THE UNIVERSITY *of* EDINBURGH
Centre for Statistics

Centre for Statistics Annual Conference 2023

Thursday, 15th June 2023

The Nucleus, Larch Lecture Theatre
University of Edinburgh, King's Buildings, EH9 3FD

https://centreforstatistics.maths.ed.ac.uk

Organized by Natalia Bochkina, Timothy Cannings, Ozan Evkaya, and Maarya Sharif.

The **Centre for Statistics** unites data-driven researchers from across the University of Edinburgh and Associated Institutions, building the capacity needed to address key scientific and societal challenges. The CfS fosters interdisciplinary research and knowledge exchange through cross-disciplinary seminars; small grants to develop collaborative research projects; extended research visits of world leading statisticians in academic units across Scotland; establishing connections between academics, and with industrial, governmental and commercial users; and training the next generation of statisticians and data scientists.

The **Centre for Statistics Annual Conference** showcases the most innovative interdisciplinary research and applications and brings together researchers from across the University of Edinburgh and Associated Institutions. The one day conference features a number of **invited speakers**, a **new members session**, a **poster presentations** and plenty of time for **networking**.

PROGRAMME (all talks will take place in the Larch Lecture Theatre):

- 09:15-09:30 Registration and arrival.

- 09:30-09:40 **Welcome**: Natalia Bochkina (Director of Centre for Statistics).

- 09:40-10:10 **Extending Entropy Analysis Techniques to Irregularly Sampled Data: From Multivariate Time Series to Graphs Data and Beyond.** Javier Escudero (School of Engineering).

- 10:15-10:45 **AI in Health: a Computer Vision perspective.** Miguel O. Bernabeu (Usher Institute).

- 10:45-11:15 Coffee break.

- 11:15-11:45 **A distributed block-coordinate Gibbs sampler for image recovery.** Audrey Repetti (Heriot-Watt University).

- 11:50-12:20 **Financial data, just for finance?** Mike Spencer (Smart Data Foundry).

- 12:25-12:40 **New Members Session**: Short talks from new CfS members

- 12:40-14:00 Lunch

- 14:00-14:10 **New Members Session ctd.**: Short talks from new CfS members

- 14:10-14:40 **Dispensing with unnecessary assumptions in population genetics analysis.** Ava Khamseh (School of Informatics and IGC).

- 14:45-15:15 **What, Why, How: Data visualization for Exploring and Communicating Data.** Benjamin Bach (School of Informatics).

- 15:15-15:35 Break and Group Photo.

- 15:35-16:05 **Social inequalities in occupational attainment: using sibling data to estimate the total effect of family of origin and the role of education.** Cristina Iannelli (Moray House School of Education and Sport).

- 16:10-16:20 Closing remarks.

https://centreforstatistics.maths.ed.ac.uk/cfs/events/upcoming-events/cfsc-2023

Centre for Statistics

THE UNIVERSITY *of* EDINBURGH

Edinburgh, UK

https://centreforstatistics.maths.ed.ac.uk

Follow us on Twitter at https://twitter.com/Ed_CfS

CfS Annual Conference 2023: Invited Talks

Invited talks feature speakers from the **University of Edinburgh** and **Heriot Watt University**.

Extending entropy analysis techniques to irregularly sampled data: from multivariate time series to graphs data and beyond

Javier Escudero (Javier.Escudero@ed.ac.uk)

School of Engineering, University of Edinburgh

Abstract: Nonlinear analysis metrics based on entropy and information theory have found an increasing interest in the analysis of dynamical systems. They have recently extended from univariate to multivariate signals and images, and their use to characterise complexity in a variety of systems is now widespread. In this talk, I will present very recent advances extending definitions of nonlinear analysis metrics for graph signals, that is, data sampled on irregular networks. I will show that our graph-based algorithms does not only generalise previously available uni-variate methods, but also allow new applications relevant to a wide variety of fields.   †**Time slot**: 09:40-10:10

AI in Health: a Computer Vision perspective

Miguel O. Bernabeu (Miguel.Bernabeu@ed.ac.uk)

Usher Institute, University of Edinburgh

Abstract: Recent advances in Computer Vision and Machine Learning promise to revolutionise medical image interpretation. In the words of Prof. Geoff Hinton, a Turing Awardee in 2018, it is "quite obvious that we should stop training radiologists", radiologists are "the coyote already over the edge of the cliff who hasn't yet looked down". Unsurprisingly, radiologists, amongst others, have strongly disputed such statements and current data does not support any reduction in radiology job openings. In this talk, I will argue that while Computer Vision has taken a strong foothold in the area of medical image interpretation, successful translation of recent advances to clinical practice require a sophisticated interplay of multiple disciplines. Considerations around data quality, implementation, and evaluation are as relevant as model development, typically the area making the headlines. To illustrate challenges and approaches, I will present work undertaken at The University of Edinburgh around these three domains involving large multi-disciplinary teams.   †**Time slot**: 10:15-10:45

A distributed block-coordinate Gibbs sampler for image recovery

Audrey Repetti (A.Repetti@hw.ac.uk)

School of Mathematical and Computer Sciences and the school of Engineering and Physical Sciences, Heriot-Watt University

