

**Computational Actuarial Science with R**

A. CHARPENTIER (ed.), 2015

Boca Raton, CRC Press

xxxii + 618 pp., \$73.79

ISBN 978-1-466-59259-9

Actuarial science helps to assess and manage financial risks faced by governments, businesses and individuals when they plan for the future. The traditional actuarial skill set includes computational techniques from mathematics, probability, statistics, econometrics, finance and demography. The book aims at providing a broad introduction to these techniques by using the R language, which has become the *de facto* standard for statistical computing and data analysis for statisticians and has gained much popularity among actuaries and other scientists working with data.

The book's first chapter presents an interesting and useful introduction to the R language, after which 15 chapters follow in four parts. Part I, 'Methodology', consists of five chapters covering standard parametric inference for univariate and multivariate statistical distributions that are commonly used in actuarial science, the Bayesian philosophy, statistical methods of learning from data and predictive modelling of a binary random variable, spatial analysis with an application to car accidents and some techniques from extreme value theory applied to reinsurance pricing. Part II, 'Life insurance', comprises four chapters covering standard financial and actuarial calculations for life contingent risk, various extrapolative methods for forecasting mortality rates, an operational framework for constructing and validating prospective mortality tables specific to an insurer and techniques used in survival analysis when dealing with censored data. Part III, 'Finance' consists of three chapters covering modelling univariate time series of financial returns, some standard techniques for modelling yield curves and interest rates, and portfolio optimization problems using alternative risk measures. Part IV, 'Non-life insurance' has three chapters covering generalized linear models fitted to cross-sectional data for general insurance pricing, linear models fitted to longitudinal data and model-based forecasting of claims based on historical data for reserving.

As it is impossible to review all the chapters in any detail here, I have picked the following three for a little more detailed review. The chapters selected are conceptually very different (Chapter 4 is largely a review of the predictive modelling techniques that are suitable for a particular type of data, Chapter 7 is essentially a review of functions in a popular R package and Chapter 11 is built around a well-known time series model), which highlights

the value of the book as an instrument of learning a broad range of topics encountered in actuarial practice.

In Chapter 4, a credit scoring rule to classify a new applicant as a good credit risk, or a bad one (not creditworthy), based on values for one or more predictors from a well-known data set of past credit applications is developed. The first section describes the data and reviews techniques and ideas that are pertinent to measuring the predictive power of models, including scoring tools, receiver operating characteristic curves, the area under the receiver operating characteristic curve, the Kolmogorov–Smirnov statistic, sensitivity, specificity and more. The next two sections are devoted to logistic regression, a special type of the generalized linear model where the distribution of the response variable is binary, and its penalized versions known as ridge regression and lasso regression. The last two sections deal with decision trees and methods based on multiple decision trees, namely bagging, boosting and random forest. It can be mentioned here that the `caret` package that has been developed to streamline the process of creating various predictive models is not mentioned in these sections.

In Chapter 7, the standard financial and actuarial calculations for life contingent risks implemented in the R package `lifecontingencies` are discussed by a means of examples of computations required to answer some commonly asked questions in this area of actuarial science. The first 10 examples are from financial mathematics, which deals with the time value of money. In addition to financial calculations, actuarial mathematics involves probabilities that are commonly referred to as life tables. The next six examples show how life tables can be created and some basic demographic analysis performed by using functions in the package. Taking into account the time value of money and the life-related probabilities, the expected values of the future cash flows currently valued, which are also known as actuarial present values, can be calculated. These are then used in pricing life insurance, annuity and endowment contracts, which is covered by the next nine examples. There are also examples on reserving life insurances, multiple-life-insurance theory, incorporating expenses in calculating premiums and reserves, and stochastic simulation of life contingent insurances. The total of 42 examples is supported by 27 exercises covering all of the aforementioned topics. The standard actuarial notation is used through the chapter.

In Chapter 11, the common properties of financial time series are first discussed by using a data set of daily exchange rates as an example. Using

simple code snippets, it is explained why for these data the normal distribution is not adequate to describe the log-returns and shown that the absolute values of return, which can be used as a proxy for volatility, are auto-correlated, indicating the time dependence of the volatility. An implementation of the generalized auto-regressive conditional heteroscedasticity GARCH(1,1) model with Student t -distributed innovations is then detailed by using the same approach of providing snippets of R code. The chapter proceeds by showing how to define print and predict and other common methods for the newly created class of an R object, how to profile the R code with `Rprof()` and `summaryRprof()` functions to identify the most time-consuming parts of the code, and how to make the recursive estimation of the conditional variance more than 10 times faster via a direct translation of the relevant R function to C or C++ code by using the `Rccp` package. The chapter ends with an example of application of the fitted GARCH model for estimating the value at risk, where the results of several alternative approaches to this problem are compared. No exercises are provided.

Having examined all the other chapters, I can confirm that the entire book is certainly worthwhile reading and can be recommended to anyone who is interested in the computational aspects of actuarial science. The book contains many detailed worked examples, with R code fully integrated into the text. Its index is useful and the references are copious. Although many chapters are well written and free of typographical errors, some others have omissions and inadequacies even when describing their own structures (e.g. pages 289 and 410). Contrary to what the book says in the preface, the `CASdatasets` package is not available from the Comprehensive R Archive Network, simply because some of the data sets are too large. The package is available from <http://cas.uqam.ca>. The book's code repository at <https://github.com/CASwithR> is far from complete, with R code available for only three chapters.

Overall, the book provides information and code that readers with any quantitative background can gain something from. It will naturally appeal to actuaries of all calibres, but it has a much wider audience of quantitative analysts using R for statistical modelling and data analysis in various fields. There are also good reasons to recommend this book to any science library.

Andrey Kostenko
Western Power
Perth
E-mail: akoste01@gmail.com

Analysis of Multivariate and High Dimensional Data

I. KOCH, 2013
Cambridge, Cambridge University Press
526 pp., £55.00
ISBN 978-0-521-88793-9

The need for tools to conduct the analysis of high dimensional data has increased significantly in the last decade owing to both computational and data advances. The development of such tools is mostly governed by computationally based techniques and such techniques are in turn based on statistical theory. Such theory has either been developed specifically for high dimensional data situations or has been either extended or derived from classical concepts and further modified for a high dimensional data environment. This book represents a successful attempt to connect and compile such developments from past to recent in one place and to illustrate their interconnections.

The book is developed in 13 chapters which are suitably divided into three sections. The first two sections comprising eight chapters are devoted to the classical methods and the third section consists of five chapters which explore recent methods.

The book begins with an explanation of the basics of random vectors and matrices in Chapter 1 which are required to understand later topics. Chapters 2 and 3 describe principal component analysis and canonical correlation analysis respectively along with their standard statistical properties. Their applications are illustrated by using several data-based analyses with graphics. Such analysis is extended for a high dimensional data environment and some applications are illustrated in detail. More recent results are presented which are connected to the classical results. Some suggestions and guidelines for using these classical tools in high dimensional data situations are made for data analysts. Next Chapter 4 considers discriminant analysis in a classical set-up and then details more recent developments. The connection between discriminant analysis and both principal component analysis and canonical correlation analysis is explained in this chapter. Related issues are illustrated by using data-based examples. The presentations on principal component analysis, canonical correlation analysis and discriminant analysis in Chapters 2–4 are quite different from other books on multivariate analysis. Furthermore, Chapters 6 and 7 concern topics and issues in cluster analysis and factor analysis respectively. The theoretical concepts are clearly