estimated probability function is

$$p_{94,935}(x) = \begin{cases} 81,714/94,935 = 0.860736, & x = 0, \\ 11,306/94,935 = 0.119092, & x = 1, \\ 1,618/94,935 = 0.017043, & x = 2, \\ 250/94,935 = 0.002633, & x = 3, \\ 40/94,935 = 0.000421, & x = 4, \\ 7/94,935 = 0.000074, & x = 5, \end{cases}$$

where the values add to 0.999999 due to rounding. The distribution function is a step function with jumps at each data point.

$$F_{94,935}(x) = \begin{cases} 0/94,935 = 0.000000, & x < 0, \\ 81,714/94,935 = 0.860736, & 0 \leq x < 1, \\ 93,020/94,935 = 0.979828, & 1 \leq x < 2, \\ 94,638/94,935 = 0.996872, & 2 \leq x < 3, \\ 94,888/94,935 = 0.999505, & 3 \leq x < 4, \\ 94,928/94,935 = 0.999926, & 4 \leq x < 5, \\ 94,935/94,935 = 1.000000, & x \geq 5. \end{cases}$$

For Data Set B,

$$p_{20}(x) = \begin{cases} 0.05, & x = 27, \\ 0.05, & x = 82, \\ 0.05, & x = 115, \\ \vdots & \vdots \\ 0.05, & x = 15,743. \end{cases}$$

□

As noted in the example, the empirical model is a discrete distribution. Therefore, the derivative required to create the density and hazard rate functions cannot be taken. The closest we can come, in an empirical sense, to estimating the hazard rate function is to estimate the cumulative hazard rate function, defined as follows.

Definition 10.7 The cumulative hazard rate function is defined as

$$H(x) = -\ln S(x).$$

The name comes from the fact that, if $S(x)$ is differentiable,

$$H'(x) = -\frac{S'(x)}{S(x)} = \frac{f(x)}{S(x)} = h(x)$$

and then

$$H(x) = \int_{-\infty}^x h(y) dy.$$

The distribution function can be obtained from $F(x) = 1 - S(x) = 1 - e^{-H(x)}$. Therefore, estimating the cumulative hazard function provides an alternative way to estimate the distribution function.

In order to define empirical estimates, some additional notation is needed. For a sample of size n , let $y_1 < y_2 < \dots < y_k$ be the k unique values that appear in the sample, where k must be less than or equal to n . Let s_j be the number of times the observation y_j appears in the sample. Thus, $\sum_{j=1}^k s_j = n$. Also of interest is the number of observations in the data set that are greater than or equal to a given value. Both the observations and the number of observations are referred to as the risk set. Let $r_j = \sum_{i=j}^k s_i$ be the number of observations greater than or equal to y_j . Using this notation, the empirical distribution function is

$$F_n(x) = \begin{cases} 0, & x < y_1, \\ 1 - \frac{r_j}{n}, & y_{j-1} \leq x < y_j, \quad j = 2, \dots, k, \\ 1, & x \geq y_k. \end{cases}$$

Example 10.8 Consider a data set containing the numbers 1.0, 1.3, 1.5, 1.5, 2.1, 2.1, 2.1, and 2.8. Determine the quantities described in the previous paragraph and then obtain the empirical distribution function.

There are five unique values and thus $k = 5$. Values of y_j , s_j , and r_j are given in Table 10.6.

Table 10.6 Values for Example 10.8

j	y_j	s_j	r_j
1	1.0	1	8
2	1.3	1	7
3	1.5	2	6
4	2.1	3	4
5	2.8	1	1

$$F_8(x) = \begin{cases} 0, & x < 1.0, \\ 1 - \frac{7}{8} = 0.125, & 1.0 \leq x < 1.3, \\ 1 - \frac{6}{8} = 0.250, & 1.3 \leq x < 1.5, \\ 1 - \frac{4}{8} = 0.500, & 1.5 \leq x < 2.1, \\ 1 - \frac{1}{8} = 0.875, & 2.1 \leq x < 2.8, \\ 1, & x \geq 2.8. \end{cases} \quad \square$$

Definition 10.9 The Nelson-Åalen estimate ([1],[99]) of the cumulative hazard rate function is

$$\hat{H}(x) = \begin{cases} 0, & x < y_1, \\ \sum_{i=1}^{j-1} \frac{s_i}{r_i}, & y_{j-1} \leq x < y_j, \quad j = 2, \dots, k, \\ \sum_{i=1}^k \frac{s_i}{r_i}, & x \geq y_k. \end{cases}$$

Because this is a step function, its derivatives (which would provide an estimate of the hazard rate function) are not interesting. An intuitive derivation of this estimator can be found on page 302.

Example 10.10 Determine the Nelson-Åalen estimate for the previous example.

We have

$$\hat{H}(x) = \begin{cases} 0, & x < 1.0, \\ \frac{1}{8} = 0.125, & 1.0 \leq x < 1.3, \\ 0.125 + \frac{1}{7} = 0.268, & 1.3 \leq x < 1.5, \\ 0.268 + \frac{2}{6} = 0.601, & 1.5 \leq x < 2.1, \\ 0.601 + \frac{3}{4} = 1.351, & 2.1 \leq x < 2.8, \\ 1.351 + \frac{1}{1} = 2.351, & x \geq 2.8. \end{cases}$$

Table 10.7 Data for Example 10.11

j	y_j	s_j	r_j	$S_{30}(x)$	$\hat{H}(x)$	$\hat{S}(x) = e^{-\hat{H}(x)}$
1	0.8	1	30	$\frac{29}{30} = 0.9667$	$\frac{1}{30} = 0.0333$	0.9672
2	2.9	2	29	$\frac{27}{30} = 0.9000$	$0.0333 + \frac{2}{29} = 0.1023$	0.9028
3	3.1	1	27	$\frac{26}{30} = 0.8667$	$0.1023 + \frac{1}{27} = 0.1393$	0.8700
4	4.0	1	26	$\frac{25}{30} = 0.8333$	$0.1393 + \frac{1}{26} = 0.1778$	0.8371
5	4.8	2	25	$\frac{23}{30} = 0.7667$	$0.1778 + \frac{2}{25} = 0.2578$	0.7727

These values can be used to produce an alternative estimate of the distribution function via exponentiation. For example, for $1.5 \leq x < 2.1$, the estimate is $\hat{F}(x) = 1 - e^{-0.601} = 0.452$, which is not equal to the empirical estimate of 0.5. When a function is estimated by other methods, the function has a caret (hat) placed on it. \square

Example 10.11 Determine the empirical survival function and Nelson-Åalen estimate of the cumulative hazard rate function for the time to death for Data Set D1. Estimate the survival function from the Nelson-Åalen estimate. Assume that the death time is known for those who surrender.

The calculations are in Table 10.7. For the empirical functions, the values given are for the interval from (and including) the current y value to (but not including) the next y value. \square

For this particular problem, where it is known that all policyholders terminate at time 5, results past 5 are not interesting. The methods of obtaining an empirical distribution that have been introduced work only when the individual observations are available and there is no truncation or censoring. The following chapter will introduce modifications for those situations.

10.2.1 Exercises

10.1 Obtain the empirical distribution function and the Nelson-Åalen estimate of the distribution function for the time to surrender using Data Set D1. Assume that the surrender time is known for those who die.

10.2 The data in Table 10.8 are from *Loss Distributions* [59], p. 128. It represents the total damage done by 35 hurricanes between the years 1949 and 1980. The losses have been adjusted for inflation (using the Residential Construction Index) to be in 1981 dollars. The entries represent all hurricanes for which the trended loss was in excess of 5,000,000.

Table 10.8 Trended hurricane losses

Year	Loss (10^3)	Year	Loss (10^3)	Year	Loss (10^3)
1964	6,766	1964	40,596	1975	192,013
1968	7,123	1949	41,409	1972	198,446
1971	10,562	1959	47,905	1964	227,338
1956	14,474	1950	49,397	1960	329,511
1961	15,351	1954	52,600	1961	361,200
1966	16,983	1973	59,917	1969	421,680
1955	18,383	1980	63,123	1954	513,586
1958	19,030	1964	77,809	1954	545,778
1974	25,304	1955	102,942	1970	750,389
1959	29,112	1967	103,217	1979	863,881
1971	30,146	1957	123,680	1965	1,638,000
1976	33,727	1979	140,136		

The federal government is considering funding a program that would provide 100% payment for all damages for any hurricane causing damage in excess of 5,000,000. You have been asked to make some preliminary estimates.

- Estimate the mean, standard deviation, coefficient of variation, and skewness for the population of hurricane losses.
- Estimate the first and second limited moments at 500,000,000.

10.3 (*) There have been 30 claims recorded in a random sampling of claims. There were 2 claims for 2,000, 6 for 4,000, 12 for 6,000, and 10 for 8,000. Determine the empirical skewness coefficient.

10.3 EMPIRICAL DISTRIBUTIONS FOR GROUPED DATA

For grouped data, construction of the empirical distribution as defined previously is not possible. However, it is possible to approximate the empirical distribution. The strategy is to obtain values of the empirical distribution function wherever possible and then connect those values in some reasonable way. For grouped data, the distribution function is usually approximated by connecting the points with straight lines. Other interpolation methods are discussed in Chapter 15. For notation, let the group boundaries be $c_0 < c_1 < \dots < c_k$, where often $c_0 = 0$ and $c_k = \infty$. The number of observations falling between c_{j-1} and c_j is denoted n_j with $\sum_{j=1}^k n_j = n$. For such data, we are able to determine the empirical distribution at each group boundary. That is, $F_n(c_j) = (1/n) \sum_{i=1}^j n_i$. Note that no rule is proposed for observations that fall on a group boundary. There is no correct approach,

but whatever approach is used, consistency in assignment of observations to groups should be used. You will note that in Data Set C it is not possible to tell how the assignments were made. If we had that knowledge, it would not affect any subsequent calculations.

Definition 10.12 For grouped data, the distribution function obtained by connecting the values of the empirical distribution function at the group boundaries with straight lines is called the *ogive*. The formula is

$$F_n(x) = \frac{c_j - x}{c_j - c_{j-1}} F_n(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}} F_n(c_j), \quad c_{j-1} \leq x \leq c_j.$$

This function is differentiable at all values except group boundaries. Therefore the density function can be obtained. To completely specify the density function, it will arbitrarily be made right-continuous.

Definition 10.13 For grouped data, the empirical density function can be obtained by differentiating the ogive. The resulting function is called a *histogram*. The formula is

$$f_n(x) = \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} = \frac{n_j}{n(c_j - c_{j-1})}, \quad c_{j-1} \leq x < c_j.$$

Many computer programs that produce histograms actually create a bar chart with bar heights proportional to n_j/n . This is acceptable if the groups have equal width, but if not, then the above formula is needed. The advantage of this approach is that the histogram is indeed a density function, and among other things, areas under the histogram can be used to obtain empirical probabilities.

Example 10.14 Construct the ogive and histogram for Data Set C.

The distribution function is

$$F_{227}(x) = \begin{cases} 0.000058150x, & 0 \leq x \leq 7,500, \\ 0.29736 + 0.000018502x, & 7,500 \leq x \leq 17,500, \\ 0.47210 + 0.000008517x, & 17,500 \leq x \leq 32,500, \\ 0.63436 + 0.000003524x, & 32,500 \leq x \leq 67,500, \\ 0.78433 + 0.000001302x, & 67,500 \leq x \leq 125,000, \\ 0.91882 + 0.000000227x, & 125,000 \leq x \leq 300,000, \\ \text{undefined,} & x > 300,000, \end{cases}$$

where, for example, for the range $32,500 \leq x \leq 67,500$ the calculation is

$$F_{227}(x) = \frac{67,500 - x}{67,500 - 32,500} \frac{170}{227} + \frac{x - 32,500}{67,500 - 32,500} \frac{198}{227}.$$

The value is undefined above 300,000 because the last interval has a width of infinity. A graph of the ogive for values up to 125,000 appears in Figure 10.1.

