

Stu Hunter Research Conference 2019

February 17 – 20, 2019, Villa Porro Pirelli, Induno Olona, Italy

Plenary talks



Day 1: February 18 – Monday

1) Willis Jensen, W.L. Gore & Associates, Inc. “Statistics=analytics?”

Discussants: Bonnie Ray, Talkspace; Nathaniel Stevens, University of Waterloo

Moderator: Stefan Steiner, University of Waterloo

Abstract

In recent years, analytics has received a lot of buzz and attention in the business world. Many conferences and publications have appeared covering the topics of data and analytics and how to unlock the insights contained in the data. Statisticians would argue that many of the tools and concepts are not really all that new. We compare and contrast some of the various definitions associated with statistics and analytics. We discuss our perspectives on analytics and share an example of project which highlights some of the statistical and non-statistical tools required to be successful in analytics projects. We also discuss the connection of statistics and analytics to other terms and disciplines such as big data, data science, industrial statistics, statistical thinking, statistical engineering and Six Sigma. We provide our answer to the question of whether or not statistics and analytics are the same thing. The connections we highlight have some implications and create some opportunities and challenges for the statistics profession.

2) Jeff Wu, Georgia Institute of Technology “Analysis-of-Marginal-Tail-Means (ATM) – a new method for robust parameter optimization”

Discussants: Brad Jones, JMP; Murat Kulahci, Technical University of Denmark

Moderator: David Edwards, Virginia Commonwealth University

Abstract:

We present a new method, called Analysis-of-marginal-Tail-Means (ATM), for effective robust optimization of discrete black-box problems. ATM has important applications to many real-world engineering problems (e.g., manufacturing optimization, product design, molecular engineering), where the objective to optimize is black-box and expensive, and the design space is inherently discrete. One weakness of existing methods is that they are not robust: these methods perform well under certain assumptions, but yield poor results when such assumptions (which are difficult to verify in black-box problems) are violated. ATM addresses this via the use of marginal tail means for optimization, which combines both rank-based and model-based methods. The trade-off between rank- and model-based optimization is tuned by first identifying important main effects and interactions, then finding a good compromise which best exploits additive structure. By adaptively tuning this trade-off from data, ATM provides improved robust optimization over

existing methods, particularly in problems with (i) a large number of factors, (ii) unordered factors, or (iii) experimental noise. We demonstrate the effectiveness of ATM in simulations and in two real-world engineering problems: the first on robust parameter design of a circular piston, and the second on product family design of a thermistor network.

Day 2: February 19 – Tuesday

1) Paolo Giudici, Università di Pavia

“Scoring Models for P2P Business Lending Platforms – A Network Approach”

Discussants: Agus Sudjianto, Wells Fargo; Michele Tumminello, Università di Palermo

Moderator: Christian Ritter, Ritter & Danielson Consulting

Abstract:

Peer-to-Peer lending platforms may lead to cost reduction, and to an improved user experience. However, these improvements may come at the price of inaccurate credit risk measurements, which can hamper lenders and endanger the stability of a financial system. In the paper, we propose how to improve credit risk accuracy of peer to peer platforms and, specifically, of those who lend to small and medium enterprises, by leveraging borrowers’ networking data. To achieve this goal, we propose to augment traditional credit scoring methods with centrality measures derived from correlation networks among borrowers, deduced from the comovement of their financial variables. Our findings show that the inclusion of network centralities does improve the predictive accuracy of credit scoring models.

2) Neil Lawrence, University of Sheffield – Amazon UK

“Data Science and Digital Systems – The 3Ds of Machine Learning System Design”

Discussants: Jean Michel Poggi, Université Paris-Sud; David Banks, Duke University

Moderator: Frederick W. Faltin – Virginia Tech

Abstract:

Machine learning solutions, in particular those based on deep learning methods, form an underpinning of the current revolution in “artificial intelligence” that has dominated popular press headlines and is having a significant influence on the wider tech agenda. Here we give an overview of the 3Ds of ML systems design: Data, Design and Deployment. By considering the 3Ds we can move towards *data first* design.

Day 3: February 20 – Wednesday

1) Enrique Del Castillo, Pennsylvania State University **“Industrial Statistics and Manifold Data”**

Discussants: Marco Reis, Universidade de Coimbra; Changliang Zou, Nankai University

Moderator: Xavier Tort-Martorell, Universitat Politècnica de Catalunya, Barcelona Tech

Abstract:

Complex and not only big data are everywhere in industry and how to control and optimize systems based on these data types is an important aspect of modern Quality Engineering. We present a new approach for statistical process monitoring of point cloud, mesh and voxel data based on intrinsic geometrical features of the 2-D manifold (surfaces) of scanned manufactured parts. Monitoring intrinsic properties avoids computationally expensive registration pre-processing of the data sets. Our proposal brings SPC closer to computer vision and computer graphics methods. We also present a review of recent approaches for analyzing and designing experiments where either the response or the covariates lie on manifolds. For SPC we offer a proof of concept rather than a finished technology; for DOE we will review recent work.

2) Laura Sangalli, Politecnico di Milano **“Functional and complex data – new methods merging statistics, scientific computing and engineering”**

Discussants: Fugee Tsung, Hong Kong University of Science and Technology; Ernst Wit, Università della Svizzera Italiana

Moderator: Peter Parker, NASA

Abstract:

Recent years have seen an explosive growth in the recording of increasingly complex and high-dimensional data. Classical statistical methods are often unfit to handle such data, whose analysis calls for the definition of new methods merging ideas and approaches from statistics, applied mathematics and engineering. I will in particular focus on data displaying complex spatial or spatio/temporal dependencies, where the complexity can for instance be due to the complex physics of the problem, the presence of external forces driving the phenomenon under study, the non-trivial conformation of the domain where the data are observed. In this context, the interplay among statistics, applied mathematics and engineering can be particularly effective. I will specifically present one class of models, at the interface between statistics and numerical analysis that can efficiently handle these complex data structures. The ideas and methods will be illustrated through a few case studies.