Insurance Data Science Conference

Stan in Insurance Workshop

16th - 17th July 2018

TABLE OF CONTENTS

Welcome to the 1st Insurance Data Science conference

The first Insurance Data Science conference at Cass Business School, on 16 July, follows on from five successful iterations of the *R in Insurance* series. This one-day conference brings together practitioners and academics to share perspectives on the advances, opportunities and challenges engendered by the expansion of data science in insurance businesses. Presentations will address areas including risk assessment, customer analytics, machine learning, pricing, capital management, catastrophe and econometric modelling.

The Stan in Insurance Workshop offers an introduction to cutting-edge techniques in multi-level regression modelling, with hands-on insurance applications.

We hope that you find the conference and workshop enjoyable and stimulating.

We are grateful to our sponsors RStudio, Mango Solutions, Verisk, Mirai Solutions, Oasis, and Milliman. Without their generous support, holding this conference would not have been possible.

Furthermore, we thank the Cass Faculty Administration team, who have worked tirelessly to make this conference a success.

See you in London!

Markus Gesmann & Andreas Tsanakas

On behalf of the Scientific Committee:

Katrien Antonio, Christophe Dutang, Markus Gesmann, Ioannis Kyriakou, Davide de March, Lola Martínez-Miranda, Jon Sedar, David Smith, Giorgio Alfredo Spedicato, Jaap Spreeuw, Andreas Tsanakas.

Insurance Data Science Conference Programme – 16/07/2018

Venue: Room LG001, Cass Business School, 106 Bunhill Row, EC1Y 8TZ

[08:30-09:00] Registration

[09:00-10:00] Keynote 1: Gareth Peters
- Modelling longevity risk for retirement and pension planning

[10:00-11:00] Session 1: Pricing / Claims modelling
- Sparsity with multi-type lasso penalties - Tom Reynkens
- Statistical analysis of weather-related property insurance claims - Christian Rohrbeck
- Claims frequency modelling using telematics car driving data - Mario Wüthrich

[11:00-11:30] Coffee

[11:30-12:30] Session 2: Lightning talks
- Machine learning for actuaries: An introduction - Valerie du Preez, Steven Perkins
- Truncated regression models for the analysis of operational losses due to fraud: A high performance computing implementation in R - Alberto Glionna
- Using Random Forest to estimate risk profiles and probability of breakdowns - Lara A. Neira Gonzalez
- Simulating economic variables using graphical models - Aniketh Pittea
- RShiny at Qatar Re: A business case study - Marc Rierola
- PnC reinsurance modeling using Python and TensorFlow - Pauli Rämö
- 'KSgeneral' : A package for fast, exact, Komogorov-Smirnov goodness of fit testing - Senren Tan

[12:30-13:30] Lunch

[13:30-14:30] Session 3: Capital and exposure modelling
- Statistical learning for portfolio tail risk measurement - Michael Ludkovski
- Reverse sensitivity testing: What does it take to break the model? - Silvana M. Pesenti
- Using R for catastrophe modelling of cyber risks in (re)insurance - Benjamin C. Dean

[14:30-15:00] Session 4: Panel discussion
- Digital transformation of insurance business: opportunities and challenges

[15:00-15:30] Coffee

[15:30-16:30] Session 5: Business Case Studies
- SFCR automated analysis using scraping, text mining and machine learning methods for benchmarking and capital modelling - Aurelien Couloumy
- Machine learning and fairness for commercial insurance - Oliver Laslett
- Getting value out of machine learning - Javier Rodriguez Zaurin

[16:30-17:30] Keynote 2: Eric Novik

- Making decisions under uncertainty using Bayesian inference and Stan

[18:00-22:00] Reception drinks & Dinner - Ironmongers' Hall

Stan in Insurance Workshop Programme – 17/07/2018

Venue: Room 2006, Cass Business School, 106 Bunhill Row, EC1Y 8TZ

[08:30-09:00] Registration

[9:00 - 10:30] Eric Novik

- Intro to Stan, including:
    - Coding linear regression to assess wine quality
    - Demonstrating important parts of the Stan program
    - Doing some basic posterior predicting checking
    - Introduction to calibration and model comparison
    - Introduction to making decisions with Bayesian models

[10:30 - 11:00] Coffee

[11:00 - 12:30] Paul-Christian Bürkner

- From classical GLMs to multi-level models
    - Comparing classical GLMs with bayesian GLMs using rstanarm
    - Building more complex multi-level models using brms
    - Examples from pricing and claims reserving

[12:30 - 13:30] Lunch

[13:30 - 14:30] Mick Cooney & Jake Morris

- Case studies from the insurance industry
    - Loss development curves in Stan (Mick Cooney)
    - Hierarchical compartmental reserving models (Jake Morris)

[14:30 - 15:00] Coffee

[15:00 - 17:00] Working in groups with support of the presenters

- Work on your own problems or work through on of the following examples:
    - Loss curve case study (Mick Cooney)
    - Extreme value case study (Aki Vehtari)
    - Various insurance related examples from Markus Gesmann

- Examples from the 'brms' vignettes (Paul-Christian Bürkner)

Abstracts

Modelling longevity risk for retirement and pension planning

Gareth Peters, Heriot-Watt University, Edinburgh

In this talk we build on a sequence of papers recently developed to enhance the modelling of life expectancy based on mortality data. Forecasting life expectancy and mortality are two important aspects for the study of demography that influence pension planning, retirement decisions and government policy.