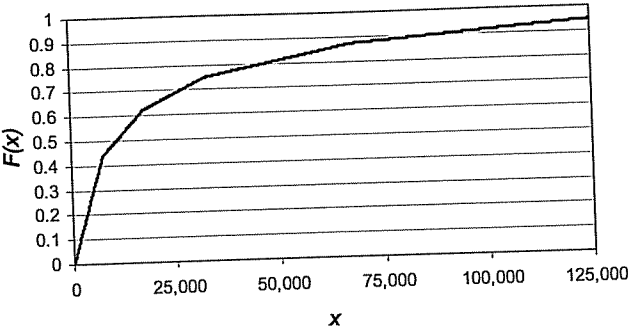


Fig. 10.1 Ogive for general liability losses.

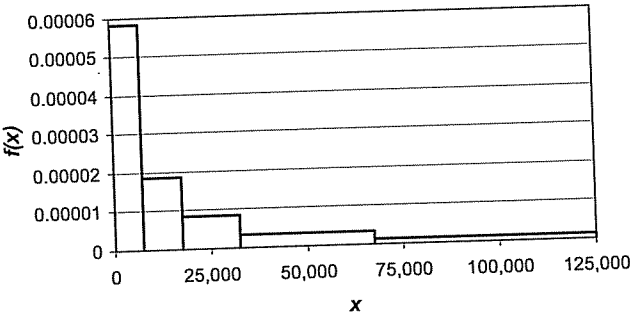


Fig. 10.2 Histogram of general liability losses.

The derivative is simply a step function with the following values.

$$f_{227}(x) = \begin{cases} 0.000058150, & 0 \leq x < 7,500, \\ 0.000018502, & 7,500 \leq x < 17,500, \\ 0.000008517, & 17,500 \leq x < 32,500, \\ 0.000003524, & 32,500 \leq x < 67,500, \\ 0.000001302, & 67,500 \leq x < 125,000, \\ 0.000000227, & 125,000 \leq x < 300,000, \\ \text{undefined,} & x \geq 300,000. \end{cases}$$

A graph of the function up to 125,000 appears in Figure 10.2.

Table 10.9 Data for Exercise 10.4

Payment range	Number of payments
0-25	6
25-50	24
50-75	30
75-100	31
100-150	57
150-250	80
250-500	85
500-1000	54
1000-2000	15
2000-4000	10
Over 4000	0

10.3.1 Exercises

10.4 Construct the ogive and histogram for the data in Table 10.9.

10.5 (*) The following 20 wind losses (in millions of dollars) were recorded in one year:

1 1 1 1 1 2 2 3 3 4
6 6 8 10 13 14 15 18 22 25

- (a) Construct an ogive based on using class boundaries at 0.5, 2.5, 8.5, 15.5, and 29.5.
- (b) Construct a histogram using the same boundaries as in part (a).

10.6 The data in Table 10.10 are from Herzog and Laverty [54]. A certain class of 15-year mortgages was followed from issue until December 31, 1993. The issues were split into those that were refinances of existing mortgages and those that were original issues. Each entry in the table provides the number of issues and the percentage of them that were still in effect after the indicated number of years. Draw as much of the two ogives (on the same graph) as is possible from the data. Does it appear from the ogives that the lifetime variable (time to mortgage termination) has a different distribution for refinanced versus original issues?

10.7 (*) The data in Table 10.11 were collected (units are millions of dollars). Construct the histogram.

10.8 (*) Forty losses have been observed. Sixteen are between 1 and $\frac{4}{3}$ and those 16 losses total 20. Ten losses are between $\frac{4}{3}$ and 2 with a total of 15. Ten more are between 2 and 4 with a total of 35. The remaining 4 losses

Table 10.10 Data for Exercise 10.6

Years	Refinances		Original	
	No. issued	Survived	No. issued	Survived
1.5	42,300	99.97	12,813	99.88
2.5	9,756	99.82	18,787	99.43
3.5	1,550	99.03	22,513	98.81
4.5	1,256	98.41	21,420	98.26
5.5	1,619	97.78	26,790	97.45

Table 10.11 Data for Exercise 10.7

Loss	No. of observations
0-2	25
2-10	10
10-100	10
100-1,000	5

are greater than 4. Using the empirical model based on these observations, determine $E(X \wedge 2)$.

10.9 (*) A sample of size 2,000 contains 1,700 observations that are no greater than 6,000, 30 that are greater than 6,000 but no greater than 7,000, and 270 that are greater than 7,000. The total amount of the 30 observations that are between 6,000 and 7,000 is 200,000. The value of $E(X \wedge 6,000)$ for the empirical distribution associated with these observations is 1,810. Determine $E(X \wedge 7,000)$ for the empirical distribution.

11

Estimation for modified data

11.1 POINT ESTIMATION

It is not unusual for data to be incomplete due to censoring or truncation. The formal definitions are as follows.

Definition 11.1 An observation is *truncated from below* (also called *left truncated*) at d if when it is below d it is not recorded but when it is above d it is recorded at its observed value.

An observation is *truncated from above* (also called *right truncated*) at u if when it is above u it is not recorded but when it is below u it is recorded at its observed value.

An observation is *censored from below* (also called *left censored*) at d if, when it is below d it is recorded as being equal to d but when it is above d it is recorded at its observed value.

An observation is *censored from above* (also called *right censored*) at u if, when it is above u it is recorded as being equal to u but when it is below u it is recorded at its observed value.

The most common occurrences are left truncation and right censoring. Left truncation occurs when an ordinary deductible of d is applied. When a policyholder has a loss below d , he or she knows no benefits will be paid and so

Loss Models: From Data to Decisions, Second Edition.

By Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot

ISBN 0-471-21577-5 Copyright © 2004 John Wiley & Sons, Inc.

does not inform the insurer. When the loss is above d , the amount of the loss will be reported. A policy limit is an example of right censoring. When the amount of the loss exceeds u , benefits beyond that value are not paid and so the exact value is not recorded. However, it is known that a loss of at least u has occurred.

When constructing a mortality table, it is impractical to follow people from birth to death. It is more common to follow a group of people of varying ages for a few years. When a person joins a study, he or she is alive at that time. This person's age at death must be at least as great as the age at entry to the study and thus has been left truncated. If the person is alive when the study ends, right censoring has occurred. The person's age at death is not known, but it is known that it is at least as large as the age when the study ended. Right censoring also affects those who leave the study prior to its end due to surrender.

Because left truncation and right censoring are the most common occurrences in actuarial work, they are the only cases that will be covered in this chapter. To save words, *truncated* will always mean truncated from below and *censored* will always mean censored from above.

When trying to construct an empirical distribution from truncated or censored data, the first task is to create notation to summarize the data. For individual data, there are three facts that are needed. First is the truncation point for that observation. Let that value be d_j for the j th observation. If there was no truncation, $d_j = 0$. The second is the observation itself. The notation used will depend on whether or not that observation was censored. If it was not censored, let its value be x_j . If it was censored, let its value be u_j . When this subject is presented more formally, a distinction is made between the case where the censoring point is known in advance and where it is not. For example, a liability insurance policy with a policy limit usually has the censoring point known prior to the receipt of any claims. On the other hand, in a mortality study of insured lives, those that surrender their policy do so at an age that was not known when the policy was sold. In this chapter no distinction will be made between the two cases.

To perform the estimation, the raw data must be summarized in a useful manner. The most interesting values are the uncensored observations. Let $y_1 < y_2 < \dots < y_k$ be the k unique values of the x_j s that appear in the sample, where k must be less than or equal to the number of uncensored observations. Let s_j be the number of times the uncensored observation y_j appears in the sample. The final important quantity is the risk set at the j th ordered observation y_j and is denoted r_j . When thinking in terms of a mortality study, the risk set comprises the individuals who are under observation at that age. Included are all who die (have x values) at that age or later and all who are censored (have u values) at that age or later. However, those who are first observed (have d values) at that age or later were not under observation

at that time. The formula is

$$r_j = (\text{number of } x_{is} \geq y_j) + (\text{number of } u_{is} \geq y_j) - (\text{number of } d_{is} \geq y_j).$$

Alternatively, because the total number of d_{is} is equal to the total number of x_{is} and u_{is} , we also have

$$r_j = (\text{number of } d_{is} < y_j) - (\text{number of } x_{is} < y_j) - (\text{number of } u_{is} < y_j). \quad (11.1)$$

This latter version is a bit easier to conceptualize because it includes all who have entered the study prior to the given age less those who have already left. The key point is that the risk set is the number of people observed alive at age y_j . If the data are loss amounts, the risk set is the number of policies with observed loss amounts (either the actual amount or the maximum amount due to a policy limit) greater than or equal to y_j less those with deductibles greater than or equal to y_j . This also leads to a recursive version of the formula,

$$\begin{aligned} r_j = & r_{j-1} + (\text{number of } d_{is} \text{ between } y_{j-1} \text{ and } y_j) \\ & - (\text{number of } x_{is} \text{ between } y_{j-1} \text{ and } y_j) \\ & - (\text{number of } u_{is} \text{ between } y_{j-1} \text{ and } y_j), \end{aligned} \quad (11.2)$$

where *between* is interpreted to mean greater than or equal to y_{j-1} and less than y_j and r_0 is set equal to 0.

Example 11.2 Using Data Set D2, calculate the above quantities using both (11.1) and (11.2).

The calculations appear in Tables 11.1 and 11.2. □

Despite all the work we have done to this point, we have yet to produce an estimator of the survival function. The one most commonly used is called the **Kaplan–Meier product-limit Estimator** [70]. Begin with $S(0) = 1$. Because no one died prior to y_1 , the survival function remains at 1 until this value. Thinking conditionally, just before y_1 , there were r_1 people available to die, of which s_1 did so. Thus, the probability of surviving past y_1 is $(r_1 - s_1)/r_1$. This becomes the value of $S(y_1)$ and the survival function remains at that value until y_2 . Again, thinking conditionally, the new survival value at y_2 is $S(y_1)(r_2 - s_2)/r_2$. The general formula is

$$S_n(t) = \begin{cases} 1, & 0 \leq t < y_1, \\ \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right), & y_{j-1} \leq t < y_j, j = 2, \dots, k, \\ \prod_{i=1}^k \left(\frac{r_i - s_i}{r_i} \right) \text{ or } 0, & t \geq y_k. \end{cases}$$

Table 11.1 Values for Example 11.2

i	d_i	x_i	u_i	i	d_i	x_i	u_i
1	0	—	0.1	16	0	4.8	—
2	0	—	0.5	17	0	—	4.8
3	0	—	0.8	18	0	—	4.8
4	0	0.8	—	19–30	0	—	5.0
5	0	—	1.8	31	0.3	—	5.0
6	0	—	1.8	32	0.7	—	5.0
7	0	—	2.1	33	1.0	4.1	—
8	0	—	2.5	34	1.8	3.1	—
9	0	—	2.8	35	2.1	—	3.9
10	0	2.9	—	36	2.9	—	5.0
11	0	2.9	—	37	2.9	—	4.8
12	0	—	3.9	38	3.2	4.0	—
13	0	4.0	—	39	3.4	—	5.0
14	0	—	4.0	40	3.9	—	5.0
15	0	—	4.1				

Table 11.2 Risk set calculations for Example 11.2

j	y_j	s_j	r_j
1	0.8	1	$32 - 0 - 2 = 30$ or $0 + 32 - 0 - 2 = 30$
2	2.9	2	$35 - 1 - 8 = 26$ or $30 + 3 - 1 - 6 = 26$
3	3.1	1	$37 - 3 - 8 = 26$ or $26 + 2 - 2 - 0 = 26$
4	4.0	2	$40 - 4 - 10 = 26$ or $26 + 3 - 1 - 2 = 26$
5	4.1	1	$40 - 6 - 11 = 23$ or $26 + 0 - 2 - 1 = 23$
6	4.8	1	$40 - 7 - 12 = 21$ or $23 + 0 - 1 - 1 = 21$

If $s_k = r_k$, then $S(t) = 0$ for $t \geq y_k$ makes sense. Everyone in the sample has died by that value and so, empirically, survival past that age is not possible. However, due to censoring, it is possible that at the age of the last death there were still people alive but all were censored prior to death. We know that survival past the last observed death age is possible, but there is no empirical data available to complete the survival function. One option (the first one used in the above formula) is to keep the function at its last value. This is clearly the largest reasonable choice. Another option is to declare the function to be zero past the last observed age, whether it is an observed death or a censored age. This is the smallest reasonable choice and makes it possible to calculate moments. An intermediate option is to use an exponential curve to reduce the value from its current level to zero. Let $w = \max\{x_1, \dots, x_n, u_1, \dots, u_n\}$.

