27-30 JUNE 2022

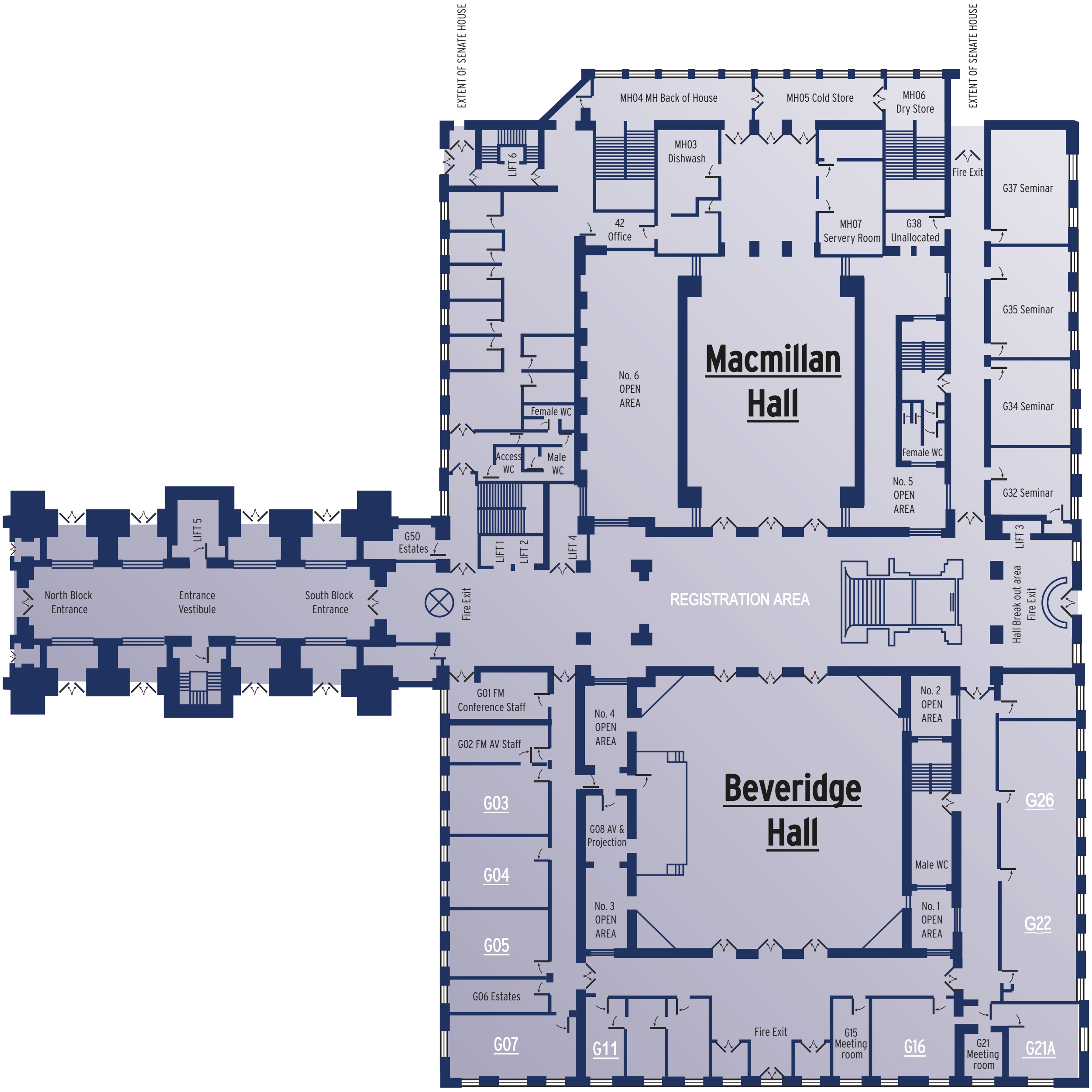
London, UK

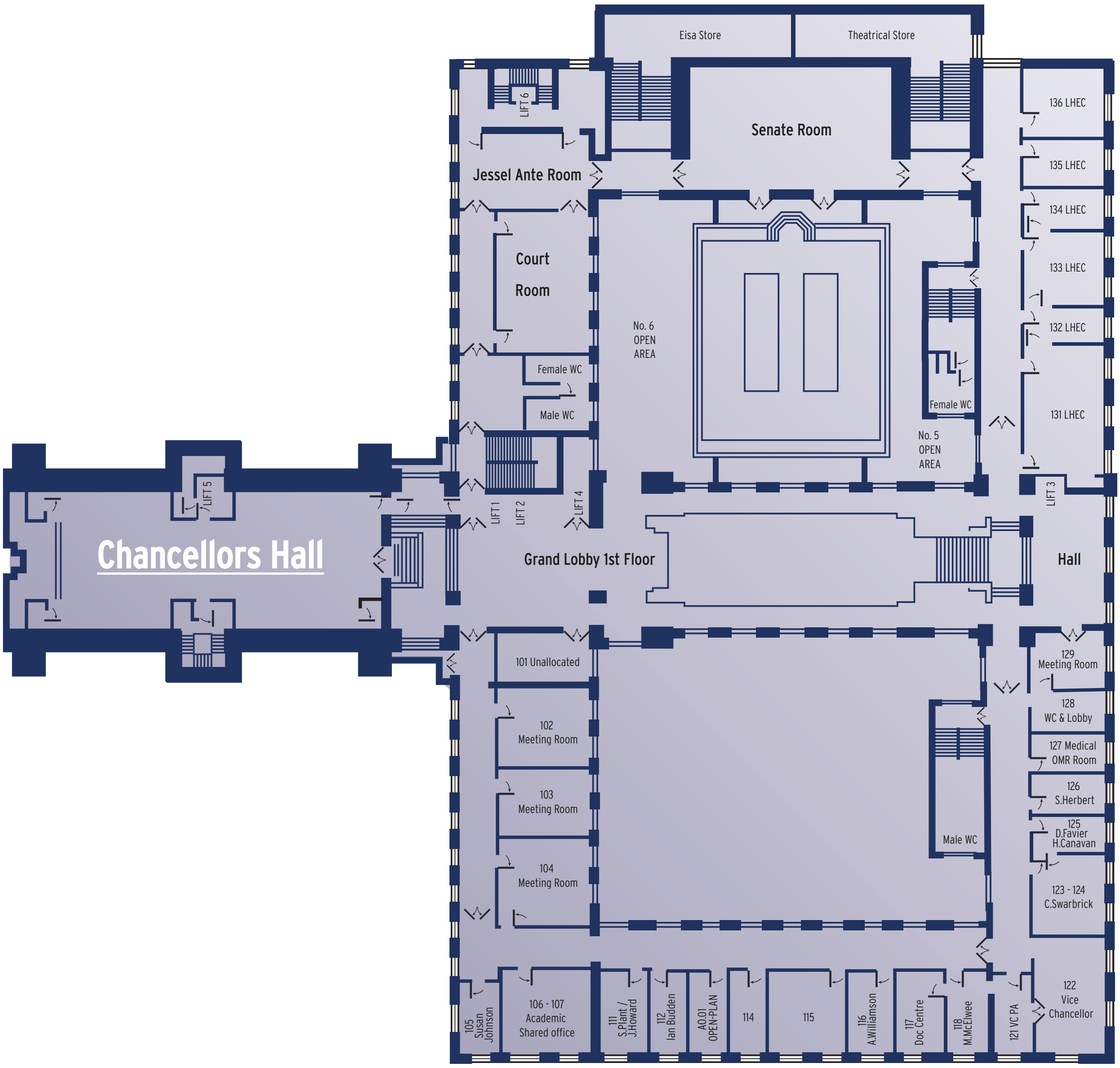
2022 IMS Annual Meeting

Probability and Statistics

Senate House, University of London, Malet St, WC1E 7HU

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE


EXTENT OF SENATE HOUSE
EXTENT OF SENATE HOUSE
MH04 MH Back of House
MH05 Cold Store
MH06 Dry Store
MH03 Dishwash
LIFT 6
Fire Exit
G37 Seminar
42 Office
MH07 Servery Room
G38 Unallocated
G35 Seminar
Macmillan Hall
No. 6 OPEN AREA
G34 Seminar
Female WC
Female WC
Access WC
Male WC
No. 5 OPEN AREA
G32 Seminar
LIFT 5
G50 Estates
LIFT 1
LIFT 2
LIFT 4
LIFT 3
North Block Entrance
Entrance Vestibule
South Block Entrance
Fire Exit
REGISTRATION AREA
Hall Break out area
Fire Exit
G01 FM Conference Staff
No. 4 OPEN AREA
No. 2 OPEN AREA
G02 FM AV Staff
Beveridge Hall
G03
G26
G08 AV & Projection
G04
Male WC
G22
No. 3 OPEN AREA
No. 1 OPEN AREA
G05
G06 Estates
Fire Exit
G15 Meeting room
G07
G11
G16
G21 Meeting room
G21A


Eisa Store
Theatrical Store
Senate Room
Jessel Ante Room
Court Room
Female WC
