1.1 Introduction to Actuarial Data Science

Actuarial Data Science Online Textbook

Fei Huang, UNSW Sydney

With the rise of AI and Big Data, every firm has effectively become a data-driven enterprise. Whether it's supermarkets, hairdressers, restaurants, or dentists, businesses of all types are collecting data on customer behavior and engagement. We are living in a data-driven era, where data has become one of the most valuable assets for any business. However, possessing data alone doesn't automatically translate into insights or business value. It is crucial to apply the right data science techniques to extract meaningful insights and effectively communicate those findings to stakeholders.

There are many textbooks available today that teach data science, statistical methods, and machine learning techniques. So, what does this textbook offer that's different? While many excellent data science textbooks exist, they often focus on individual techniques or models without addressing the entire problem-solving process in a business setting. In contrast, this textbook aims to equip readers with the skills to solve data problems in real-world business environments. We guide you through the complete process: from asking the right questions and performing exploratory data analysis, to modeling, interpreting results, communicating findings, and considering ethical implications.

This aim of this textbook is consistent with that of the Australian Actuaries Institue for their Part II Data Science Principles sysllabus.

> "Data Science Principles – aims to extend students' knowledge of modern analytical tools and techniques beyond those introduced in the Foundation Program subjects and to teach students how to apply this knowledge in real-life business settings" — Actuaries Institute

Data science is an interdisciplinary subject that covers many different subjects, including but not limited to Statistics, Machine Learning, Database,Optimization, Algorithm and Programming, Domain knowledge in a business setting. This book mainly focuses on applying the data analysis cycle with tatistical machine learning techniques to tackle actuarial applications, called Actuarial Data Science. The techniques and concepts introduced in this book can be widely applied to other business problems. So this book can also be used as a textbook to solve general business data problems.

It is easy to argue that Data Science is a science and engineering related subject with many quantitative tools. However, it is important to bear in mind that Data Scinece is also an *art* (Peng and Matsui 2015). It is not something yet that we can teach to a computer. Although there are many data analysis tools, from linear regression to classification trees and even deep learning, and these tools have all been coded as software packages, however, a data scientist must make lots of judgement calls during the problem solving process, select a set of tools for each specific data problem, and communicate the results in an appropriate format to stakeholders. Those judgements and communications are not able to be delivered by a computer, at least for now.

Data Analysis Cycle (DAC)

Actuaries apply Actuarial Control Cycle (ACC) as shown in figure Figure 1 for problem solving. Figure 1 is taken from Bellis et al. (2003).

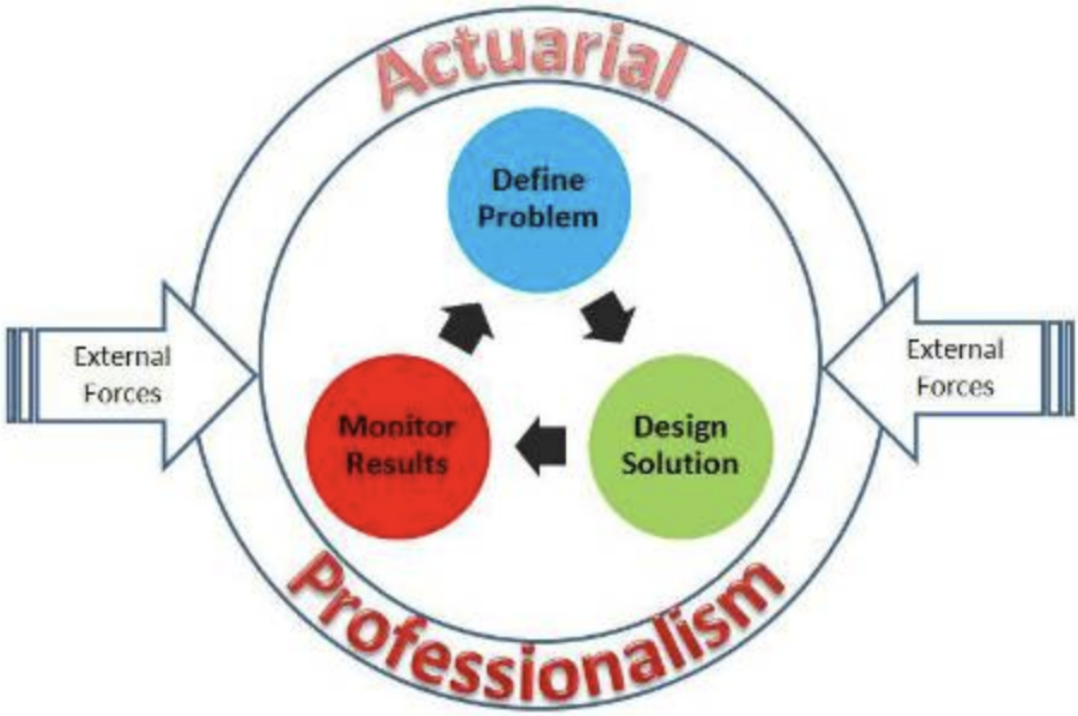

Figure 1: Actuarial Control Cycle

Similarly, we develop Data Analysis Cycle (DAC) to solve data problems. Data analysis is a highly iterative and non-linear process including six key steps. The data analysis process is a specific application of the Actuarial Control Cycle and we call it Data Analysis Cycle (DAC).

The six steps of DAC is listed below:

1. State the question
2. Explore the data
3. Build the models
4. Interpret the results
5. Communicate the results
6. Deploy the model

Throughout the process, we always need to consider ethics and professionalism. These six steps are all built on deep understanding of the business context.

In the following chapters we illustrate how to apply the Data Analysis Cycle (DAC) to solve real-world business problems.

Bellis, Clare, Richard Lyon, Stuart A Klugman, and John Shepherd. 2003. *Understanding Actuarial Management: The Actuarial Control Cycle*. The Institute of Actuaries of Australia; the Society of Actuaries.

Peng, Roger D, and Elizabeth Matsui. 2015. *The Art of Data Science: A Guide for Anyone Who Works with Data*. Skybrude Consulting, LLC.