Then, for $t \geq w$,

$$S_n(t) = e^{(t/w) \ln s^*} = (s^*)^{t/w}, \text{ where } s^* = \prod_{i=1}^k \left(\frac{r_i - s_i}{r_i} \right).$$

There is an alternative method of obtaining the values of s_j and r_j that is more suitable for Excel® spreadsheet work.¹ The steps are as follows:

1. There should be one row for each data point. The points need not be in any particular order.
2. Each row should have three entries. The first entry should be d_j , the second entry should be u_j or x_j (there can be only one of these). The third entry should be the letter "x" (without the quotes) if the second entry is an x -value (an observed value) and should be the letter "u" if the second entry is a u -value (a censored value). Assume, for example, that the d_j s occupy cells B6:B45, the u_j s and x_j s occupy C6:C45, and the x/u letters occupy D6:D45.
3. Create the columns for x , r , and s as follows.
4. For the ordered observed values (say they should begin in cell F2, start with the lowest d -value. The formula in F2 is =MIN(B6:B45).
5. Then in cell F3 enter the formula

$$=\text{MIN}(\text{IF}(\text{C\$6:C\$45}>\text{F2}, \text{IF}(\text{D\$6:D\$45}="x", \text{C\$6:C\$45}, 1\text{E}36), 1\text{E}36)).$$

Because this is an array formula, it must be entered with Ctrl-Shift-Enter. Copy this formula into cells F4, F5, and so on until the value 1E36 appears. Column F should now contain the unique, ordered, y -values.

6. In cell G3 enter the formula

$$=\text{COUNTIF}(\text{B\$6:B\$45}, "<"\&\text{F3}) - \text{COUNTIF}(\text{C\$6:C\$45}, "<"\&\text{F3}).$$

Copy this formula into cells G4, G5, and so on until values appear across from all but the last value in column F. This column contains the risk set values.

7. In cell H3 enter the formula

$$=\text{SUM}(\text{IF}(\text{C\$6:C\$45}=\text{F3}, \text{IF}(\text{D\$6:D\$45}="x", 1, 0), 0)).$$

¹This scheme was devised by Charles Thayer and improved by Margie Rosenberg. It is a great improvement over the author's scheme presented in earlier drafts. These instructions work with Office XP and should work in a similar manner for other versions.

Enter this array formula with Ctrl-Shift-Enter and then copy it into H4, H5, and so on to match the rows of column G. This column contains the s -values.

8. Begin the calculation of $S(y)$ by entering a 1 in cell I2
9. Calculate the next $S(y)$ value by entering the following formula in cell I3.

$$=I2*(G3 - H3)/G3.$$

Then copy this formula into cells I4, I5, and so on to complete the process.

Example 11.3 Determine the Kaplan-Meier estimate for Data Set D2.

Based on the previous example, we have

$$S_{40}(t) = \begin{cases} 1, & 0 \leq t < 0.8, \\ \frac{30-1}{30} = 0.9667, & 0.8 \leq t < 2.9, \\ 0.9667 \frac{26-2}{26} = 0.8923, & 2.9 \leq t < 3.1, \\ 0.8923 \frac{26-1}{26} = 0.8580, & 3.1 \leq t < 4.0, \\ 0.8580 \frac{26-2}{26} = 0.7920, & 4.0 \leq t < 4.1, \\ 0.7920 \frac{23-1}{23} = 0.7576, & 4.1 \leq t < 4.8, \\ 0.7576 \frac{21-1}{21} = 0.7215, & 4.8 \leq t < 5.0, \\ 0.7215 \text{ or } 0 \text{ or } 0.7215^{t/5.0}, & t \geq 5.0. \end{cases} \quad \square$$

An alternative to the Kaplan-Meier estimate is a modification of the Nelson-Åalen estimate introduced earlier. As before, this method directly estimates the cumulative hazard rate function. The following is an intuitive derivation of this estimator. Let $r(t)$ be the risk set at any time t and let $h(t)$ be the hazard rate function. Also, let $s(t)$ be the expected total number of observed deaths prior to time t . It is reasonable to conclude that

$$s(t) = \int_0^t r(u)h(u)du.$$

Taking derivatives,

$$ds(t) = r(t)h(t)dt.$$

Then,

$$\frac{ds(t)}{r(t)} = h(t)dt.$$

Integrating both sides yields

$$\int_0^t \frac{ds(u)}{r(u)} = \int_0^t h(u)du = H(t).$$

Now replace the true expected count $s(t)$ by $\hat{s}(t)$, the observed number of deaths by time t . It is a step function, increasing by s_i at each death time. Therefore, the left-hand side becomes

$$\sum_{t_i \leq t} \frac{s_i}{r_i},$$

which defines the estimator, $\hat{H}(t)$. The Nelson-Åalen estimator is

$$\hat{H}(t) = \begin{cases} 0, & 0 \leq t < y_1, \\ \sum_{i=1}^{j-1} \frac{s_i}{r_i}, & y_{j-1} \leq t < y_j, j = 2, \dots, k, \\ \sum_{i=1}^k \frac{s_i}{r_i}, & t \geq y_k, \end{cases}$$

and then

$$\hat{S}(t) = e^{-\hat{H}(t)}.$$

For $t \geq w$, alternative estimates are $\hat{S}(t) = 0$ and $\hat{S}(t) = \hat{S}(y_k)^{t/w}$.

Example 11.4 Determine the Nelson-Åalen estimate of the survival function for Data Set D2.

$$\hat{H}(t) = \begin{cases} 0, & 0 \leq t < 0.8, \\ \frac{1}{30} = 0.0333, & 0.8 \leq t < 2.9, \\ 0.0333 + \frac{2}{26} = 0.1103, & 2.9 \leq t < 3.1, \\ 0.1103 + \frac{1}{26} = 0.1487, & 3.1 \leq t < 4.0, \\ 0.1487 + \frac{2}{26} = 0.2256, & 4.0 \leq t < 4.1, \\ 0.2256 + \frac{1}{23} = 0.2691, & 4.1 \leq t < 4.8, \\ 0.2691 + \frac{1}{21} = 0.3167, & t \geq 4.8. \end{cases}$$

$$\hat{S}(t) = \begin{cases} 1, & 0 \leq t < 0.8, \\ e^{-0.0333} = 0.9672, & 0.8 \leq t < 2.9, \\ e^{-0.1103} = 0.8956, & 2.9 \leq t < 3.1, \\ e^{-0.1487} = 0.8618, & 3.1 \leq t < 4.0, \\ e^{-0.2256} = 0.7980, & 4.0 \leq t < 4.1, \\ e^{-0.2691} = 0.7641, & 4.1 \leq t < 4.8, \\ e^{-0.3167} = 0.7285, & 4.8 \leq t < 5.0, \\ 0.7285 \text{ or } 0 \text{ or } 0.7285^{t/5.0}, & t \geq 5.0. \end{cases} \quad \square$$

It is important to note that when the data are truncated the resulting distribution function is the distribution function for payments given that they are above the smallest truncation point (that is, the smallest d value). Empirically, there is no information about observations below that value, and thus there can be no information for that range. It should be noted that all the notation and formulas in this section are consistent with those in Section 10.2. If it turns out that there was no censoring or truncation, using the formulas in this section will lead to the same results as when using the empirical formulas in Section 10.2.

11.1.1 Exercises

11.1 Repeat Example 11.2, treating "surrender" as "death." The easiest way to do this is to reverse the x and u labels and then use the above formula. In this case death produces censoring because those who die are lost to observation and thus their surrender time is never observed. Treat those who lasted the entire five years as surrenders at that time.

11.2 Determine the Kaplan–Meier estimate for the time to surrender for Data Set D2. Treat those who lasted the entire five years as surrenders at that time.

11.3 Determine the Nelson–Åalen estimate of $H(t)$ and $S(t)$ for Data Set D2 where the variable is time to surrender.

11.4 Determine the Kaplan–Meier and Nelson–Åalen estimates of the distribution function of the amount of a workers compensation loss. First do this using the raw data from Data Set B. Then repeat the exercise, modifying the data by left truncation at 100 and right censoring at 1,000.

11.5 (*) You are given the following times of first claim for five randomly selected auto insurance policies: 1, 2, 3, 4, 5. You are later told that one of the five times given is actually the time of policy lapse but you are not told which one. The smallest product-limit estimate of $S(4)$, the probability that the first claim occurs after time 4, would result if which of the given times arose from the lapsed policy?

11.6 (*) For a mortality study with right censored data, you are given the information in Table 11.3. Calculate the estimate of the survival function at time 12 using the Nelson–Åalen estimate.

11.7 (*) Three hundred mice were observed at birth. An additional 20 mice were first observed at age 2 (days) and 30 more were first observed at age 4. There were 6 deaths at age 1, 10 at age 3, 10 at age 4, a at age 5, b at age 9, and 6 at age 12. In addition, 45 mice were lost to observation at age 7, 35 at age 10, and 15 at age 13. The following product-limit estimates were obtained: $S_{350}(7) = 0.892$ and $S_{350}(13) = 0.856$. Determine the values of a and b .

Table 11.3 Data for Exercise 11.6

Time t_j	Number of deaths, s_j	Number at risk, r_j
5	2	15
7	1	12
10	1	10
12	2	6

11.8 (*) Let n be the number of lives observed from birth. None were censored and no two lives died at the same age. At the time of the ninth death, the Nelson–Åalen estimate of the cumulative hazard rate is 0.511 and at the time of the tenth death it is 0.588. Estimate the value of the survival function at the time of the third death.

11.9 (*) All members of a study joined at birth; however, some may leave the study by means other than death. At the time of the third death, there was one death (that is, $s_3 = 1$); at the time of the fourth death there were two deaths; and at the time of the fifth death there was one death. The following product-limit estimates were obtained: $S_n(y_3) = 0.72$, $S_n(y_4) = 0.60$, and $S_n(y_5) = 0.50$. Determine the number of censored observations between times y_4 and y_5 . Assume no observations were censored at the death times.

11.2 MEANS, VARIANCES, AND INTERVAL ESTIMATION

When all of the information is available, working with the empirical estimate of the survival function is straightforward.

Example 11.5 *Demonstrate that for complete data the empirical estimator of the survival function is unbiased and consistent.*

Recall that the empirical estimate of $S(x)$ is $S_n(x) = Y/n$, where Y is the number of observations in the sample that are greater than x . Then Y must have a binomial distribution with parameters n and $S(x)$. Then,

$$E[S_n(x)] = E\left(\frac{Y}{n}\right) = \frac{nS(x)}{n} = S(x),$$

demonstrating that the estimator is unbiased. The variance is

$$\text{Var}[S_n(x)] = \text{Var}\left(\frac{Y}{n}\right) = \frac{S(x)[1 - S(x)]}{n},$$

which has a limit of zero, thus verifying consistency. \square

In order to make use of the result, the best we can do for the variance is estimate it. It is unlikely we know the value of $S(x)$ because that is the quantity we are trying to estimate. The estimated variance is given by

$$\widehat{\text{Var}}[S_n(x)] = \frac{S_n(x)[1 - S_n(x)]}{n}.$$

The same results hold for empirically estimated probabilities. Let $p = \Pr(a < X \leq b)$. The empirical estimate of p is $\hat{p} = S_n(a) - S_n(b)$. Arguments similar to those used in the example above verify that \hat{p} is unbiased and consistent, with $\text{Var}(\hat{p}) = p(1 - p)/n$.

When doing mortality studies or evaluating the effect of deductibles, we sometimes are more interested in estimating conditional quantities.

Example 11.6 Using the full information from the observations in Data Set D1, empirically estimate q_2 and estimate the variance of this estimator.

