Data Analytics for Actuarial and Financial Modeling

Emiliano A. Valdez *
University of Connecticut, Storrs, Connecticut USA 06269-1009

VNU University of Science, Hanoi, Vietnam
3 January 2019

Abstract

This workshop provides a survey of the use of modern statistical techniques for actuarial and financial applications. There has been an increased interest in the use of data science to deliver predictive models applicable in all areas of insurance and financial security programs: life and health, property and casualty, pensions. Such predictive models are generally used for improved decision making in: (1) delivering consumer products that are financially sound, (2) developing better relationships with consumers to meet their evolving needs, and (3) evaluating sufficient funding for promised benefits from insurance and pension programs, to name a few. We will conduct this workshop by introducing a little bit of the theory, but we will also simultaneously incorporate practical case studies and how data and models can be analyzed using R. In particular, we cover:

- background of regression and generalized linear models (GLMs)

  For background, we will review the basic concepts of regression and its extension to generalized linear models. These models have become popular actuarial tools for better understanding claims prediction, pricing, risk classification, and even claims reserving.

- penalized regression and likelihood

  Traditional methods of estimating regression and generalized linear models are extended to introduce the concept of regularization by introducing a penalty function. This is typically accomplished because estimates from traditional methods of least squares or likelihood may be distorted or meaningless.

- regression and classification trees

  We explore the usefulness of decision trees as a predictive model. In decision trees, the regions of the explanatory variables are repeatedly partitioned to create a tree-based structure that can be used to predict the outcome variable. This is an alternative to classical regression and GLMs that have been popularized for decades, and we will assess how trees perform relative to these classical models.

- data clustering

  A form of unsupervised learning, data clustering is the process of dividing a group of observations into homogeneous groups or clusters. Observations in the same cluster share "similar" features while those from different clusters are "dissimilar". This method has many uses in "big data" analytics but it has a large potential in actuarial applications: understanding and monitoring claims, valuation of large portfolios of variable annuities.

*email: emiliano.valdez@uconn.edu website: http://www.math.uconn.edu/~valdez/

- principal component analysis (PCA)

  PCA is a technique to simplify a high dimensional dataset by transforming or reducing the dimension of the data while keeping to preserve its original features. Such high dimensional data presents computational challenges and PCA has been used to reduce the burden of such datasets. Here, we present the procedure used to calculate principal components and we demonstrate how this technique can be used for forecasting currency exchange rates.

About the Speaker

Emiliano (Emil) A. Valdez, Ph.D., FSA, is a Professor of Actuarial Science in the Department of Mathematics at the University of Connecticut, USA. He is a Fellow of the Society of Actuaries, holds a Ph.D. from the University of Wisconsin in Madison, and is also the interim director of the graduate programs in actuarial science at Connecticut. His academic experience includes several years of teaching and conducting research in actuarial science in three different continents: North America, Australia and Asia. His previous academic posts include working for the Nanyang Business School in Singapore and for the University of New South Wales in Sydney, Australia. From 2013 to 2015, he was the Director of the actuarial science program at Michigan State University. He has been awarded the Edward A. Lew Award, the Halmstad Memorial Prize, and recently in 2010, the Charles A. Hachemeister Prize, in recognition for his significant contributions to the actuarial literature. His current research interest includes copula models and dependencies, managing post-retirement assets, and risk measures and capital requirements related to enterprise risk management. In addition, he has several years of industry experience working as an actuary for Connecticut Mutual in Hartford and held summer actuarial positions at Price Waterhouse.

REFERENCES

Gan, G. and Valdez, E.A. (2018), *Actuarial Statistics with R: Theory and Case Studies*, ACTEX publications: New Hartford, Connecticut.