We demonstrate how to develop regression models incorporating stochastic factors such as graduation temporal effects, period effects, cohort effects, stochastic volatility and long memory to enhance the forecasting and estimation of life tables. In addition, we show the mispricing that occurs in standard annuities, pure endowments and Guaranteed Annuity Options (GAOs).

Sparsity with multi-type lasso penalties

Tom Reynkens, KU Leuven

Insurance companies use predictive models for a variety of analytic tasks including pricing, marketing campaigns, and fraud and churn detection. In practice, these predictive models often use a selection of continuous, ordinal, nominal and spatial variables to detect different risks.

Such models should not only be competitive, but also interpretable by stakeholders (including the policyholder and the regulator), and easy to implement and maintain in a production environment. Therefore, current actuarial literature puts focus on generalised linear models (GLMs) where ad hoc techniques or professional expertise are applied to bin (recombine levels) or remove variables. In the statistical and machine learning literature, penalised regression is often used to obtain an automatic data-driven method for variable selection and binning in predictive modelling. Most penalisation strategies work for data where all predictors are of the same type, such as Lasso for continuous variables and Fused Lasso for ordinal variables.

We design an estimation strategy for regularised GLMs which includes variable selection and binning through the use of multi-type Lasso penalties. We consider the joint presence of different types of variables and specific penalties for each type. Using the theory of proximal operators, our estimation procedure is computationally efficient, splitting the overall optimisation problem into easier to solve subproblems per variable and its associated penalty. As such, we are able to simultaneously select, estimate and group, in a statistically sound way, any combination of continuous, ordinal, nominal and spatial variables. We illustrate the approach, which is implemented in an R package, using a case-study on motor insurance pricing.

Statistical analysis of weather-related property insurance claims

Christian Rohrbeck, Lancaster University

This talk considers the association between weather events, such as rainfall or snow-melt, and the number of water-related property insurance claims. Weather events which cause severe damages are of general interest; decision makers want to take efficient actions against them while the insurance companies want to set adequate premiums. The modelling is challenging since the underlying dynamics vary across geographical regions due to differences in topology, construction designs and climate.

This talk presents new methodology which improves upon the existing models which in particular fail to model high numbers of claims. The approach includes a clustering algorithm which aggregates claims over consecutive days based on the observed weather metrics and derives more informative explanatory variables. In combination with a statistical framework based on mixture modelling, we achieve a better understanding of the association between claims and weather events. Both the clustering algorithm and the estimation of the statistical model are performed using R. The benefits of the methodology are illustrated by applying it to insurance and weather data from 1997 to 2006 for three Norwegian cities: Oslo, Bergen and Bærum.

Claims frequency modelling using telematics car driving data

Mario Wüthrich, RiskLab, ETH Zurich

We investigate the predictive power of covariates extracted from telematics car driving data using the v-a heatmaps of Gao and Wüthrich (2017) for claims frequency modeling. These telematics covariates include the K-means classification, the principal components, and the bottleneck activations from a bottleneck neural network. It turns out that the predictive power of the first principal component and the bottleneck activations are more significant in predicting claims frequencies than driver's age. For this reason we recommend the use of these telematics covariates for car insurance pricing.

Machine learning for actuaries: An introduction

Valerie du Preez, Dupro

Co-authors: Steven Perkins, Zhixin Lim

Improvements in computational power has given rise to the use of machine learning techniques in a wide variety of areas, including finance, driverless cars, image detection, speech recognition etc. In a world of high volume and varied datasets, machine learning techniques are an essential toolkit to provide actionable insights from the data.

The exponential increase in data generation, capture and storage along with improved computer power is likely to benefit actuaries in two key ways. Firstly, improved data and

computational capabilities is likely to mean that traditional actuarial tasks can be tackled with increasingly sophisticated approaches. The second opportunity arises because many actuaries will have the necessary skills to capitalise on new opportunities which arise to expand the profession into new areas.

This session will cover machine learning at a high level and will include an overview of case studies performed by the Modelling Analytics and Insights from Data Working Party from the Institute of Actuaries in the UK, to explore how additional techniques around data analysis could be utilised in the future. The views presented will be the presenters' own.