For this data set, $n = 30$. By duration 2, one had died, and by duration 3, three had died. Thus, $S_{30}(2) = \frac{29}{30}$ and $S_{30}(3) = \frac{27}{30}$. The empirical estimate is

$$\hat{q}_2 = \frac{S_{30}(2) - S_{30}(3)}{S_{30}(2)} = \frac{2}{29}.$$

The challenge is in trying to evaluate the mean and variance of this estimator. Let X be the number of deaths between durations 0 and 2 and let Y be the number of deaths between durations 2 and 3. Then $\hat{q}_2 = Y/(30 - X)$. It should be clear that it is not possible to evaluate $E(\hat{q}_2)$ because with positive probability X will equal 30 and the estimator will not be defined.² The usual solution is to obtain a conditional estimate of the variance. That is, given there were 29 people alive at duration 2, determine the variance. Then the only random quantity is Y and we have

$$\widehat{\text{Var}}[\hat{q}_2 | S_{30}(2) = \frac{29}{30}] = \frac{(2/29)(27/29)}{29}.$$

□

In general, let n be the initial sample, n_x the number alive at age x , and n_y the number alive at age y . Then,

$$\widehat{\text{Var}}_{(y-x)}\hat{q}_x | n_x = \widehat{\text{Var}}_{(y-x)}\hat{p}_x | n_x = \frac{(n_x - n_y)(n_y)}{n_x^3}.$$

²This is a situation where the Bayesian approach (introduced in Section 12.4) works better. Bayesian analyses proceed from the data as observed and are not concerned with other values that might have been observed. If $X = 30$, there is no estimate to worry about, while if $X < 30$, analysis can proceed.

Example 11.7 Using Data Set B, empirically estimate the probability that a payment will be at least 1,000 when there is a deductible of 250.

Empirically, there were 13 losses above the deductible, of which 4 exceeded 1,250 (the loss that produces a payment of 1,000). The empirical estimate is $\frac{4}{13}$. Using survival function notation, the estimator is $S_{20}(1250)/S_{20}(250)$. Once again, only a conditional variance can be obtained. The estimated variance is $4(9)/13^3$. □

For grouped data, there is no problem if the survival function is to be estimated at a boundary. For interpolated values using the ogive, it is a bit more complex.

Example 11.8 Determine the expected value and variance of the estimators of the survival function and density function using the ogive and histogram, respectively.

Suppose the value of x is between the boundaries c_{j-1} and c_j . Let Y be the number of observations at or below c_{j-1} and let Z be the number of observations above c_{j-1} and at or below c_j . Then

$$S_n(x) = 1 - \frac{Y(c_j - c_{j-1}) + Z(x - c_{j-1})}{n(c_j - c_{j-1})}$$

and

$$\begin{aligned} E[S_n(x)] &= 1 - \frac{n[1 - S(c_{j-1})](c_j - c_{j-1}) + n[S(c_{j-1}) - S(c_j)](x - c_{j-1})}{n(c_j - c_{j-1})} \\ &= S(c_{j-1})\frac{c_j - x}{c_j - c_{j-1}} + S(c_j)\frac{x - c_{j-1}}{c_j - c_{j-1}}. \end{aligned}$$

This estimator is biased (although it is an unbiased estimator of the true interpolated value). The variance is

$$\text{Var}[S_n(x)] = \frac{(c_j - c_{j-1})^2 \text{Var}(Y) + (x - c_{j-1})^2 \text{Var}(Z) + 2(c_j - c_{j-1})(x - c_{j-1}) \text{Cov}(Y, Z)}{[n(c_j - c_{j-1})]^2},$$

where $\text{Var}(Y) = nS(c_{j-1})[1 - S(c_{j-1})]$, $\text{Var}(Z) = n[S(c_{j-1}) - S(c_j)][1 - S(c_{j-1}) + S(c_j)]$, and $\text{Cov}(Y, Z) = -n[1 - S(c_{j-1})][S(c_{j-1}) - S(c_j)]$. For the density estimate,

$$f_n(x) = \frac{Z}{n(c_j - c_{j-1})}$$

and

$$E[f_n(x)] = \frac{S(c_{j-1}) - S(c_j)}{c_j - c_{j-1}},$$

which is biased for the true density function. The variance is

$$\text{Var}[f_n(x)] = \frac{[S(c_{j-1}) - S(c_j)][1 - S(c_{j-1}) + S(c_j)]}{n(c_j - c_{j-1})^2}.$$

□

Example 11.9 For Data Set C estimate $S(10,000)$, $f(10,000)$, and the variance of your estimators.

The point estimates are

$$\begin{aligned} S_{227}(10,000) &= 1 - \frac{99(17,500 - 7,500) + 42(10,000 - 7,500)}{227(17,500 - 7,500)} = 0.51762, \\ f_{227}(10,000) &= \frac{42}{227(17,500 - 7,500)} = 0.000018502. \end{aligned}$$

The estimated variances are

$$\begin{aligned} \widehat{\text{Var}}[S_{227}(10,000)] &= \frac{1}{227(10,000)^2} \left[10,000^2 \frac{99}{227} \frac{128}{227} + 2,500^2 \frac{42}{227} \frac{185}{227} \right. \\ &\quad \left. - 2(10,000)(2,500) \frac{99}{227} \frac{42}{227} \right] \\ &= 0.00094713 \end{aligned}$$

and

$$\widehat{\text{Var}}[f_{227}(10,000)] = \frac{\frac{42}{227} \frac{185}{227}}{227(10,000)^2} = 6.6427 \times 10^{-12}.$$

□

Discrete data such as in Data Set A can be considered a special form of grouped data. Each discrete possibility is similar to an interval.

Example 11.10 Demonstrate that for a discrete random variable the empirical estimator of a particular outcome is unbiased and consistent and derive its variance.

Let N_j be the number of times the value x_j was observed in the sample. Then N_j has a binomial distribution with parameters n and $p(x_j)$. The empirical estimator is $p_n(x_j) = N_j/n$ and

$$E[p_n(x_j)] = E\left(\frac{N_j}{n}\right) = \frac{np(x_j)}{n} = p(x_j),$$

demonstrating that it is unbiased. Also,

$$\text{Var}[p_n(x_j)] = \text{Var}\left(\frac{N_j}{n}\right) = \frac{np(x_j)[1 - p(x_j)]}{n^2} = \frac{p(x_j)[1 - p(x_j)]}{n},$$

which goes to zero as $n \rightarrow \infty$, demonstrating consistency. □

Example 11.11 For Data Set A determine the empirical estimate of $p(2)$ and estimate its variance.

The empirical estimate is

$$p_{94,935}(2) = \frac{1,618}{94,935} = 0.017043$$

and its estimated variance is

$$\frac{0.017043(0.982957)}{94,935} = 1.76466 \times 10^{-7}.$$

□

It is possible to use the variances to construct confidence intervals for the unknown probability.

Example 11.12 Use (9.3) and (9.4) to construct approximate 95% confidence intervals for $p(2)$ using Data Set A.

From (9.3),

$$0.95 = \Pr\left(-1.96 \leq \frac{p_n(2) - p(2)}{\sqrt{p(2)[1 - p(2)]/n}} \leq 1.96\right).$$

Solve this by making the inequality an equality and then squaring both sides to obtain [dropping the argument of (2) for simplicity],

$$\begin{aligned} \frac{(p_n - p)^2 n}{p(1 - p)} &= 1.96^2, \\ np_n^2 - 2npnp_n + np^2 &= 1.96^2 p - 1.96^2 p^2, \\ 0 &= (n + 1.96^2)p^2 - (2np_n + 1.96^2)p + np_n^2. \end{aligned}$$

The solution is

$$p = \frac{2np_n + 1.96^2 \pm \sqrt{(2np_n + 1.96^2)^2 - 4(n + 1.96^2)np_n^2}}{2(n + 1.96^2)},$$

which provides the two endpoints of the confidence interval. Inserting the numbers from Data Set A ($p_n = 0.017043$, $n = 94,935$) produces a confidence interval of (0.016239, 0.017886).

Equation (9.4) provides the confidence interval directly as

$$p_n \pm 1.96 \sqrt{\frac{p_n(1 - p_n)}{n}}.$$

Inserting the numbers from Data Set A gives 0.017043 ± 0.000823 for an interval of $(0.016220, 0.017866)$. The answers for the two methods are very similar, which is to be expected when the sample size is large. The results are reasonable because it is well known that the normal distribution is a good approximation to the binomial. \square

When data are censored or truncated, the matter becomes more complex. Counts no longer have the binomial distribution and therefore the distribution of the estimator is harder to obtain. While there are proofs available to back up the results presented here, they will not be provided. Instead, an attempt will be made to indicate why the results are reasonable.

Consider the Kaplan-Meier product-limit estimator of $S(t)$. It is the product of a number of terms of the form $(r_j - s_j)/r_j$, where r_j was viewed as the number available to die at age y_j and s_j is the number who actually did so. Assume that the death ages and the number available to die are fixed, so that the value of s_j is the only random quantity. As a random variable, S_j has a binomial distribution based on a sample of r_j lives and success probability $[S(y_{j-1}) - S(y_j)]/S(y_{j-1})$. The probability arises from the fact that those available to die were known to be alive at the previous death age. For one of these terms,

$$E\left(\frac{r_j - S_j}{r_j}\right) = \frac{r_j - r_j[S(y_{j-1}) - S(y_j)]/S(y_{j-1})}{r_j} = \frac{S(y_j)}{S(y_{j-1})}.$$

That is, this ratio is an unbiased estimator of the probability of surviving from one death age to the next one. Furthermore,

$$\begin{aligned} \text{Var}\left(\frac{r_j - S_j}{r_j}\right) &= \frac{r_j \frac{S(y_{j-1}) - S(y_j)}{S(y_{j-1})} \left[1 - \frac{S(y_{j-1}) - S(y_j)}{S(y_{j-1})}\right]}{r_j^2} \\ &= \frac{[S(y_{j-1}) - S(y_j)]S(y_j)}{r_j S(y_{j-1})^2}. \end{aligned}$$

Now consider the estimated survival probability at one of the death ages. Its expected value is

$$\begin{aligned} E[\hat{S}(y_j)] &= E\left[\prod_{i=1}^j \left(\frac{r_i - S_i}{r_i}\right)\right] = \prod_{i=1}^j E\left(\frac{r_i - S_i}{r_i}\right) \\ &= \prod_{i=1}^j \frac{S(y_i)}{S(y_{i-1})} = \frac{S(y_j)}{S(y_0)}, \end{aligned}$$

where y_0 is the smallest observed age in the sample. In order to bring the expectation inside the product, it was assumed that the S values are independent. The result demonstrates that at the death ages the estimator is unbiased.