Abstract: Sampling-based algorithms are classical approaches to perform Bayesian inference in inverse problems. They provide estimators with associated credibility intervals, for uncertainty quantifi- cation. A main drawback of these methods is that they usually hardly scale to high dimensional problems. This issue has been partly addressed recently, by pairing them with opti- misation tech- niques, such as proximal and splitting approaches. Such approaches pave the way to distributed samplers, that can split computation costs for faster and more scalable inference. We propose a distributed Gibbs sampler to efficiently solve such problems, considering posterior distributions with multiple smooth and non-smooth functions composed with linear operators. The proposed approach leverages a recent approximate augmentation technique reminiscent of primal-dual op- timisation methods. It is further combined with a block-coordinate approach to split the primal and dual variables into blocks, leading to a distributed block-coordinate Gibbs sampler. The re- sulting algorithm exploits the hypergraph structure of the involved linear operators to efficiently distribute the variables over multiple workers. Experiments on an image deblurring problem show the performance of the proposed approach in high dimension, to produce high quality estimates with credibility intervals, while reducing drastically the computation cost compared to its non- distributed version. Joint work with P.A. Thouvenin and P. Chainais.

†**Time slot**: 11:15-11:45

Financial data, just for finance?

Mike Spencer (michael.spencer@smartdatafoundry.com)

Smart Data Foundry

Abstract: The financial services sector have a huge presence in Edinburgh, from banks to fintechs. The data they hold has huge value to businesses and individuals, but what use is it for research and policy making? In this talk Mike will discuss how Smart Data Foundry (https://smartdatafoundry.com) works with retail banks, accounting firms and fintechs to make their financial data available for the public good. He will illustrate how accademics can work safely with these data and provide policy impact. Financial data are a newly available resource which tell us much about the way society functions. Come and hear how Smart Data Foundry link and use these data and how you can too.

†**Time slot**: 11:50-12:20

New Members Session

Short Talks from new Centre for Statistics members.

Speakers include

Nicole Augustin (School of Mathematics): *Research overview: Spatio-temporal modelling of environmental data, modelling of complex health data and model selection uncertainty*

Ozan Evkaya (School of Mathematics): *Dependence everywhere: Copulas and its applications*

Man Ho Suen (School of Mathematics): *Poster preview*

David Sterratt (School of Informatics): *Visualisation principles and guidance for a data science course*

More to be confirmed

Five minute talks either side of lunch   †**Time slot**: 12:20-14:10

Dispensing with unnecessary assumptions in population genetics analysis

Ava Khamseh (ava.khamseh@ed.ac.uk)

School of Informatics & Institute for Genetics and Cancer, University of Edinburgh

Abstract: Parametric assumptions in population genetics analysis – including linearity, sources of population stratification and additivity of variance as part of a Gaussian noise – are often made, yet their (approximate) validity depends on variant and traits of interest, as well as genetic ancestry and population dependence structure of the sample cohort. We present a unified statistical workflow, called TarGene, for targeted estimation of effect sizes, as well as two-point and higher-order epistatic interactions of genomic variants on polygenic traits, which dispenses with these unnecessary assumptions. Our approach is founded on Targeted Learning, a framework for estimation that integrates mathematical statistics, machine learning and causal inference. TarGene maximises power whilst simultaneously maximising control over false discoveries by: (i) guaranteeing optimal bias-variance trade-off, (ii) taking into account potential covariate non-linearities, sources of population stratification and dependence structure, and (iii) detecting genetic non-linearities. The necessity of this model-independent approach is demonstrated via simulations, with application to the UK Biobank genotype-phenotype data.   †**Time slot**: 14:10-14:40

What, Why, How: Data visualization for Exploring and Communicating Data

Benjamin Bach (bbach@exseed.ed.ac.uk)

VisHub, School of Informatics, University of Edinburgh

Abstract: This talk gives a broad overview into applications, techniques, and tools for data visualization, aimed to inspire uptake and application. I discuss why data visualization works and how it complements purely quantitive analyses. I discuss what problems it can help solve, e.g., around complex relational data (i.e., networks), temporal data or the communication of scientific insights. Eventually, I will discuss what tools exist and how to decide which tools to use. I am co-leading the VisHub research group at the School of Informatics: https://vishub.net.   †**Time slot**: 14:45-15:15

Social inequalities in occupational attainment: using sibling data to estimate the total effect of family of origin and the role of education

Cristina Iannelli (C.Iannelli@ed.ac.uk)

Moray House School of Education and Sport, University of Edinburgh

Abstract: This study provides new evidence about the extent to which individual occupational status is determined by family of origin (ascription) and by educational attainment (achievement). Existing social mobility research has examined the degree of intergenerational transmission of advantage and disadvantage by measuring the associations between parent and child social class, education and income. However, this approach is unable to capture the full extent of the family-of-origin effect because available data often contain limited information about family background. In our study, we overcome this issue by using sibling data from the Scottish Longitudinal Study, containing linked administrative data from three Censuses (1991, 2001 and 2011). We use random effects and fixed effects modelling to estimate the importance of family background for occupational outcomes in Scotland and we assess the effect of education on these outcomes by examining between- and within-sibling differences.

†**Time slot**: 15:35-16:05