Truncated regression models for the analysis of operational losses due to fraud: A high performance computing implementation in R

Alberto Glionna, Generali, Italy

Co-authors: Giovanni Millo and Nicola Torelli

In operational risks modeling, losses arising from fraud are often collected only when exceeding a given monetary threshold. With reference to regression-like problems, in econometrics and statistics, those type of observations are known as truncated. In general, great care is needed when truncated data are used in predictive analysis, or more broadly risk modeling, which are aiming at inferring the model for the data. With respect to this, the dualistic frequency-severity nature of fraud and the inevitable truncation of the information challenge the traditional methods which may lead to inaccurate results. In this paper it is shown that more accurate analyses are possible whenever some conditions are met. A technique called Augmented (Generalized) Linear Model (AGLM) have been developed: it is aimed at managing the problem of truncated fraud data and allow for theoretically unbiased estimates of operational losses. This technique has been implemented and used in Assicurazioni Generali.

Unfortunately, in this context inference on relevant quantities (such as confidence intervals) can be obtained by using computationally heavy solutions with complex and time consuming algorithms. In this paper we will discuss elements of the above mentioned models alongside a cross-platform, user friendly high performance computing implementation in R that considerably reduces the computational time, overcoming one of the main limitations of AGLMs. In particular the packages 'parallel' and 'shiny' will be exploited to speed up both parametric and non-parametric bootstrap on any multicore architecture, up to distributed networks of workstations, through a user-friendly application with an intuitive graphical interface, not requiring any previous knowledge of R.

Using Random Forest to estimate risk profiles and probability of breakdowns

Lara A. Neira Gonzalez, University of Edinburgh

Co-authors: Martin Kreer, Jose-Maria Guerra, Alfredo D Egidio dos Reis

Over the last decade, machine learning techniques have become more popular on different background such as genetics, finance and health. The issues that classical statistical techniques have on Big Data situations such as p << n, make Machine Learning algorithms quite attractive for usage in real application studies. A precise estimation of breakdowns cannot only be applied in predictive maintenance but also for the calculation of insurance premiums of industrial equipment. The aim of our study is to estimate the probability of breakdowns using a Machine Learning technique on machine data using training and test datasets.

Random Forest, a supervised non-parametric technique based on the AUC variable importance measure, was applied 1000 times under the null hypothesis and once under the alternative on our training sample in order to calculate an empirical p-value. Then, the empirically significant variables from the training sample were tested for significance using general linear regression on the independent test sample corrected by multiple testing. After obtaining the risk factors, the probability of breakdowns was estimated.

All of 32 variables of one machine and all of 6 variables of another machine were empirically significant. After testing on the independent test dataset, 30 variables showed significance after Bonferroni correction (considering 90%, 95% and 90% of confidence levels); 29 with a p-value < 0.01 and 1 with a p-value < 0.10 on the first machine. The most significant variable showed a $R^2 = 18.2\%$. Considering a model with all significant variables, our model reached an accuracy greater than 80% in an independent test dataset when predicting a breakdown. In the other machine, 5 variables were significant, 3 having a significance level of 0.01, one at 0.05 significant level and another one at 0.10. The most significant variable explained more than 85% of the variance ($R^2 = 85.93$). Considering models with only this variable, we predicted the probability of breakdown with an accuracy of 100% in the independent test dataset (AUC = 1).

Both our studies found significant factors which help us to better understand the insights of the failure of the two machines. Calculating the empirical p-value based on 1000 outputs under the null makes the estimation stable and reliable. Using Random Forest for diagnosis helps us to reach high levels of accuracy on predicting breakdowns on both machines. In both cases our model can predict better than a human would do it, predicting perfectly the breakdowns of one machine. The probability of breakdowns based on our models is a good estimation to use when calculating premiums, and they can be used for creating risk profiles as our models could detect different significant factors which are playing an important role in the process of each machine.

Simulating economic variables using graphical models

Aniketh Pittea, University of Kent

Co-authors: Jaideep Oberoi and Pradip Tapadar

Actuaries and financial risk managers use an Economic Scenario Generator (ESG) to identify, manage and mitigate risks at a range of horizons. In particular, pension schemes and other long term businesses require ESGs to simulate projections of assets and liabilities in order to devise adequate risk mitigation mechanisms. This requires ESGs to provide reasonable simulations of the joint distribution of several variables that enter the calculation of assets and liabilities. In this paper, we discuss how a graphical model approach is used to develop an ESG, and also provide a specific application.

A wide range of ESGs are currently in use in industry. These models have varying levels of complexity and are often proprietary. They are periodically recalibrated, and tend to incorporate a forecasting dimension. For instance, they may incorporate a Vector Auto Regression model. Alternatively, many rely on a cascading structure, where the forecast of one or more variables is then used to generate values for other variables, and so on. In each case, these models balance the difficult trade-off between accurately capturing both short and long term dynamics and interdependences. We argue that, for the purpose of risk calculations over very long periods, it may be easier and more transparent to use a simpler approach that captures the underlying correlations between the variables in the model. Graphical models achieve this in a parsimonious manner, making them useful for simulating data in larger dimensions. In graphical models, dependence between variables is represented by "edges" in a graph connecting the variables or "nodes". This approach allows us to assume conditional independence between variables and to set their partial correlations to zero. The two variables could then be connected via one or more intermediate variables, so that they could still be weakly correlated.