With regard to the variance, we first need a general result concerning the variance of a product of independent random variables. Let X_1, \dots, X_n be independent random variables where $E(X_j) = \mu_j$ and $\text{Var}(X_j) = \sigma_j^2$. Then,

$$\begin{aligned} \text{Var}(X_1 \cdots X_n) &= E(X_1^2 \cdots X_n^2) - E(X_1 \cdots X_n)^2 \\ &= E(X_1^2) \cdots E(X_n^2) - E(X_1)^2 \cdots E(X_n)^2 \\ &= (\mu_1^2 + \sigma_1^2) \cdots (\mu_n^2 + \sigma_n^2) - \mu_1^2 \cdots \mu_n^2. \end{aligned}$$

For the product-limit estimator,

$$\begin{aligned} \text{Var}[S_n(y_j)] &= \text{Var}\left[\prod_{i=1}^j \left(\frac{r_i - S_i}{r_i}\right)\right] \\ &= \prod_{i=1}^j \left[\frac{S(y_i)^2}{S(y_{i-1})^2} + \frac{[S(y_{i-1}) - S(y_i)]S(y_i)}{r_i S(y_{i-1})^2} \right] - \frac{S(y_j)^2}{S(y_0)^2} \\ &= \prod_{i=1}^j \left[\frac{r_i S(y_i)^2 + [S(y_{i-1}) - S(y_i)]S(y_i)}{r_i S(y_{i-1})^2} \right] - \frac{S(y_j)^2}{S(y_0)^2} \\ &= \prod_{i=1}^j \left[\frac{S(y_i)^2}{S(y_{i-1})^2} \frac{r_i S(y_i) + [S(y_{i-1}) - S(y_i)]}{r_i S(y_i)} \right] - \frac{S(y_j)^2}{S(y_0)^2} \\ &= \frac{S(y_j)^2}{S(y_0)^2} \left\{ \prod_{i=1}^j \left[1 + \frac{S(y_{i-1}) - S(y_i)}{r_i S(y_i)} \right] - 1 \right\}. \end{aligned}$$

This formula is unpleasant to work with, so the following approximation is often used. It is based on the fact that for any set of small numbers a_1, \dots, a_n the product $(1 + a_1) \cdots (1 + a_n)$ is approximately $1 + a_1 + \cdots + a_n$. This follows because the missing terms are all products of two or more of the a_i s. If they are all small to begin with, the products will be even smaller and so can be ignored. Applying this produces the approximation

$$\text{Var}[S_n(y_j)] \doteq \left[\frac{S(y_j)}{S(y_0)} \right]^2 \sum_{i=1}^j \frac{S(y_{i-1}) - S(y_i)}{r_i S(y_i)}.$$

Because it is unlikely that the survival function is known, an estimated value needs to be inserted. Recall that the estimated value of $S(y_j)$ is actually conditional on being alive at age y_0 . Also, $(r_i - s_i)/r_i$ is an estimate of $S(y_i)/S(y_{i-1})$. Then,

$$\widehat{\text{Var}}[S_n(y_j)] \doteq S_n(y_j)^2 \sum_{i=1}^j \frac{s_i}{r_i(r_i - s_i)}. \quad (11.3)$$

Equation (11.3) is known as Greenwood's approximation. It is the only version that will be used in this text.

Example 11.13 Using Data Set D1, estimate the variance of $S_{30}(3)$ both directly and using Greenwood's formula. Do the same for ${}_2\hat{q}_3$.

Because there is no censoring or truncation, the empirical formula can be used to directly estimate this variance. There were three deaths out of 30 individuals, and therefore

$$\widehat{\text{Var}}[S_{30}(3)] = \frac{(3/30)(27/30)}{30} = \frac{81}{30^3}.$$

For Greenwood's approximation, $r_1 = 30$, $s_1 = 1$, $r_2 = 29$, and $s_2 = 2$. The approximation is

$$\left(\frac{27}{30}\right)^2 \left(\frac{1}{30(29)} + \frac{2}{29(27)}\right) = \frac{81}{30^3}.$$

It can be demonstrated that when there is no censoring or truncation the two formulas will always produce the same answer. Recall that the development of Greenwood's formula produced the variance only at death ages. The convention for non-death ages is to take the sum up to the last death age that is less than or equal to the age under consideration.

With regard to ${}_2\hat{q}_3$, arguing as in Example 11.6 produces an estimated (conditional) variance of

$$\widehat{\text{Var}}({}_2\hat{q}_3) = \frac{(4/27)(23/27)}{27} = \frac{92}{27^3}.$$

For Greenwood's formula, we first must note that we are estimating

$${}_2q_3 = \frac{S(3) - S(5)}{S(3)} = 1 - \frac{S(5)}{S(3)}.$$

As with the empirical estimate, all calculations must be done given the 27 people alive at duration 3. Furthermore, the variance of ${}_2\hat{q}_3$ is the same as the variance of $\hat{S}(5)$ using only information from duration 3 and beyond. Starting from duration 3 there are three death times, 3.1, 4.0, and 4.8, with $r_1 = 27$, $r_2 = 26$, $r_3 = 25$, $s_1 = 1$, $s_2 = 1$, and $s_3 = 2$. Greenwood's approximation is

$$\left(\frac{23}{27}\right)^2 \left(\frac{1}{27(26)} + \frac{1}{26(25)} + \frac{2}{25(23)}\right) = \frac{92}{27^3}.$$

□

Example 11.14 Repeat the previous example, this time using all 40 observations in Data Set D2 and the incomplete information due to censoring and truncation.

For this example, the direct empirical approach is not available. That is because it is unclear what the sample size is (it varies over time as subjects

enter and leave due to truncation and censoring). From Example 11.2, the relevant values within the first three years are $r_1 = 30$, $r_2 = 26$, $s_1 = 1$, and $s_2 = 2$. From Example 11.3, $S_{40}(3) = 0.8923$. Then, Greenwood's estimate is

$$(0.8923)^2 \left(\frac{1}{30(29)} + \frac{2}{26(24)}\right) = 0.0034671.$$

An approximate 95% confidence interval can be constructed using the normal approximation. It is

$$0.8923 \pm 1.96\sqrt{0.0034671} = 0.8923 \pm 0.1154,$$

which corresponds to the interval (0.7769, 1.0077). For small sample sizes, it is possible that the confidence intervals admit values less than 0 or greater than 1.

With regard to ${}_2\hat{q}_3$, the relevant quantities are (starting at duration 3, but using the subscripts from the earlier examples for these data) $r_3 = 26$, $r_4 = 26$, $r_5 = 23$, $r_6 = 21$, $s_3 = 1$, $s_4 = 2$, $s_5 = 1$, and $s_6 = 1$. This gives an estimated variance of

$$\left(\frac{0.7215}{0.8923}\right)^2 \left(\frac{1}{26(25)} + \frac{2}{26(24)} + \frac{1}{23(22)} + \frac{1}{21(20)}\right) = 0.0059502. \quad \square$$

The previous example indicated that the usual method of constructing a confidence interval can lead to an unacceptable result. An alternative approach can be constructed as follows. Let $Y = \ln[-\ln S_n(t)]$. Using the delta method (see Theorem 12.17), the variance of Y can be approximated as follows. The function of interest is $g(x) = \ln(-\ln x)$. Its derivative is

$$g'(x) = \frac{1}{-\ln x} \frac{-1}{x} = \frac{1}{x \ln x}.$$

According to the delta method, the variance of Y can be approximated by

$$\{g'[E(S_n(t))]\}^2 \text{Var}[S_n(t)] = \frac{\text{Var}[S_n(t)]}{[S_n(t) \ln S_n(t)]^2}$$

where we have used the fact that $S_n(t)$ is an unbiased estimator of $S(t)$. Then, an estimated 95% confidence interval for $\theta = \ln[-\ln S(t)]$ is

$$\ln[-\ln S_n(t)] \pm 1.96 \frac{\sqrt{\widehat{\text{Var}}[S_n(t)]}}{S_n(t) \ln S_n(t)}.$$

Because $S(t) = \exp(-e^\theta)$, putting each endpoint through this formula will provide a confidence interval for $S(t)$. For the upper limit we have (where

$$\hat{v} = \widehat{\text{Var}}[S_n(t)]$$

$$\begin{aligned} & \exp \left\{ -e^{\ln[-\ln S_n(t)] + 1.96\sqrt{\hat{v}}/[\ln S_n(t)]} \right\} \\ &= \exp \left\{ [\ln S_n(t)] e^{1.96\sqrt{\hat{v}}/[\ln S_n(t)]} \right\} \\ &= S_n(t)^U, \quad U = \exp \left[\frac{1.96\sqrt{\hat{v}}}{S_n(t) \ln S_n(t)} \right]. \end{aligned}$$

Similarly, the lower limit is $S_n(t)^{1/U}$. This interval will always be inside the range zero to 1 and is referred to as the log-transformed confidence interval.

Example 11.15 Obtain the log-transformed confidence interval for $S(3)$ as in Example 11.14.

We have

$$U = \exp \left[\frac{1.96\sqrt{0.0034671}}{0.8923 \ln(0.8923)} \right] = 0.32142.$$

The lower limit of the interval is $0.8923^{1/0.32142} = 0.70150$ and the upper limit is $0.8923^{0.32142} = 0.96404$. \square

Similar results are available for the Nelson–Åalen estimator. An intuitive derivation of a variance estimate proceeds as follows. As in the derivation for the Kaplan–Meier estimator, all results are obtained assuming the risk set numbers are known, not random. The number of deaths at death time t_i has approximately a Poisson distribution³ with parameter $r_i h(t_i)$ and so its variance is $r_i h(t_i)$, which can be approximated by $r_i(s_i/r_i) = s_i$. Then (also assuming independence),

$$\widehat{\text{Var}}[\hat{H}(y_j)] = \widehat{\text{Var}} \left(\sum_{i=1}^j \frac{s_i}{r_i} \right) = \sum_{i=1}^j \frac{\widehat{\text{Var}}(s_i)}{r_i^2} = \sum_{i=1}^j \frac{s_i}{r_i^2}.$$

The linear confidence interval is simply

$$\hat{H}(t) \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}[\hat{H}(y_j)]}.$$

A log-transformed interval similar to the one developed for the survival function⁴ is

$$\hat{H}(t)U, \quad \text{where } U = \exp \left[\pm \frac{z_{\alpha/2} \sqrt{\widehat{\text{Var}}[\hat{H}(y_j)]}}{\hat{H}(t)} \right].$$

³A binomial assumption (as was used for the Kaplan–Meier derivation) could also have been made. Similarly, a Poisson assumption could have been used for the Kaplan–Meier derivation. The formulas given here are the ones most commonly used.

⁴The derivation of this interval uses the transformation $Y = \ln \hat{H}(t)$.

Example 11.16 Construct an approximate 95% confidence interval for $H(3)$ by each formula using all 40 observations in Data Set D2.

The point estimate is $\hat{H}(3) = \frac{1}{30} + \frac{2}{26} = 0.11026$. The estimated variance is $\frac{1}{30^2} + \frac{2}{26^2} = 0.0040697$. The linear confidence interval is

$$0.11026 \pm 1.96\sqrt{0.0040697} = 0.11026 \pm 0.12504$$

for an interval of $(-0.01478, 0.23530)$. For the log-transformed interval,

$$U = \exp \left[\pm \frac{1.96(0.0040697)^{1/2}}{0.11026} \right] = \exp(\pm 1.13402) = 0.32174 \text{ to } 3.10813.$$

The interval is $0.11026(0.32174) = 0.03548$ to $0.11026(3.10813) = 0.34270$. \square

11.2.1 Exercises

11.10 Using the full information from Data Set D1, empirically estimate q_j for $j = 0, \dots, 4$ and ${}_5p_0$ where the variable of interest is time of surrender. Estimate the variance of each of your estimators. Identify which estimated variances are conditional. Interpret ${}_5q_0$ as the probability of surrendering before the five years expire.

11.11 For Data Set A determine the empirical estimate of the probability of having two or more accidents and estimate its variance.

11.12 Repeat Example 11.13 using time to surrender as the variable.

11.13 Repeat Example 11.14 using time to surrender as the variable. Interpret ${}_2q_3$ as the probability of surrendering before the five years expire.

11.14 Obtain the log-transformed confidence interval for $S(3)$ in Exercise 11.13.

11.15 Construct 95% confidence intervals for $H(3)$ by each formula using all 40 observations in Data Set D2 with surrender being the variable of interest.

11.16 (*) Ten individuals were observed from birth. All were observed until death. Table 11.4 gives the death ages. Let V_1 denote the estimated conditional variance of ${}_3\hat{q}_7$ if calculated without any distribution assumption. Let V_2 denote the conditional variance of ${}_3\hat{q}_7$ if calculated knowing that the survival function is $S(t) = 1 - t/15$. Determine $V_1 - V_2$.

11.17 (*) For the interval from zero to one year, the exposure (r) is 15 and the number of deaths (s) is 3. For the interval from one to two years the exposure is 80 and the number of deaths is 24. For two to three years the values are 25 and 5; for three to four years they are 60 and 6; and for four to

Table 11.4 Data for Exercise 11.16

Age	Number of deaths
2	1
3	1
5	1
7	2
10	1
12	2
13	1
14	1

five years they are 10 and 3. Determine Greenwood's approximation to the variance of $\hat{S}(4)$.

