Loss Data Analytics

An open text authored by the Actuarial Community

Contents

Preface

Date: 01 October 2018

Book Description

Loss Data Analytics is an interactive, online, freely available text.

- The online version contains many interactive objects (quizzes, computer demonstrations, interactive graphs, video, and the like) to promote *deeper learning*.
- A subset of the book is available for *offline reading* in pdf and EPUB formats.
- The online text will be available in multiple languages to promote access to a *worldwide audience*.

What will success look like?

The online text will be freely available to a worldwide audience. The online version will contain many interactive objects (quizzes, computer demonstrations, interactive graphs, video, and the like) to promote deeper learning. Moreover, a subset of the book will be available in pdf format for low-cost printing. The online text will be available in multiple languages to promote access to a worldwide audience.

How will the text be used?

This book will be useful in actuarial curricula worldwide. It will cover the loss data learning objectives of the major actuarial organizations. Thus, it will be suitable for classroom use at universities as well as for use by independent learners seeking to pass professional actuarial examinations. Moreover, the text will also be useful for the continuing professional development of actuaries and other professionals in insurance and related financial risk management industries.

Why is this good for the profession?

An online text is a type of open educational resource (OER). One important benefit of an OER is that it equalizes access to knowledge, thus permitting a broader community to learn about the actuarial profession. Moreover, it has the capacity to engage viewers through active learning that deepens the learning process, producing analysts more capable of solid actuarial work. Why is this good for students and teachers and others involved in the learning process?

Cost is often cited as an important factor for students and teachers in textbook selection (see a recent post on the $400 textbook). Students will also appreciate the ability to "carry the book around" on their mobile devices.

Why loss data analytics?

Although the intent is that this type of resource will eventually permeate throughout the actuarial curriculum, one has to start somewhere. Given the dramatic changes in the way that actuaries treat data, loss data seems like a natural place to start. The idea behind the name *loss data analytics* is to integrate classical loss data models from applied probability with modern analytic tools. In particular, we seek to recognize that big data (including social media and usage based insurance) are here and high speed computation s readily available.

Project Goal

The project goal is to have the actuarial community author our textbooks in a collaborative fashion.

To get involved, please visit our Loss Data Analytics Project Site.

Acknowledgements

Edward Frees acknowledges the John and Anne Oros Distinguished Chair for Inspired Learning in Business which provided seed money to support the project. Frees and his Wisconsin colleagues also acknowledge a Society of Actuaries Center of Excellence Grant that provided funding to support work in dependence modeling and health initiatives.

We acknowledge the Society of Actuaries for permission to use problems from their examinations.

We thank Rob Hyndman, Monash University, for allowing us to use his excellent style files to produce the online version of the book.

We thank Yihui Xie and his colleagues at Rstudio for the R bookdown package that allows us to produce this book.

We also wish to acknowledge the support and sponsorship of the International Association of Black Actuaries in our joint efforts to provide actuarial educational content to all.

Contributors

The project goal is to have the actuarial community author our textbooks in a collaborative fashion. The following contributors have taken a leadership role in developing *Loss Data Analytics*.

- **Zeinab Amin** is the Director of the Actuarial Science Program and Associate Dean for Undergraduate Studies of the School of Sciences and Engineering at the American University in Cairo (AUC). Amin holds a PhD in Statistics and is an Associate of the Society of Actuaries. Amin is the recipient of the 2016 Excellence in Academic Service Award and the 2009 Excellence in Teaching Award from AUC. Amin has designed and taught a variety of statistics and actuarial science courses. Amin's current area of research includes quantitative risk assessment, reliability assessment, general statistical modelling, and Bayesian statistics.
- **Katrien Antonio**, KU Leuven
- **Jan Beirlant**, KU Leuven
- **Carolina Castro** - University of Buenos Aires