We compare different algorithms to select a graphical model, based on p-values, AIC, BIC, and deviance using R (and provide the relevant packages). We find them to yield reasonable results and relatively stable structures in our example. The graphical approach is fairly easy to implement, is flexible and transparent when incorporating new variables, and thus easier to apply across different datasets (e.g. countries). Similar to other reduced form approaches, it may require some constraints to avoid violation of theoretical rules. It is also easy to use this model to introduce arbitrary economic shocks. We provide an example in which we identify a suitable ESG for a pension fund in United Kingdom that invests in equities and bonds, and pays defined benefits. While more complex modelling of the short term dynamics of processes is certainly feasible, our focus is on the joint distribution of innovations over the long term. To this end, we simply fit an autoregressive process to each of the series in our model and then estimate the graphical structure of the contemporaneous residuals. We find that simulations from this simple structure provide plausible distributions that are comparable to existing models. We also discuss how these models can be used to introduce nonlinear dependence through regime shifts in a simple way.

Overall, we argue that this approach to developing ESGs is a useful tool for actuaries and financial risk managers concerned about long term portfolios.

RShiny at Qatar Re: A business case study

Marc Rierola, Qatar Re

Qatar Re as a global multi-line reinsurer has to deal with a multitude of different loss files (result files from probabilistic catastrophe models) from different sources. Loss files can be provided by the clients/brokers, created by our actuaries, the Catastrophe Pricing Analytics team or Risk Management. While all those loss files are being presented in a different format and being stored in different places using different systems it becomes rather difficult to make good use of all the available information across the different departments. We are currently building a Loss File Database within Qatar Re which shall function as a central storage place for all different forms of Loss files. A place where each department can upload their data in a predefined format and any other department can easily access said data. RShiny plays a crucial role in this process as we are building an app that allows to upload, examine, scale, manipulate, download Loss files. This RShiny application shall become the single point of contact for anybody in the company that is seeking Loss File information.

PnC reinsurance modeling using Python and TensorFlow

Pauli Rämö, Mirai Solutions

Insurance companies require detailed (and ideally actionable) insights into the sources and contributors of their risks arising from losses modeled given premium exposure. It is crucial to have good monitoring and valid predictions of the evolving risks, for example when determining an adequate reinsurance strategy. For rare events incurring large losses, particularly large scale simulations are required in order to obtain realistic and stable risk estimates, such as Expected Shortfall (ES / CVaR) or Value at Risk (VaR).

Google's TensorFlow is a modern IT framework designed for big data analytics, particularly in the field of (deep) machine learning. TensorFlow has attracted a lot of attention over the past couple of years and provides several advantages: definition of computational graphs, lazy execution, improved performance, established framework with active development community support, visualization and profiling tools, and seamless CPU/GPU/TPU support/switching out-of-the-box.

Somewhat out of focus, with all the buzz around deep learning and artificial intelligence, TensorFlow can also be applied to any other computational task suitable to "tensor mathematics". In this talk we demonstrate how this can work in the case of reinsurance modeling. We simulate large-scale gross loss data through convolution of frequency and severity distributions and then apply a realistic set of proportional reinsurance contracts (excess of loss, surplus share) to calculate net losses and various risk measures. We show hands-on implementation using Python and TensorFlow and discuss the advantages and challenges this creates.

'KSgeneral' : A package for fast, exact, Komogorov-Smirnov goodness of fit testing

Senren Tan

We have developed an R package 'KSgeneral' (available from https://CRAN.R-project.org/package=KSgeneral) that computes the (complementary) cumulative distribution function (cdf) of the one-sample two-sided (or one-sided, as a special case) Kolmogorov-Smirnov (KS) statistic, for any fixed critical level, and an arbitrary, possibly large sample size for a pre-specified purely discrete, mixed or continuous cdf under the null hypothesis. If a data sample is supplied, 'KSgeneral' also computes the p-value corresponding to the value of the KS test statistic computed based on the user provided data sample. The package 'KSgeneral' implements a novel, accurate and efficient method named Exact-KS-FFT, developed by Dimitrova, Kaishev, Tan (2017), available together with the underlying C++ code from http://openaccess.city.ac.uk/18541. The p-value is expressed as a double-boundary non-crossing probability for a homogeneous Poisson process, which is then efficiently computed using Fast Fourier Transform (FFT).