11.18 (*) Observations can be censored, but there is no truncation. Let y_j and y_{j+1} be consecutive death ages. A 95% linear confidence interval for $H(y_j)$ using the Nelson-Åalen estimator is (0.07125, 0.22875) while a similar interval for $H(y_{j+1})$ is (0.15607, 0.38635). Determine s_{j+1} .

11.19 (*) A mortality study is conducted on 50 lives, all observed from age 0. At age 15 there were two deaths; at age 17 there were three censored observations; at age 25 there were four deaths; at age 30 there were c censored observations; at age 32 there were eight deaths; and at age 40 there were two deaths. Let S be the product-limit estimate of $S(35)$ and let V be the Greenwood estimate of this estimator's variance. You are given $V/S^2 = 0.011467$. Determine the value of c .

11.20 (*) Fifteen cancer patients were observed from the time of diagnosis until the earlier of death or 36 months from diagnosis. Deaths occurred as follows: At 15 months there were two deaths; at 20 months there were three deaths; at 24 months there were two deaths; at 30 months there were d deaths; at 34 months there were two deaths; and at 36 months there was one death. The Nelson-Åalen estimate of $H(35)$ is 1.5641. Determine the variance of this estimator.

11.21 (*) You are given the values in Table 11.5. Determine the standard deviation of the Nelson-Åalen estimator of the cumulative hazard function at time 20.

11.3 KERNEL DENSITY MODELS

One problem with empirical distributions is that they are always discrete. If it is known that the true distribution is continuous, the empirical distribution

Table 11.5 Data for Exercise 11.21

y_j	r_j	s_j
1	100	15
8	65	20
17	40	13
25	31	31

may be viewed as a poor approximation. In this section, a method of obtaining a smooth, empirical-like distribution is introduced. Recall from Definition 10.4 that the idea is to replace each discrete piece of probability by a continuous random variable. While not necessary, it is customary that the continuous variable have a mean equal to the value of the point it replaces. This ensures that the kernel estimate has the same mean as the empirical estimate. One way to think about such a model is that it produces the final observed value in two steps. The first step is to draw a value at random from the empirical distribution. The second step is to draw a value at random from a continuous distribution whose mean is equal to the value drawn at the first step. The selected continuous distribution is called the kernel.

For notation, let $p(y_j)$ be the probability assigned to the value y_j ($j = 1, \dots, k$) by the empirical distribution. Let $K_y(x)$ be a distribution function for a continuous distribution such that its mean is y . Let $k_y(x)$ be the corresponding density function.

Definition 11.17 A kernel density estimator of a distribution function is

$$\hat{F}(x) = \sum_{j=1}^k p(y_j) K_{y_j}(x)$$

and the estimator of the density function is

$$\hat{f}(x) = \sum_{j=1}^k p(y_j) k_{y_j}(x).$$

The function $k_y(x)$ is called the kernel. Three kernels will now be introduced.

Definition 11.18 The uniform kernel is given by

$$k_y(x) = \begin{cases} 0, & x < y - b, \\ \frac{1}{2b}, & y - b \leq x \leq y + b, \\ 0, & x > y + b, \end{cases}$$

$$K_y(x) = \begin{cases} 0, & x < y - b, \\ \frac{x - y + b}{2b}, & y - b \leq x \leq y + b, \\ 1, & x > y + b. \end{cases}$$

The triangular kernel is given by

$$k_y(x) = \begin{cases} 0, & x < y - b, \\ \frac{x - y + b}{b^2}, & y - b \leq x \leq y, \\ \frac{y + b - x}{b^2}, & y \leq x \leq y + b, \\ 0, & x > y + b, \end{cases}$$

$$K_y(x) = \begin{cases} 0, & x < y - b, \\ \frac{(x - y + b)^2}{2b^2}, & y - b \leq x \leq y, \\ 1 - \frac{(y + b - x)^2}{2b^2}, & y \leq x \leq y + b, \\ 1, & x > y + b. \end{cases}$$

The gamma kernel is given by letting the kernel have a gamma distribution with shape parameter α and scale parameter y/α . That is,

$$k_y(x) = \frac{x^{\alpha-1} e^{-x\alpha/y}}{(y/\alpha)^\alpha \Gamma(\alpha)}.$$

Note that the gamma distribution has a mean of $\alpha(y/\alpha) = y$ and a variance of $\alpha(y/\alpha)^2 = y^2/\alpha$.

In each case there is a parameter that relates to the spread of the kernel. In the first two cases it is the value of $b > 0$, which is called the bandwidth. In the gamma case, the value of α controls the spread, with a larger value indicating a smaller spread. There are other kernels that cover the range from zero to infinity.

Example 11.19 Determine the kernel density estimate for Example 10.8 using each of the three kernels.

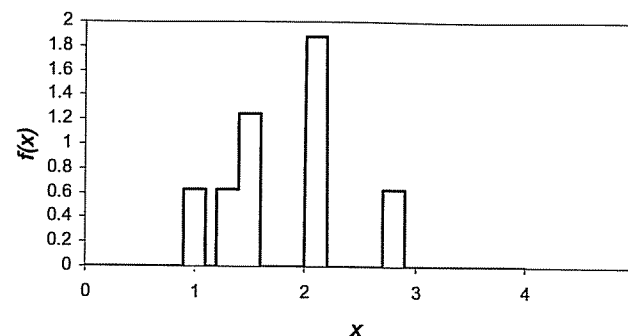


Fig. 11.1 Uniform kernel density with bandwidth 0.1.

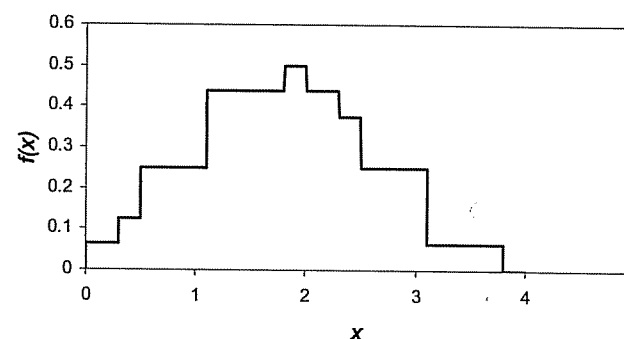


Fig. 11.2 Uniform kernel density with bandwidth 1.0.

The empirical distribution places probability $\frac{1}{8}$ at 1.0, $\frac{1}{8}$ at 1.3, $\frac{2}{8}$ at 1.5, $\frac{3}{8}$ at 2.1, and $\frac{1}{8}$ at 2.8. For a uniform kernel with a bandwidth of 0.1 we do not get much separation. The data point at 1.0 is replaced by a horizontal density function running from 0.9 to 1.1 with a height of $\frac{1}{8} \frac{1}{2(0.1)} = 0.625$. On the other hand, with a bandwidth of 1.0 that same data point is replaced by a horizontal density function running from 0.0 to 2.0 with a height of $\frac{1}{8} \frac{1}{2(1)} = 0.0625$. Figures 11.1 and 11.2 provide plots of the density functions.

It should be clear that the larger bandwidth provides more smoothing. In the limit, as the bandwidth approaches zero, the kernel density estimate matches the empirical estimate. Note that, if the bandwidth is too large, probability will be assigned to negative values, which may be an undesirable result. Methods exist for dealing with that issue, but they will not be presented here.

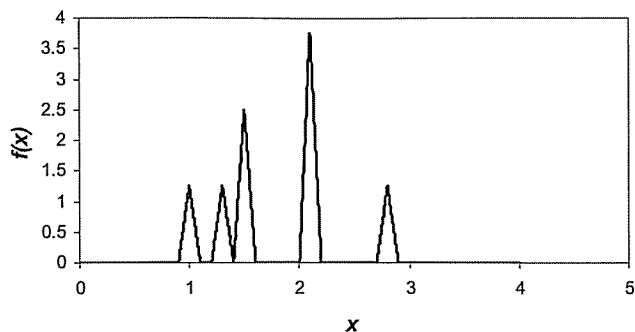


Fig. 11.3 Triangular kernel density with bandwidth 0.1.

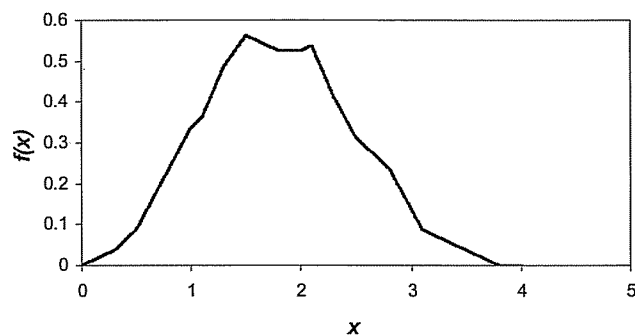
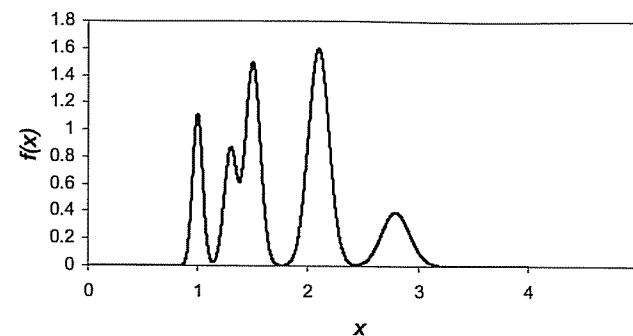
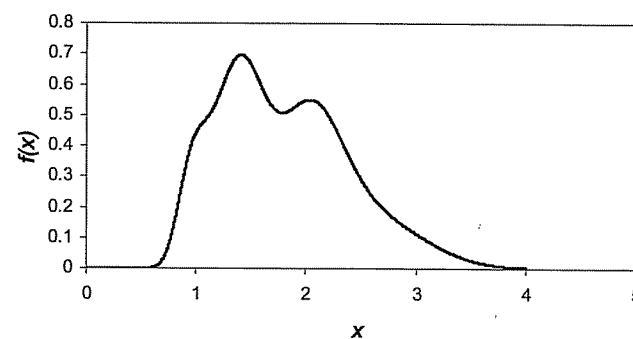


Fig. 11.4 Triangular kernel density with bandwidth 1.0.

For the triangular kernel, each point is replaced by a triangle. Pictures for the same two bandwidths used previously appear in Figures 11.3 and 11.4.

Once again, the larger bandwidth provides more smoothing. The gamma kernel simply provides a mixture of gamma distributions where each data point provides the mean and the empirical probabilities provide the weights. The density function is

$$f_{\alpha}(x) = \sum_{j=1}^5 p(y_j) \frac{x^{\alpha-1} e^{-x/y_j}}{(y_j/\alpha)^{\alpha} \Gamma(\alpha)}$$

Fig. 11.5 Gamma kernel density with $\alpha = 500$.Fig. 11.6 Gamma kernel density with $\alpha = 50$.

and is graphed in Figures 11.5 and 11.6 for two α values.⁵ For this kernel, decreasing the value of α increases the amount of smoothing. Further discussion of the gamma kernel can be found in [23], where the author recommends $\alpha = \sqrt{n}/(\hat{\mu}'_4/\hat{\mu}'_2 - 1)^{1/2}$. \square

11.3.1 Exercises

11.22 Provide the formula for the Pareto kernel.

⁵When computing values of the density function, overflow and underflow problems can be reduced by computing the logarithm of the elements of the ratio, that is, $(\alpha - 1) \ln x - x\alpha/y_j - \alpha \ln(y_j/\alpha) - \ln \Gamma(\alpha)$, and then exponentiating the result.

Table 11.6 Data for Exercise 11.24

t_j	s_j	r_j
10	1	20
34	1	19
47	1	18
75	1	17
156	1	16
171	1	15

11.23 Construct a kernel density estimate for the time to surrender for Data Set D2. Be aware of the fact that this is a mixed distribution (probability is continuous from 0 to 5 but is discrete at 5).

11.24 (*) You are given the data in Table 11.6 on time to death. Using the uniform kernel with a bandwidth of 60, determine $\hat{f}(100)$.