- **Curtis Gary Dean** is the Lincoln Financial Distinguished Professor of Actuarial Science at Ball State University. He is a Fellow of the Casualty Actuarial Society and a CFA charterholder. He has extensive practical experience as an actuary at American States Insurance, SAFECO, and Travelers. He has served the CAS and actuarial profession as chair of the Examination Committee, first editor-in-chief for *Variance: Advancing the Science of Risk*, and as a member of the Board of Directors and the Executive Council. He contributed a chapter to *Predictive Modeling Applications in Actuarial Science* published by Cambridge University Press.
- **Edward W. (Jed) Frees** is an emeritus professor, formerly the Hickman-Larson Chair of Actuarial Science at the University of Wisconsin-Madison. He is a Fellow of both the Society of Actuaries and the American Statistical Association. He has published extensively (a four-time winner of the Halmstad and Prize for best paper published in the actuarial literature) and has written three books. He also is a co-editor of the two-volume series *Predictive Modeling Applications in Actuarial Science* published by Cambridge University Press.
- **Guojun Gan** is an assistant professor in the Department of Mathematics at the University of Connecticut, where he has been since August 2014. Prior to that, he worked at a large life insurance company in Toronto, Canada for six years. He received a BS degree from Jilin University, Changchun, China, in 2001 and MS and PhD degrees from York University, Toronto, Canada, in 2003 and 2007, respectively. His research interests include data mining and actuarial science. He has published several books and papers on a variety of topics, including data clustering, variable annuity, mathematical finance, applied statistics, and VBA programming.
- **Lisa Gao** is a doctoral student at the University of Wisconsin-Madison.
- **José Garrido**, Concordia University
- **Noriszura Ismail** is a Professor and Head of Actuarial Science Program, Universiti Kebangsaan Malaysia (UKM). She specializes in Risk Modelling and Applied Statistics. She obtained her BSc and MSc (Actuarial Science) in 1991 and 1993 from University of Iowa, and her PhD (Statistics) in 2007 from UKM. She also passed several papers from Society of Actuaries in 1994. She has received several research grants from Ministry of Higher Education Malaysia (MOHE) and UKM, totaling about MYR1.8 million. She has successfully supervised and co-supervised several PhD students (13 completed and 11 on-going). She currently has about 180 publications, consisting of 88 journals and 95 proceedings.
- **Joseph H.T. Kim**, Ph.D., FSA, CERA, is Associate Professor of Applied Statistics at Yonsei University, Seoul, Korea. He holds a Ph.D. degree in Actuarial Science from the University of Waterloo, at which he taught as Assistant Professor. He also worked in the life insurance industry. He has published papers in *Insurance Mathematics and Economics*, *Journal of Risk and Insurance*, *Journal of Banking and Finance*, *ASTIN Bulletin*, and *North American Actuarial Journal*, among others.
- **Shyamalkumar Nariankadu** - University of Iowa
- **Nii-Armah Okine** is a dissertator at the business school of University of Wisconsin-Madison with a major in actuarial science. He obtained his master's degree in Actuarial science from Illinois State University. His research interests includes micro-level reserving, joint longitudinal-survival modeling, dependence modelling, micro insurance and machine learning.
- **Margie Rosenberg** - University of Wisconsin
- **Emine Selin Sarıdaş** is a doctoral candidate in the Statistics department of Mimar Sinan University. She holds a bachelor degree in Actuarial Science with a minor in Economics and a master degree in Actuarial Science from Hacettepe University. Her research interest includes dependence modeling, regression, loss models and life contingencies.
- **Peng Shi** - University of Wisconsin - Madison
- **Jianxi Su**, Purdue University

- **Tim Verdonck**, KU Leuven
- **Krupa Viswanathan** is an Associate Professor in the Risk, Insurance and Healthcare Management Department in the Fox School of Business, Temple University. She is an Associate of the Society of Actuaries. She teaches courses in Actuarial Science and Risk Management at the undergraduate and graduate levels. Her research interests include corporate governance of insurance companies, capital management, and sentiment analysis. She received her Ph.D. from The Wharton School of the University of Pennsylvania.