Statistical learning for portfolio tail risk measurement

Michael Ludkovski, UC Santa Barbara Dept of Statistics; Visiting Professor, London School of Economics

Co-authors: Jimmy Risk (Cal Poly Pomona Dept of Mathematical Sciences)

We consider calculation of VaR/TVaR capital requirements when the underlying economic scenarios are determined by simulatable risk factors. This problem involves computationally expensive nested simulation, since evaluating expected portfolio losses of an outer scenario (aka computing a conditional expectation) requires inner-level Monte Carlo. We introduce several inter-related machine learning techniques to speed up this computation, in particular by properly accounting for the simulation noise. Our main workhorse is an advanced Gaussian Process (GP) regression approach co-authored by the second author as part of the hetGP R package. hetGP constructs a heteroskedastic spatial model to efficiently learn the relationship between the stochastic factors defining scenarios and corresponding portfolio value. Leveraging this emulator, we develop sequential algorithms that adaptively allocate inner simulation budgets to target the quantile region, akin to Bayesian contour-finding. The GP framework also yields better uncertainty quantification for the resulting VaR/estimators that reduces bias and variance compared to existing methods.

Reverse sensitivity testing: What does it take to break the model?

Silvana M. Pesenti, Cass Business School

Co-authors: Pietro Millossovich, Andreas Tsanakas

Sensitivity analysis is an important component of model building, interpretation and validation. A model comprises a vector of random input factors, an aggregation function mapping input factors to a random output, and a (baseline) probability measure. A risk measure, such as Value-at-Risk and Expected Shortfall, maps the distribution of the output to the real line. As is common in risk management, the value of the risk measure applied to the output is a decision variable. Therefore, it is of interest to associate a critical increase in the risk measure to specific input factors. We propose a global and model-independent framework, termed 'reverse sensitivity testing', comprising three steps:

- an output stress is specified, corresponding to an increase in the risk measure(s);
- a (stressed) probability measure is derived, minimising the Kullback-Leibler divergence with respect to the baseline probability, under constraints generated by the output stress;
- changes in the distributions of input factors are evaluated.

We argue that a substantial change in the distribution of an input factor corresponds to high sensitivity to that input and introduce a novel sensitivity measure to formalise this insight. Implementation of reverse sensitivity testing in a Monte-Carlo setting can be performed on a single set of input/output scenarios, simulated under the baseline model. Thus the approach circumvents the need for additional computationally expensive evaluations of the aggregation function. We illustrate the proposed approach through a numerical example of a simple insurance portfolio and a model of a London Insurance Market portfolio used in industry.

Using R for catastrophe modelling of cyber risks in (re)insurance

Benjamin C. Dean, Iconoclast Tech LLC

Modelling of cyber risks is an emerging domain in the field of catastrophe modelling. Insurers and re-insurers increasingly wish to model cyber risks given the fast growth of stand-alone cyber insurance product lines and the potential 'silent' cyber risk exposure of existing property and casualty policies.

Analysts face a number of challenges in modelling these risks. For instance, the cyber ecosystem contains complex dependencies, which creates systemic risks linked to contagion and cascading effects in the wake of cyber incidents. This results in highly-skewed distributions and the potential for 'black swan' phenomena.

Extreme value theory (EVT) has traditionally been used in catastrophe modelling of phenomena such as hurricanes, floods, etc. However, cyber risks are different to traditional catastrophic phenomena in a number of fundamental ways, which necessitates adaptation of analytical techniques such as EVT when modelling cyber risks.

This presentation will explain how cyber risks differ from traditional catastrophe risks and the implications this has when modelling these risks. It will use examples related to severity and cost of data breaches to illustrate these points with the help of R packages such as evd, extRemes and ismev.

SFCR automated analysis using scraping, text mining and machine learning methods for benchmarking and capital modelling

Aurelien Couloumy, Reacfin

The aim of this paper is to propose a new process developed in R to manage easily both unstructured textual data and quantitative data for insurance purposes. The approach is threefold:

- Using scraping methods to gather content automatically;
- Using text mining and natural language processing to analyze documents;
- Exploiting data using both supervised and unsupervised machine leaning methods.

To show the benefits of this process, we propose to analyze Solvency & Financial Condition Reports (SFCR) that have been published since 18th May 2017 by insurers under Solvency II regulation. We will deduce and analyze many information from this work such as:

- Descriptive statistics about insurers;
- Unconventional insurers clustering;
- Machine learning models to predict key risk drivers.

Based on this application, we will also make some recommendation about data visualization methods. We will finally suggest other applicable uses of this process (such as insurance general conditions or reinsurance treaties management).

Machine learning and fairness for commercial insurance

Oliver Laslett, Cytora

With the growing speed and scale of data generation, insurers are increasingly deploying advanced data analytics and machine learning techniques to assess and price risks. The application of these techniques promises to transform external data into insight for commercial underwriting. But can machine learning deliver fair underwriting results? The answer crucially depends on whether or not fairness is included in the design and analysis of the machine learning algorithms.