11.4 APPROXIMATIONS FOR LARGE DATA SETS

11.4.1 Introduction

When there are large amounts of data, constructing the Kaplan–Meier estimate may require more sorting and counting than can be justified by the results. This is especially true if values of the distribution function are not needed at each value. For example, if the goal is to construct a mortality table, values are needed only at integral ages. The finer details of mortality table construction and alternative methods can be found in the texts by Batten [11] and London [85]. While the context for the examples presented here will be the construction of mortality tables, the methods can apply anytime the data have been rounded.

Suppose there are intervals given as $c_0 < c_1 < \dots < c_k$. Let d_j be the number of observations that are left truncated at a value in the interval $[c_j, c_{j+1})$. In a mortality study, this would be a count of the lives that were first observed at an age in the given range. Similarly, let u_j be the number of observations that are right censored at a value in the interval $(c_j, c_{j+1}]$. Note the difference in the endpoints of the two intervals. This is necessary because truncation is possible at the left end of the first interval but not at the right end of the last interval. The reverse is true for censoring. Note that all observations are to contribute to some d_j , but only observations that are actually censored can contribute to some u_j . Let x_j be the number of uncensored observations in the interval $(c_j, c_{j+1}]$. Then $n = \sum_{j=0}^{k-1} d_j = \sum_{j=0}^{k-1} (u_j + x_j)$ where n is the

sample size. The computational advantage is that one pass through the data set allows these values to be accumulated and from there only this reduced set of values needs to be processed.

11.4.2 Kaplan–Meier type approximations

In order to apply the Kaplan–Meier formula, assumptions must be made about the location of the values within each interval. The simplest is to assume that all of the uncensored observations in an interval occur at the same value within that interval, say c_j^* , that all left truncated values are less than c_j^* , and that all right censored values are greater than or equal to c_j^* . The risk set at c_j^* is $r_j = \sum_{i=0}^j d_i - \sum_{i=0}^{j-1} (u_i + x_i)$. Rather than place all probability at the c_j^* values (as is done by the Kaplan–Meier method), it is customary to evaluate the distribution function at the given endpoints and then smooth the function by interpolation (usually linear) between successive values. Then,

$$\begin{aligned}\hat{F}(c_0) &= 0, \\ \hat{F}(c_j) &= 1 - \prod_{i=0}^{j-1} \left(1 - \frac{x_i}{r_i}\right), \quad j = 1, 2, \dots, k.\end{aligned}\quad (11.4)$$

Let

$$q_j = c_{j+1} - c_j q_{c_j} = \frac{S(c_j) - S(c_{j+1})}{S(c_j)}.$$

From (11.4),

$$\hat{q}_j = \frac{\prod_{i=0}^{j-1} \left(1 - \frac{x_i}{r_i}\right) - \prod_{i=0}^j \left(1 - \frac{x_i}{r_i}\right)}{\prod_{i=0}^{j-1} \left(1 - \frac{x_i}{r_i}\right)} = 1 - \left(1 - \frac{x_j}{r_j}\right) = \frac{x_j}{r_j}. \quad (11.5)$$

This is the traditional form of a life table estimator where the numerator has the number of observed deaths and the denominator is a measure of exposure (the number of lives available to die). For this formula, all who enter the study prior to or during the current interval are given a chance to die and all who have left prior to the interval are removed from consideration. If dollar amounts are being studied and the boundaries include all the possible values for deductibles and limits, this formula produces the exact product-limit estimate at the given values. For mortality studies, this is equivalent to having all lives enter on their birthdays (which may be true if insuring ages are used, see Exercise 11.26) and surrender on birthdays.

Equation (11.5) can be generalized as follows. Let $P_j = \sum_{i=0}^{j-1} (d_i - u_i - x_i)$, the number of people under observation at age c_j . Continue to assume that all the uncensored observations occur at a fixed point in the interval. Further assume that at this time 100 α % of those who will enter (be counted in d_j) have

done so and that 100β% of those who will be censored (be counted in u_j) have done so. Then the risk set is $r_j = P_j + \alpha d_j - \beta u_j$. Equation (11.5) is obtained by setting $\alpha = 1$ and $\beta = 0$. An alternative is to set $\alpha = \beta = 0.5$. This is equivalent to having the entrants and exits spread uniformly throughout the interval. The result is $r_j = P_j + 0.5(d_j - u_j)$. Because $P_{j+1} = P_j + d_j - u_j - x_j$, this can be rewritten as

$$r_j = 0.5(P_j + P_{j+1} + x_j), \quad (11.6)$$

a formula commonly used in constructing mortality tables. At times it may be necessary to make different assumptions for different intervals, as illustrated in Example 11.20.

11.4.3 Multiple-decrement tables

The goal of all the estimation procedures in this text is to deduce the probability distribution for the variable of interest in the absence of truncation and censoring. For loss data, that would be the probabilities if there were no deductible or limit. For lifetime data it would be the probabilities if we could follow people from birth to death. In the language of *Actuarial Mathematics* [16], these are called *single-decrement rates* and are denoted q'_j . It is often desired to create a table of *multiple-decrement probabilities*, denoted q_j . A superscript identifies the decrement of interest. For example, suppose the decrements were death (d), withdrawal (w), and retirement (r). Then $q'_j{}^{(w)}$ is the probability that a person age c_j withdraws prior to age c_{j+1} in an environment where death and retirement are not possible while $q'_j{}^{(r)}$ is the probability that a person age c_j retires prior to age c_{j+1} in an environment where prior death or withdrawal eliminates the chance of retirement. Multiple-decrement tables are often constructed by obtaining the single-decrement rates from different sources and then combining with a formula such as

$$q_j^{(g)} = \frac{\ln(1 - q'_j{}^{(g)})}{\ln(1 - q'_j{}^{(T)})} q_j^{(T)}, \quad q_j^{(T)} = 1 - \prod_{g=1}^m (1 - q'_j{}^{(g)}), \quad (11.7)$$

where g represents an arbitrary decrement and m is the total number of decrements.

Example 11.20 Estimate both single- and multiple-decrement quantities using Data Set D2 and the methods of this section. Make reasonable assumptions.

First consider the decrement death. In the notation of this section, the relevant quantities are in Table 11.7. For this setting, more than one assumption is needed. For $d_0 = 32$ it is clear that the 30 values that are exactly zero should be treated as such (policies followed from the beginning) while

Table 11.7 Single-decrement calculations for Example 11.20

j	d_j	u_j	x_j	P_j	r_j	$q'_j{}^{(d)}$	$\hat{F}(j)$
0	32	3	1	0	29.5	0.0339	0.0000
1	2	2	0	28	28.0	0.0000	0.0339
2	3	3	2	28	28.0	0.0714	0.0339
3	3	3	3	26	26.0	0.1154	0.1029
4	0	21	2	23	21.0	0.0952	0.2064
5							0.2820

Table 11.8 Multiple-decrement calculations for Example 11.20

j	$q'_j{}^{(d)}$	$q'_j{}^{(w)}$	$q'_j{}^{(T)}$	$q_j^{(d)}$	$q_j^{(w)}$
0	0.0339	0.0984	0.1289	0.0322	0.0967
1	0.0000	0.0690	0.0690	0.0000	0.0690
2	0.0714	0.1053	0.1692	0.0676	0.1015
3	0.1154	0.1154	0.2175	0.1087	0.1087
4	0.0952	0.1818	0.2597	0.0864	0.1733

the 2 policies that entered after issue require an assumption. It makes sense to assume a uniform spread. Then $r_0 = 30 + 0.5(2) - 0.5(3) = 29.5$. The other r values are calculated using (11.6). Also note that the 17 policies that were still active after five years are all assumed to be censored at time 5, rather than be spread uniformly through the fifth year.

For withdrawals the values of $q'_j{}^{(w)}$ are given in Table 11.8 along with the calculated multiple decrement values using (11.7). \square

Example 11.21 Loss data for policies with deductibles of 0, 250, and 500 and policy limits of 5,000, 7,500, and 10,000 were collected. The data are in Table 11.9. Use the methods of this section to estimate the distribution function for losses.

The calculations appear in Table 11.10. Because the deductibles and limits are at the endpoints of intervals, the only reasonable assumption is $\alpha = 1$ and $\beta = 0$. \square

11.4.4 Exercises

11.25 Verify the calculations in Table 11.8.

11.26 When life insurance policies are issued, it is customary to assign a whole-number age to the insured and then the premium associated with that age is charged. The interpretation of q_x changes from "What is the probability

Table 11.9 Data for Example 11.21

Range	Deductible			Total
	0	250	500	
0-100	15			15
100-250	16			16
250-500	34	96		130
500-1,000	73	175	251	499
1,000-2,500	131	339	478	948
2,500-5,000	83	213	311	607
5,000-7,500	12	48	88	148
7,500-10,000	1	4	11	16
At 5,000	7	17	18	42
At 7,500	5	10	15	30
At 10,000	2	1	4	7
Total	379	903	1,176	2,458

Table 11.10 Calculations for Example 11.21

c_j	d_j	u_j	x_j	P_j	r_j	$q_j^{(d)}$	$\hat{F}(c_j)$
0	379	0	15	0	379	0.0396	0.0000
100	0	0	16	364	364	0.0440	0.0396
250	903	0	130	348	1,251	0.1039	0.0818
500	1,176	0	499	1,121	2,297	0.2172	0.1772
1,000	0	0	948	1,798	1,798	0.5273	0.3560
2,500	0	42	607	850	850	0.7141	0.6955
5,000	0	30	148	201	201	0.7363	0.9130
7,500	0	7	16	23	23	0.6957	0.9770
10,000							0.9930

someone having their x th birthday dies in the next year?" to "What is the probability that a person who was assigned age $x - k$ at issue k years ago dies in the next year?" Such age assignments are called "insuring ages" and effectively move the birthday to the policy issue date. As a result, insured lives tend to enter observation on their birthday (artificially assigned as the policy issue date). This makes the d value exact (with $\alpha = 1$) rather than approximate. However, withdrawal can take place at any time, making $\beta =$

Table 11.11 Data for Exercise 11.26

d	u	x	d	u	x
45	46.0		45	45.8	
45	46.0		46	47.0	
45		45.3	46	47.0	
45		46.7	46	46.3	
45		45.4	46		46.2
45	47.0		46		46.4
45	45.4		46	46.9	

Table 11.12 Data for Exercise 11.27

Deductible	Payment	Deductible	Payment
250	2,221	500	3,660
250	2,500*	500	215
250	207	500	1,302
250	3,735	500	10,000*
250	5,000*	1,000	1,643
250	517	1,000	3,395
250	5,743	1,000	3,981
500	2,500*	1,000	3,836
500	525	1,000	5,000*
500	4,393	1,000	1,850
500	5,000*	1,000	6,722

*Amount paid was at the policy limit

0.5 a reasonable assumption.⁶ For the data in Table 11.11 estimate q_{45} and q_{46} using both the exact Kaplan-Meier estimate and the method of this section.

11.27 Twenty-two insurance payments are recorded in Table 11.12. Use the fewest reasonable number of intervals and the method of this section to estimate the probability that a policy with a deductible of 500 will have a payment in excess of 5,000.

⁶However, just as in Example 11.20 where different d values received different treatment, here there can be u values that deserve different treatment. Observation of an individual can end because the individual leaves or because the observation period ends. For studies of insured lives it is common for observation to end at a policy anniversary and thus at a whole-number insuring age. For them, $\beta = 0$ is an appropriate assumption.

Part IV

*Parametric statistical
methods*

12

Parameter estimation

If a phenomenon is to be modeled using a parametric model, it is necessary to assign values to the parameters. This could be done arbitrarily, but it would seem to be more reasonable to base the assignment on observations from that phenomenon. In particular, we will assume that n independent observations have been collected. For some of the techniques it will be further assumed that all the observations are from the same random variable. For others, that restriction will be relaxed.

The methods introduced in Section 12.1 are relatively easy to implement but tend to give poor results. Section 12.2 covers maximum likelihood estimation. This method is more difficult to use but has superior statistical properties and is considerably more flexible.

