Loss Data Analytics

An open text authored by the Actuarial Community

Contents

Preface

Date: 20 April 2019

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?

The intent is that this type of resource will eventually permeate throughout the actuarial curriculum. 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 recognize that big data (including social media and usage based insurance) are here to stay and that high speed computation is 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 Open Actuarial Textbooks 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
- **Arthur Charpentier**, Université du Quebec á Montreal

- **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

- **Lei (Larry) Hua** is an Associate Professor of Actuarial Science at Northern Illinois University. He earned a PhD degree in Statistics from the University of British Columbia. He is an Associate of the Society of Actuaries. His research work focuses on multivariate dependence modeling for non-Gaussian phenomena and innovative applications for financial and insurance industries.

- **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

Our 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:

- Yair Babab
- 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
- Runhun Feng, University of Illinois
- Liang (Jason) Hong, Robert Morris University
- Fei Huang, Australian National University
- Hirokazu (Iwahiro) Iwasawa
- Himchan Jeong, University of Connecticut
- Min Ji, Towson University
- Paul Herbert Johnson, University of Wisconsin - Madison
- Samuel Kolins, Lebonan Valley College
- Andrew Kwon-Nakamura, Zurich North America
- 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
- Lu Yang, University of Amersterdam
- Jorge Yslas, University of Copenhagen
- Jeffrey Zheng, Temple University

- Hongjuan Zhou, Arizona State University

For our Readers

We hope that you find this book worthwhile and even enjoyable. For your convenience, at our Github Landing site (https://openacttexts.github.io/), you will find links to the book that you can (freely) download for offline reading, including a pdf version (for Adobe Acrobat) and an EPUB version suitable for mobile devices. Data for running our examples are available at the same site.

In developing this book, we are emphasizing the online version that has lots of great features such as a glossary, code and solutions to examples that you can be revealed interactively. For example, you will find that the statistical code is hidden and can only be seen by clicking on terms such as

R Code for Frequency Table

```
Insample <- read.csv("Insample.csv", header=T,  na.strings=c("."), stringsAsFactors=FALSE)
Insample2010 <- subset(Insample, Year==2010)
table(Insample2010$Freq)
```

We hide the code because we don't want to insist that you use the `R` statistical software (although we like it). Still, we encourage you to try some statistical code as you read the book – we have opted to make it easy to learn `R` as you go. We have even set up a separate R Code for Loss Data Analytics site to explain more of the details of the code.

Freely available, interactive textbooks represent a new venture in actuarial education and we need your input. Although a lot of effort has gone into the development, we expect hiccoughs. Please let your instructor know about opportunities for improvement, write us through the discussion features in the online text, or contact chapter contributors directly with suggested improvements.

Chapter 1

Frequency Modeling

Placeholder

1.1 Frequency Distributions

1.1.1 How Frequency Augments Severity Information

Basic Terminology

The Importance of Frequency

Why Examine Frequency Information

1.2 Basic Frequency Distributions

1.2.1 Foundations

1.2.2 Moment and Probability Generating Functions

1.2.3 Important Frequency Distributions

Binomial Distribution

Poisson Distribution

Negative Binomial Distribution

1.3 The (a, b, 0) Class

1.4 Estimating Frequency Distributions

1.4.1 Parameter Estimation

1.4.2 Frequency Distributions MLE

1.5 Other Frequency Distributions

1.5.1 Zero Truncation or Modification

1.6 Mixture Distributions

1.7 Goodness of Fit

1.8 Exercises

1.9 R Code for Plots in this Chapter

1.10 Further Resources and Contributors

Contributors