Reviewers

The project goal is to have the actuarial community author our textbooks in a collaborative fashion. Part of the writing process involves many reviewers who generously donated their time to help make this book better. They are:

- Chunsheng Ban, Ohio State University
- Vytaras Brazauskas, University of Wisconsin - Milwaukee
- Chun Yong Chew, Universiti Tunku Abdul Rahman (UTAR)
- Eren Dodd, University of Southampton
- Gordon Enderle, University of Wisconsin - Madison
- Rob Erhardt, Wake Forest University
- Liang (Jason) Hong, Robert Morris University
- Hirokazu (Iwahiro) Iwasawa
- Himchan Jeong, University of Connecticut
- Paul Herbert Johnson, University of Wisconsin - Madison
- Samuel Kolins, Lebonan Valley College
- Andrew Soon-Yong Kwon, Zurich Re
- Ambrose Lo, University of Iowa
- Mark Maxwell, University of Texas at Austin
- Tatjana Miljkovic, Miami University
- Bell Ouelega, American University in Cairo
- Zhiyu (Frank) Quan, University of Connecticut
- Jiandong Ren, Western University
- Rajesh V. Sahasrabuddhe, Oliver Wyman
- Ranee Thiagarajah, Illinois State University
- Ping Wang, Saint Johns University
- Chengguo Weng, University of Waterloo
- Toby White, Drake University
- Michelle Xia, Northern Illinois University
- Di (Cindy) Xu, University of Nebraska - Lincoln
- Lina Xu, Columbia University
- Jorge Yslas, University of Copenhagen
- Jeffrey Zheng, Temple University
- Hongjuan Zhou, Arizona State University