12.1 METHOD OF MOMENTS AND PERCENTILE MATCHING

For these methods we assume that all n observations are from the same parametric distribution. In particular, let the distribution function be given by

$$F(x|\theta), \quad \theta^T = (\theta_1, \theta_2, \dots, \theta_p)$$

where θ^T is the transpose of θ . That is, θ is a column vector containing the p parameters to be estimated. Furthermore, let $\mu'_k(\theta) = E(X^k|\theta)$ be the k th raw moment and let $\pi_g(\theta)$ be the 100 g th percentile of the random variable. That is, $F[\pi_g(\theta)|\theta] = g$. If the distribution function is continuous, there will be at least one solution to that equation.

For a sample of n independent observations from this random variable, let $\hat{\mu}'_k = \frac{1}{n} \sum_{j=1}^n x_j^k$ be the empirical estimate of the k th moment and let $\hat{\pi}_g$ be the empirical estimate of the 100 g th percentile

Definition 12.1 A method-of-moments estimate of θ is any solution of the p equations

$$\mu'_k(\theta) = \hat{\mu}'_k, \quad k = 1, 2, \dots, p.$$

The motivation for this estimator is that it produces a model that has the same first p raw moments as the data (as represented by the empirical distribution). The traditional definition of the method of moments uses positive integers for the moments. Arbitrary negative or fractional moments could also be used. In particular, when estimating parameters for inverse distributions, matching negative moments may be a superior approach.¹

Example 12.2 Use the method of moments to estimate parameters for the exponential, gamma, and Pareto distributions for Data Set B from Chapter 10.

The first two sample moments are

$$\begin{aligned}\hat{\mu}'_1 &= \frac{1}{20}(27 + \dots + 15,743) = 1,424.4, \\ \hat{\mu}'_2 &= \frac{1}{20}(27^2 + \dots + 15,743^2) = 13,238,441.9.\end{aligned}$$

For the exponential distribution the equation is

$$\theta = 1,424.4$$

with the obvious solution, $\hat{\theta} = 1,424.4$.

For the gamma distribution, the two equations are

$$\begin{aligned}E(X) &= \alpha\theta = 1,424.4, \\ E(X^2) &= \alpha(\alpha + 1)\theta^2 = 13,238,441.9.\end{aligned}$$

Dividing the second equation by the square of the first equation yields

$$\frac{\alpha + 1}{\alpha} = 6.52489, \quad 1 = 5.52489\alpha$$

¹One advantage is that with appropriate moments selected the equations must have a solution within the range of allowable parameter values.

and so $\hat{\alpha} = 1/5.52489 = 0.18100$ and $\hat{\theta} = 1,424.4/0.18100 = 7,869.61$.

For the Pareto distribution, the two equations are

$$\begin{aligned}E(X) &= \frac{\theta}{\alpha - 1} = 1,424.4, \\ E(X^2) &= \frac{2\theta^2}{(\alpha - 1)(\alpha - 2)} = 13,238,441.9.\end{aligned}$$

Dividing the second equation by the square of the first equation yields

$$\frac{2(\alpha - 1)}{(\alpha - 2)} = 6.52489$$

with a solution of $\hat{\alpha} = 2.442$ and then $\hat{\theta} = 1,424.4(1.442) = 2,053.985$. \square

There is no guarantee that the equations will have a solution or, if there is a solution, that it will be unique.

Definition 12.3 A percentile matching estimate of θ is any solution of the p equations

$$\pi_{g_k}(\theta) = \hat{\pi}_{g_k}, \quad k = 1, 2, \dots, p,$$

where g_1, g_2, \dots, g_p are p arbitrarily chosen percentiles. From the definition of percentile, the equations can also be written

$$F(\hat{\pi}_{g_k}|\theta) = g_k, \quad k = 1, 2, \dots, p.$$

The motivation for this estimator is that it produces a model with p percentiles that match the data (as represented by the empirical distribution). As with the method of moments, there is no guarantee that the equations will have a solution or, if there is a solution, that it will be unique. One problem with this definition is that percentiles for discrete random variables (such as the empirical distribution) are not always well defined. For example, Data Set B has 20 observations. Any number between 384 and 457 has 10 observations below and 10 above and so could serve as the median. The convention is to use the midpoint. However, for other percentiles, there is no "official" interpolation scheme.² The following definition will be used here.

Definition 12.4 The smoothed empirical estimate of a percentile is found by

$$\begin{aligned}\hat{\pi}_g &= (1 - h)x_{(j)} + hx_{(j+1)}, \text{ where} \\ j &= \lfloor (n + 1)g \rfloor \text{ and } h = (n + 1)g - j.\end{aligned}$$

²Hyndman and Fan [65] present nine different methods. They recommend a slight modification of the one presented here using $j = \lfloor g(n + \frac{1}{3}) + \frac{1}{3} \rfloor$ and $h = g(n + \frac{1}{3}) + \frac{1}{3} - j$.

Here $\lfloor \cdot \rfloor$ indicates the greatest integer function and $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ are the order statistics from the sample.

Unless there are two or more data points with the same value, no two percentiles will have the same value. One feature of this definition is that $\hat{\pi}_g$ cannot be obtained for $g < 1/(n+1)$ or $g > n/(n+1)$. This seems reasonable as we should not expect to be able to infer the value of large or small percentiles from small samples. We will use the smoothed version whenever an empirical percentile estimate is called for.

Example 12.5 Use percentile matching to estimate parameters for the exponential and Pareto distributions for Data Set B.

For the exponential distribution, select the 50th percentile. The empirical estimate is the traditional median of $\hat{\pi}_{0.5} = (384 + 457)/2 = 420.5$ and the equation to solve is

$$\begin{aligned} 0.5 &= F(420.5|\theta) = 1 - e^{-420.5/\theta}, \\ \ln 0.5 &= \frac{-420.5}{\theta}, \\ \hat{\theta} &= \frac{-420.5}{\ln 0.5} = 606.65. \end{aligned}$$

For the Pareto distribution, select the 30th and 80th percentiles. The smoothed empirical estimates are found as follows:

$$\begin{aligned} \text{30th: } j &= \lfloor 21(0.3) \rfloor = \lfloor 6.3 \rfloor = 6, h = 6.3 - 6 = 0.3, \\ \hat{\pi}_{0.3} &= 0.7(161) + 0.3(243) = 185.6, \\ \text{80th: } j &= \lfloor 21(0.8) \rfloor = \lfloor 16.8 \rfloor = 16, h = 16.8 - 16 = 0.8, \\ \hat{\pi}_{0.8} &= 0.2(1,193) + 0.8(1,340) = 1,310.6. \end{aligned}$$

The equations to solve are

$$\begin{aligned} 0.3 &= F(185.6) = 1 - \left(\frac{\theta}{185.6 + \theta} \right)^\alpha, \\ 0.8 &= F(1,310.6) = 1 - \left(\frac{\theta}{1,310.6 + \theta} \right)^\alpha, \\ \ln 0.7 &= -0.356675 = \alpha \ln \left(\frac{\theta}{185.6 + \theta} \right), \\ \ln 0.2 &= -1.609438 = \alpha \ln \left(\frac{\theta}{1,310.6 + \theta} \right), \\ \frac{-1.609438}{-0.356675} &= 4.512338 = \frac{\ln \left(\frac{\theta}{1,310.6 + \theta} \right)}{\ln \left(\frac{\theta}{185.6 + \theta} \right)}. \end{aligned}$$

Any of the methods from Appendix F can be used to solve this equation for $\hat{\theta} = 715.03$. Then, from the first equation,

$$0.3 = 1 - \left(\frac{715.03}{185.6 + 715.03} \right)^\alpha,$$

which yields $\hat{\alpha} = 1.54559$. □

The estimates are much different from those obtained in Example 12.2. This is one indication that these methods may not be particularly reliable.

12.1.1 Exercises

12.1 Determine the method-of-moments estimate for an exponential model for Data Set B with observations censored at 250.

12.2 Determine the method-of-moments estimate for a lognormal model for Data Set B.

12.3 (*) The 20th and 80th percentiles from a sample are 5 and 12 respectively. Using the percentile matching method, estimate $S(8)$ assuming the population has a Weibull distribution.

12.4 (*) From a sample you are given that the mean is 35,000, the standard deviation is 75,000, the median is 10,000, and the 90th percentile is 100,000. Using the percentile matching method, estimate the parameters of a Weibull distribution.

12.5 (*) A sample of size 5 produced the values 4, 5, 21, 99, and 421. You fit a Pareto distribution using the method of moments. Determine the 95th percentile of the fitted distribution.

12.6 (*) In year 1 there are 100 claims with an average size of 10,000 and in year 2 there are 200 claims with an average size of 12,500. Inflation increases the size of all claims by 10% per year. A Pareto distribution with $\alpha = 3$ and θ unknown is used to model the claim size distribution. Estimate θ for year 3 using the method of moments.

12.7 (*) From a random sample the 20th percentile is 18.25 and the 80th percentile is 35.8. Estimate the parameters of a lognormal distribution using percentile matching and then use these estimates to estimate the probability of observing a value in excess of 30.

12.8 (*) A claim process is a mixture of two random variables A and B , where A has an exponential distribution with a mean of 1 and B has an exponential distribution with a mean of 10. A weight of p is assigned to distribution A and

1 - p to distribution B . The standard deviation of the mixture is 2. Estimate p by the method of moments.

12.9 (*) A random sample of 20 observations has been ordered as follows:

12 16 20 23 26 28 30 32 33 35
36 38 39 40 41 43 45 47 50 57

Determine the 60th sample percentile using the smoothed empirical estimate.

12.10 (*) The following 20 wind losses (in millions of dollars) were recorded in one year:

1 1 1 1 1 2 2 3 3 4
6 6 8 10 13 14 15 18 22 25

Determine the sample 75th percentile using the smoothed empirical estimate.

12.11 (*) The observations 1,000, 850, 750, 1,100, 1,250, and 900 were obtained as a random sample from a gamma distribution with unknown parameters α and θ . Estimate these parameters by the method-of-moments.

12.12 (*) A random sample of claims has been drawn from a loglogistic distribution. In the sample, 80% of the claims exceed 100 and 20% exceed 400. Estimate the loglogistic parameters by percentile matching.

12.13 (*) Let x_1, \dots, x_n be a random sample from a population with cdf $F(x) = x^p$, $0 < x < 1$. Determine the method of moments estimate of p .

12.14 (*) A random sample of 10 claims obtained from a gamma distribution is given below:

1,500 6,000 3,500 3,800 1,800 5,500 4,800 4,200 3,900 3,000

Estimate α and θ by the method of moments.

12.15 (*) A random sample of five claims from a lognormal distribution is given below:

500 1,000 1,500 2,500 4,500.

Estimate μ and σ by the method of moments. Estimate the probability that a loss will exceed 4,500.

12.16 (*) The random variable X has pdf $f(x) = \beta^{-2}x \exp(-0.5x^2/\beta^2)$, $x, \beta > 0$. For this random variable, $E(X) = (\beta/2)\sqrt{2\pi}$ and $\text{Var}(X) = 2\beta^2 - \pi\beta^2/2$. You are given the following five observations:

4.9 1.8 3.4 6.9 4.0

Table 12.1 Data for Exercise 12.18

No. of claims	No. of policies
0	9,048
1	905
2	45
3	2
4+	0

Table 12.2 Data for Exercise 12.19

No. of claims	No. of policies
0	861
1	121
2	13
3	3
4	1
5	0
6	1
7+	0

Determine the method-of-moments estimate of β .

12.17 The random variable X has pdf $f(x) = \alpha\lambda^\alpha(\lambda + x)^{-\alpha-1}$, $x, \alpha, \lambda > 0$. It is known that $\lambda = 1,000$. You are given the following five observations:

43 145 233 396 775

Determine the method of moments estimate of α .

12.18 Use the data in Table 12.1 to determine the method-of-moments estimate of the parameters of the negative binomial model.

12.19 Use the data in Table 12.2 to determine the method-of-moments estimate of the parameters of the negative binomial model.

12.2 MAXIMUM LIKELIHOOD ESTIMATION

12.2.1 Introduction

Estimation by the method of moments and percentile matching is often easy to do, but these estimators tend to perform poorly. The main reason for this is that they use a few features of the data, rather than the entire set of observations. It is particularly important to use as much information as possible