In this talk, we present methods for evaluating the fairness of decisions made by automated algorithms using a number of frequentist and Bayesian approaches. We demonstrate how fairness can be included in the design of machine learning algorithms, by controlling for protected features. Finally, we highlight how applying transparent modelling approaches to

external data can deliver fair underwriting in commercial insurance. By implementing these methods, insurers can price risk more accurately, while delivering fairer prices to customers.

Getting value out of machine learning

Javier Rodriguez Zaurin, Simply Business

During the last 4 years Simply Business has undergone an tech transformation, gradually decommissioning legacy systems and introducing a series of new technologies across the company. This was part of a long-term strategy designed to change the ways data were collected and used in the company. As a result, and just to cite some examples, the Customer Insights teams runs analytics on Jupyter notebooks, the Pricing team runs and validates their model using Databricks and pySpark, and the entire data platform relies on cloud-based (AWS) tools streaming constantly into storage systems like S3 or Redshift. It is now the time when the company is ready to efficiently utilise machine learning. The challenge now is, however, as complex if not harder than the previously described technology transformation. The challenge now is "human". During this presentation I will illustrate with some precise examples the difficulties of productionising machine learning algorithms even when all the technology requirements are in place. I will also illustrate why there is no point in investing in data science/machine learning if the mentioned tech-transformation has not happened beforehand.

Making decisions under uncertainty using Bayesian inference and Stan

Eric Novik, Generable

In this talk, I will introduce the key components of the Bayesian workflow including gathering prior knowledge, formulating a generative model, simulating data, fitting the model, evaluating the quality of the model, and making decisions conditional on model inferences.

I will talk about how Stan fits into this workflow and discuss its strengths and weaknesses. I will use a few examples from product pricing and clinical research to highlight the applicability of the tools and methods to difficult applied problems.

Presenter Profiles

Gareth Peters, Heriot-Watt University

Gareth Peters

Chair Professor for Statistics in Risk and Insurance Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh.

Dr. Gareth W. Peters is the 'Chair Professor for Risk and Insurance' in the Department of Actuarial Mathematics and Statistics, in Heriot-Watt University in Edinburgh. Previously he held tenured positions in the Department of Statistical Sciences, University College London, UK and the Department of Mathematics and Statistics in University of New South Wales, Sydney, Australia.

Dr. Peters is the incoming Director of the Scottish Financial Risk Association.

He has published in excess of 150 peer reviewed articles on risk and insurance modelling, 2 research text books on Operational Risk and Insurance as well as being the editor and contributor to 3 edited text books on spatial statistics and Monte Carlo methods.

He currently holds positions as:

- Honorary Prof. of Statistics at University College London, 2018+
- Affiliated Prof. of Statistics in University of New South Wales Australia 2015+
- Affiliate Member of Systemic Risk Center, London School of Economics 2014+
- Affiliate Member of Oxford Mann Institute, Oxford University (OMI) 2013+

He previously held positions as:

- Honorary Prof. of Peking University, Beijing, China 2014-2016
- Adjunct Scientist in the Mathematics, Informatics and Statistics, Commonwealth Scientific and - Industrial Research Organisation (CSIRO) 2009-2017

Tom Reynkens, KU Leuven

Tom Reynkens

Tom Reynkens is currently a postdoctoral researcher in actuarial science at KU Leuven, Belgium, and holds a PhD in Mathematics on the subject of extreme value theory from the same university. His research interests are statistical methods for non-life insurance pricing and loss modelling, extreme value theory and fraud detection. He is the main author of the ReIns package which complements the book "Reinsurance: actuarial and statistical aspects". Together with the authors of this book, Tom gave workshops for actuaries in Switzerland, Poland and South Africa where he was responsible for the practical sessions with R.

Christian Rohrbeck, Lancaster University

Christian Rohrbeck

Christian Rohrbeck is an AXA Post-Doctoral Research Fellow at Lancaster University. His research aims to develop new methodology in spatial statistics, nonparametric regression and extreme value theory in order to analyze the association between insurance claims and weather events. He was awarded an AXA Research Fund postdoctoral grant in 2017 and the EPSRC Doctoral Prize of the Faculty of Science and Technology at Lancaster University for 2016/2017. Christian holds both a PhD and MRes in Statistics and Operational Research from Lancaster University and a Diplom in Mathematics from Heidelberg University.

Mario Wüthrich, ETH Zurich

Mario Wüthrich

Mario Wüthrich is Professor in the Department of Mathematics at ETH Zurich, Honorary Visiting Professor at City, University of London, and Honorary Professor at University College London. He holds a Ph.D. in Mathematics from ETH Zurich. From 2000 to 2005, he held an actuarial position at Winterthur Insurance, Switzerland. He is fully qualified actuary SAA, serves on the board of the Swiss Association of Actuaries, and is Editor-in-Chief of ASTIN Bulletin.

Steven Perkins

Steven Perkins

Steve is a qualified actuary working in UK motor insurance who also holds a PhD in statistics. His current role involves elements of traditional actuarial work as well as driving the understanding and utilisation of data science techniques across the whole insurance group. This involves developing bespoke data science solutions for a wide range of business tasks as well as for use in more standard actuarial modelling.