Chapter 1

Main Time Line

Placeholder

1.1 Relevance of Analytics

1.1.1 What is Analytics?

1.1.2 Short and Long-term Insurance

1.1.3 Insurance Processes

1.2 Insurance Company Operations

1.2.1 Initiating Insurance

1.2.2 Renewing Insurance

1.2.3 Claims and Product Management

1.2.4 Loss Reserving

1.3 Case Study: Wisconsin Property Fund

1.3.1 Fund Claims Variables: Frequency and Severity

1.3.2 Fund Rating Variables

1.3.3 Fund Operations

1.4 Further Resources and Contributors

Chapter 2

Frequency Modeling

Placeholder

2.1 Frequency Distributions

2.1.1 How Frequency Augments Severity Information

Basic Terminology

The Importance of Frequency

Why Examine Frequency Information

2.2 Basic Frequency Distributions

2.2.1 Foundations

2.2.2 Moment and Probability Generating Functions

2.2.3 Important Frequency Distributions

Binomial Distribution

Poisson Distribution

Negative Binomial Distribution

2.3 The (a, b, 0) Class

2.4 Estimating Frequency Distributions

2.4.1 Parameter estimation

2.4.2 Frequency Distributions MLE

2.5 Other Frequency Distributions

2.5.1 Zero Truncation or Modification

2.6 Mixture Distributions

2.7 Goodness of Fit

2.8 Exercises

2.9 Quiz

2.10 R Code for Plots in this Chapter

2.11 Further Resources and Contributors

Contributors

Chapter 3

Varying Scale Gamma Densities

Placeholder

3.1 Further Resources and Contributors

Chapter 4

Model Selection and Estimation

Placeholder

4.1 Nonparametric Inference

4.1.1 Nonparametric Estimation

Moment Estimators

Empirical Distribution Function

Quantiles

Density Estimators

4.1.2 Tools for Model Selection

Graphical Comparison of Distributions

Statistical Comparison of Distributions

4.1.3 Starting Values

Method of Moments

Percentile Matching

4.2 Model Selection

4.2.1 Iterative Model Selection

4.2.2 Model Selection Based on a Training Dataset

4.2.3 Model Selection Based on a Test Dataset

4.2.4 Model Selection Based on Cross-Validation

4.3 Estimation using Modified Data

4.3.1 Parametric Estimation using Modified Data

Parametric Estimation using Grouped Data

Censored Data

Truncated Data

Parametric Estimation using Censored and Truncated Data

4.3.2 Nonparametric Estimation using Modified Data

Grouped Data

Right-Censored Empirical Distribution Function

Right-Censored, Left-Truncated Empirical Distribution Function

4.4 Bayesian Inference

Chapter 5

Lorenz Curve

Placeholder

Technical Supplement A. Gini Statistic

TS A.1. The Classic Lorenz Curve

TS A.2. Ordered Lorenz Curve and the Gini Index

Ordered Lorenz Curve

Gini Index

TS A.3. Out-of-Sample Validation

Discussion

Chapter 6

Aggregate Loss Models

Placeholder

6.1 Introduction

6.2 Individual Risk Model

6.3 Collective Risk Model

6.3.1 Moments and Distribution

6.3.2 Stop-loss Insurance

6.3.3 Analytic Results

6.3.4 Tweedie Distribution

6.4 Computing the Aggregate Claims Distribution

6.4.1 Recursive Method

6.4.2 Simulation

6.5 Effects of Coverage Modifications

6.5.1 Impact of Exposure on Frequency

6.5.2 Impact of Deductibles on Claim Frequency

6.5.3 Impact of Policy Modifications on Aggregate Claims

6.6 Further Resources and Contributors

Exercises

6.7 Generating Independent Uniform Observations

6.8 Inverse Transform

6.9 How Many Simulated Values?

Chapter 7

Premium Calculation Fundamentals

This is a placeholder file

Chapter 8

Risk Classification

Placeholder

8.1 Introduction

8.2 Poisson Regression Model

8.2.1 Need for Poisson Regression

8.2.2 Poisson Regression

8.2.3 Incorporating Exposure

8.2.4 Exercises

8.3 Categorical Variables and Multiplicative Tariff

8.3.1 Rating Ractors and Tariff

8.3.2 Multiplicative Tariff Model

8.3.3 Poisson Regression for Multiplicative Tariff

8.3.4 Numerical Examples

8.4 Contributors and Further Resources

Further Reading and References

Contributor

8.5 Technical Supplement – Estimating Poisson Regression Models

Chapter 9

Experience Rating Using Credibility Theory

Placeholder

9.1 Introduction to Applications of Credibility Theory

9.2 Limited Fluctuation Credibility

9.2.1 Full Credibility for Claim Frequency

9.2.2 Full Credibility for Aggregate Losses and Pure Premium

9.2.3 Full Credibility for Severity

9.2.4 Partial Credibility

9.3 Bühlmann Credibility

9.3.1 Credibility Z, *EPV*, and *VHM*

9.4 Bühlmann-Straub Credibility

9.5 Bayesian Inference and Bühlmann

9.5.1 Gamma-Poisson Model

9.5.2 Exact Credibility

9.6 Estimating Credibility Parameters

9.6.1 Full Credibility Standard for Limited Fluctuation Credibility

9.6.2 Nonparametric Estimation for Bühlmann and Bühlmann-Straub Models

9.6.3 Semiparametric Estimation for Bühlmann and Bühlmann-Straub Models

9.6.4 Balancing Credibility Estimators

9.7 Further Resources and Contributors

Exercises

Chapter 10

For the gamma distributions, use

Placeholder

10.0.1 Classification Based on Moments

10.0.2 Comparison Based on Limiting Tail Behavior

10.1 Risk Measures

10.1.1 Value-at-Risk

10.1.2 Tail Value-at-Risk

10.1.3 Properties of risk measures

10.1.4 Proportional Reinsurance

Quota Share is Desirable for Reinsurers

Optimizing Quota Share Agreements for Insurers

10.1.5 Non-Proportional Reinsurance

Excess of Loss

10.1.6 Additional Reinsurance Treaties

Surplus Share Proportional Treaty

Layers of Coverage

Chapter 11

Loss Reserving

This is a placeholder file

Chapter 12

Experience Rating using Bonus-Malus

This is a placeholder file

Bonus-Malus

Bonus-malus system, which is used interchangeably as "no-fault discount", "merit rating", "experience rating" or "no-claim discount" in different countries, is based on penalizing insureds who are responsible for one or more claims by a premium surcharge, and awarding insureds with a premium discount if they do not have any claims (Frangos and Vrontos, 2001). Insurers use bonus-malus systems for two main purposes; firstly, to encourage drivers to drive more carefully in a year without any claims, and secondly, to ensure insureds to pay premiums proportional to their risks which are based on their claims experience.

NCD and Experience Rating

No Claim Discount (NCD) system is an experience rating system commonly used in motor insurance. NCD system represents an attempt to categorize insureds into homogeneous groups who pay premiums based on their claims experience. Depending on the rules in the scheme, new policyholders may be required to pay full premium initially, and obtain discounts in the future years as a results of claim-free years.