Valerie du Preez, Dupro
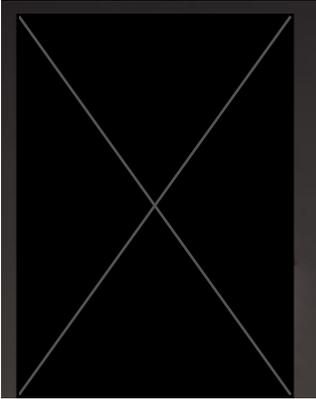
Valerie du Preez

Valerie is a qualified actuary working within the life insurance industry. She is a member of the MAID working party from the Institute of Actuaries. Within the insurance industry she has focussed her efforts on leveraging various tools and disciplines to effect change. As an actuary, she is focused on developing sustainable finance and risk operating strategies and optimising key business drivers impacting organisational performance and financial accuracy, with the use of innovative technologies, such as machine learning. She has extensive experience implementing actuarial systems; implementing and streamlining workflow processes and strengthening the financial control and financial close processes.

Alberto Glionna, Generali

Alberto Glionna

Alberto has earned a MSc in actuarial science from the University of Trieste while he was working in the UK as an actuarial analyst. During his 4 years at Assicurazioni Generali in both London and Trieste, Alberto focused on loss modeling for non-life insurance and reinsurance, and lately also for operational risk.

Lara A. Neira, Inrobin, University of Edinburgh

Lara A. Neira

Dr. Lara A. Neira Gonzalez co-founded Inrobin while in her last year of PhD in Machine Learning applied to Psychiatric Genetics at the University of Edinburgh (last two years) and Trinity College Dublin (first two years). Currently, Lara is the CTO and Data Scientist in Inrobin. She has a track record in quantitative finances having worked for Banco Pastor and Trinity College Dublin. She has held numerous advanced developer positions and masters several programming languages, driving Inrobin's technology. She holds two Master degrees: one in Quantitative Finance at IFFE Business School and another one in Statistical techniques at Universidad de Santiago de Compostela (Spain). She got an Honours Bachelor's Degree in Mathematics at the Universidad de Santiago de Compostela (5 years programme) with specialization in Statistics and Operational Research.

Aniketh Pittea, University of Kent

Aniketh Pittea

I am currently doing a PhD in Actuarial Science at the University of Kent. The aim of my research is to carry a risk assessment of pension schemes in the UK, US and Canada as the baby boomers retire and the population demographics shift in those countries. Before starting my PhD, I was working as Risk Analyst in one of the leading insurance companies in Mauritius. I hold a BSc (Hons) Actuarial Science from the University of Technology, Mauritius and an MSc in Applied

Actuarial Science from the University of Kent. In addition, I have completed the professional examinations of the Institute and Faculty of Actuaries.

Marc Rierola, Qatar Re

Marc Rierola

At Qatar Re I am working in the Risk Management team located in Zurich where I focus on

- Cat related financial reporting
- Accumulation control
- Large Loss Event reporting
- Portfolio Roll-up
- RShiny development

I have 10+ years in the reinsurance industry where I worked in various roles such as catastrophe modeler (using all major vendor models), model developer (using R for the first time) and Risk Management. At Qatar Re I introduced R and RShiny to the company and we are currently building our Lossfile database user interface in RShiny. In 2005 I received my degree as Geophysicist and Seismologist from ETH Zurich.

Pauli Rämö, Miari Solutions

Pauli Rämö

Pauli holds a master degree in Computer Science and a PhD in Systems Biology. During his academic experience in the domain of Systems Biology, he focused on dynamical modeling of gene regulatory networks and on large-scale analysis of microscope images. After working at ETH Zurich and Biozentrum Basel, Pauli joined Mirai Solutions in 2014. Since then he has been involved in several consulting projects for the Insurance industry, where he has designed and implemented technical solutions for areas such as financial risk modeling, Solvency II Standard Model, and technical reporting. Besides working on challenging coding tasks, Pauli enjoys the mountain life and is an avid gardener.

Senren Tan, Cass Business School

Senren Tan

Senren is a 3rd year PhD candidate in actuarial science at Cass Business School. He has a Bachelor of Commerce (Honours) and a Master of Commerce (Actuarial Science) in the University of Melbourne. His research interest is on boundary-crossing probabilities for stochastic processes and their applications in insurance and finance. These probabilities are not easy to compute, but have numerous applications in option pricing, operations research, statistics, insurance and many other contexts. The aim of his research is to develop methods or algorithms that are both theoretically justified and computationally efficient in calculating the boundary-crossing probabilities.