Hunger for Bonus

An NCD system rewards policyholders for not making any claims during a year, or in other words, it grants a bonus to a careful driver. This bonus principle may affect policy holders' decisions whether to claim or not to claim, especially when involving accidents with slight damages, which is known as 'hunger for bonus' phenomenon (Philipson, 1960). The option of 'hunger for bonus' implemented on insureds under an NCD system may reduce insurers' claim costs, and may be able to offset the expected decrease in premium income.

Chapter 13

Data Systems

Placeholder

13.1 Data

13.1.1 Data Types and Sources

13.1.2 Data Structures and Storage

13.1.3 Data Quality

13.1.4 Data Cleaning

13.2 Data Analysis Preliminary

13.2.1 Data Analysis Process

13.2.2 Exploratory versus Confirmatory

13.2.3 Supervised versus Unsupervised

13.2.4 Parametric versus Nonparametric

13.2.5 Explanation versus Prediction

13.2.6 Data Modeling versus Algorithmic Modeling

13.2.7 Big Data Analysis

13.2.8 Reproducible Analysis

13.2.9 Ethical Issues

13.3 Data Analysis Techniques

13.3.1 Exploratory Techniques

13.3.2 Descriptive Statistics

Principal Component Analysis

13.3.3 Cluster Analysis

13.3.4 Confirmatory Techniques

Linear Models

Generalized Linear Models

Tree-based Models

13.4 Some R Functions

13.5 Summary

13.6 Further Resources and Contributors

Chapter 14

Dependence Modeling

Placeholder

14.1 Variable Types

14.1.1 Qualitative Variables

14.1.2 Quantitative Variables

14.1.3 Multivariate Variables

14.2 Classic Measures of Scalar Associations

14.2.1 Association Measures for Quantitative Variables

Pearson Correlation

14.2.2 Pearson correlation between Claim and Coverage

14.2.3 Pearson correlation between Claim and log(Coverage)

14.2.4 Rank Based Measures

Spearman's Rho

14.2.5 Spearman correlation between Claim and Coverage

14.2.6 Spearman correlation between Claim and log(Coverage)

Kendall's Tau

14.2.7 Kendall's tau correlation between Claim and Coverage

14.2.8 Kendall's tau correlation between Claim and log(Coverage)

14.2.9 Nominal Variables

Bernoulli Variables

Categorical Variables

Ordinal Variables

Parametric Approach Using Normal Based Correlations

Interval Variables

Discrete and Continuous Variables

14.3 Introduction to Copulas

14.4 Application Using Copulas

14.4.1 Data Description

14.4.2 Marginal Models

Chapter 15

Appendix A: Review of Statistical Inference

Placeholder

15.1 Basic Concepts

15.1.1 Random Sampling

15.1.2 Sampling Distribution

15.1.3 Central Limit Theorem

15.2 Point Estimation and Properties

15.2.1 Method of Moments Estimation

15.2.2 Maximum Likelihood Estimation

15.3 Interval Estimation

15.3.1 Exact Distribution for Normal Sample Mean

15.3.2 Large-sample Properties of MLE

15.3.3 Confidence Interval

15.4 Hypothesis Testing

15.4.1 Basic Concepts

15.4.2 Student-t test based on MLE

15.4.3 Likelihood Ratio Test

15.4.4 Information Criteria

Chapter 16

Appendix B: Iterated Expectations

Placeholder

16.1 Conditional Distribution and Conditional Expectation

16.1.1 Conditional Distribution

Discrete Case

Continuous Case

16.1.2 Conditional Expectation and Conditional Variance

Discrete Case

Continuous Case

16.2 Iterated Expectations and Total Variance

16.2.1 Law of Iterated Expectations

16.2.2 Law of Total Variance

16.2.3 Application

Chapter 17

Appendix C: Maximum Likelihood Theory

Placeholder

17.1 Likelihood Function

17.1.1 Likelihood and Log-likelihood Functions

17.1.2 Properties of Likelihood Functions

17.2 Maximum Likelihood Estimators

17.2.1 Definition and Derivation of MLE

17.2.2 Asymptotic Properties of MLE

17.2.3 Use of Maximum Likelihood Estimation

17.3 Statistical Inference Based on Maximum Likelhood Estimation

17.3.1 Hypothesis Testing

17.3.2 MLE and Model Validation