Mike Ludkovski, University of Santa Barbara

Mike Ludkovski

Mike Ludkovski is a Professor of Statistics and Applied Probability at University of California Santa Barbara where he co-directs the Center for Financial Mathematics and Actuarial Research. Among his research interests are Monte Carlo techniques for optimal stopping/stochastic control, applications of machine learning in longevity and non-life insurance and non-zero-sum stochastic games. During 2016-17 he chaired the SIAM Activity Group on Financial Mathematics and Engineering.

Silvana Pesenti, Cass Business School

Silvana Pesenti

Silvana Pesenti is a 3rd year PhD student in actuarial science at Cass Business School. She has a Master of Science in Mathematics from ETH Zurich with distinction and worked prior to commencing her PhD for 6 months at the life insurance company Prudential, plc, in London. Her research interest lies in sensitivity analysis of risk management for insurance companies and financial institutions, in particular its application to capital requirements, solvency and regulation.

Benjamin C. Dean, Iconoclast Tech LLC

Benjamin C. Dean

Benjamin C. Dean works at the intersection of technology, economics and public policy. Through Iconoclast Tech LLC, Mr. Dean develops cyber risk models for insurers and reinsurers. Mr. Dean also contributes to an initiative to develop business digital risk management metrics with the Organisation for Economic Co-operation and Development's (OECD's) Working Party on Security and Privacy in the Digital Economy.

Previously, Mr. Dean contributed a paper to inform the European Parliament on the economic implications of EU-US cooperation in cybersecurity and cybercrime. During 2017 he was a Fellow at the Center for Democracy and Technology in Washington DC.

From 2014-16, Mr. Dean was a fellow for cybersecurity at Columbia University and a policy analyst at the OECD. Mr. Dean completed a MA International Affairs at Columbia University's School of International and Public Affairs. He is also a graduate of the University of Sydney with a BA Economics and Social Sciences (Hons.)

Aurélien Couloumy, Reacfin

Aurélien Couloumy

Aurélien Couloumy is currently associate lecturer at ISFA, Université Lyon 1. He is also responsible of the Data Science department in the consulting firm Reacfin. Previously, he

worked as Head of Models in an international actuarial consulting firm. He has acquired a strong experience in insurance modelling, insurance and financial regulations, machine learning technics, text mining, natural language processing and scraping methodologies. Aurélien is also fellow of the French Institute of Actuaries and the Institute of Actuaries in Belgium (IA|BE).

Oliver Laslett, Cytora

Oliver Laslett

Oliver Laslett is a Machine Learning Scientist who works with insurers to develop novel risk modelling techniques. Oliver has supported risk modelling in a range of industries from aerospace to medical. He is shortly to complete his PhD in Computational Physics from the University of Southampton.

Javier Rodriguez Zaurin, Simply Business

Javier Rodriguez Zaurin

I completed my PhD in astrophysics at the University of Sheffield, UK, focusing on studying galaxy formation and evolution. After my PhD I did a number of post doctoral position until I jumped to the private sector early in 2014. Currently, I head the data science team at Simply Business, where we use ML to address a wide variety of data-related problems, from marketing channel attribution models to risk modeling. Previously I have worked in the fashion and retail industries mainly developing recommendation algorithms. In addition I am an advisor at alea.ai, an early-stage start-up focused on the insurance industry. Finally, I am still involved in academic

research as a member of the DAWN project (Deep Learning to Analyze Webb-detected Nascent-galaxies), where we use deep learning to understand how galaxies formed in the early universe.

Eric Novik, Generable

Eric Novik

Eric Novik is the CEO of Generable leading the team that is developing tools for precision medicine and leading-edge clinical research. Prior to Generable, Eric was a Data Scientist at TIBCO Spotfire where he built statistical machine learning applications for customers in pharma, financial services, energy, and CPG. In 2010, Eric started Risktail, an options analytics software company for retail traders. During the financial crisis of 2007-2009, he traded listed options and later spent two years in the Equity Derivatives Technology group at Barclays Capital. He has an MA in Statistics from Columbia University.

Paul Bürkner, University of Münster

Paul Bürkner

Paul Bürkner has studied psychology and mathematics and is currently doing his postdoc at the chair of statistics and methods at the department of psychology in Münster. Since his early years as a student, he is enthusiastic about Bayesian statistics and a lot of his work is dedicated to bringing Bayesian methods into scientific practice. As a member of the Stan development

team with focus on the R language, he works on better software for fitting, evaluating and visualizing Bayesian multilevel models.

Jake Morris, ArgoGlobal

Jake Morris

Jake is an Analytics Actuary at ArgoGlobal, where he employs a range of classical, Bayesian and machine learning techniques to help colleagues get the most from their data. He is a Fellow of the Institute and Faculty of Actuaries

Mick Cooney, Agrippa Data Consulting

Mick Cooney

Mick Cooney has a background in quantitative finance and has been working in the insurance industry for four years as a consultant on statistical modelling, data science. He has a particular interest in the use of Bayesian modelling and Monte Carlo simulation and its applications within the industry.