Male WC
No. 6 OPEN AREA
No. 5 OPEN AREA
LIFT 6
LIFT 5
LIFT 1
LIFT 2
LIFT 4
LIFT 3
Chancellors Hall
Grand Lobby 1st Floor
Hall
136 LHEC
135 LHEC
134 LHEC
133 LHEC
132 LHEC
131 LHEC
101 Unallocated
102 Meeting Room
103 Meeting Room
104 Meeting Room
Male WC
129 Meeting Room
128 WC & Lobby
127 Medical OMR Room
126 S.Herbert
125 D.Favier H.Canavan
123 - 124 C.Swarbrick
122 Vice Chancellor
105 Susan Johnson
106 - 107 Academic Shared office
111 S.Plant / J.Howard
112 Ian Budden
A0.01 OPEN-PLAN
114
115
116 A.Williamson
117 Doc Centre
118 M.McElwee
121 VC PA

General Conference Information

IMS Annual Meeting 2022 Organising Committee

Scientific Committee Chairs

- Thomas Mikosch (University of Copenhagen, Denmark)
- Qiwei Yao (London School of Economics and Political Science, UK)

Scientific Committee

- Rina Foygel Barber (University of Chicago, USA)
- Pauline Barrieu (London School of Economics and Political Science, UK)
- Pierre Bellec (Rutgers University, USA)
- Jean Bertoin (University of Zurich, Switzerland)
- Jose Blanchet (Stanford University, USA)
- Elisabetta Candellero (Rome Tre University, Italy)
- Ying Chen (National University of Singapore, Singapore)
- Jian Ding (University of Pennsylvania, USA)
- Yingying Fan (University of Southern California, USA)
- Paul Fearnhead (Lancaster University, UK)
- Piotr Fryzlewicz (London School of Economics and Political Science, UK)
- Yuta Koike (University of Tokyo, Japan)
- Axel Munk (University of Goettingen, Germany)
- Gennady Samorodnitsky (Cornell University, USA)
- Marta Sanz-Sole (University of Barcelona, Spain)
- Rajen Shah (University of Cambridge, UK)
- Stilian Stoev (University of Michigan, USA)
- Botond Szabo (Vrije Universiteit Amsterdam, Netherlands)
- Maria Eulalia Vares (Federal University of Rio de Janeiro, Brazil)
- Rongchan Zhu (Beijing Institute of Technology, China)

Local Organising Committee

Local Organising Committee

(London School of Economics and Political Science)

Academic Members:

- Clifford Lam (co-chair)
- Erik Baurdoux
- Yining Chen
- Zezhun Chen
- Qin Fang
- Gabriel Wallin

Professional Staff Team:

- Imelda Noble-Andolfo (co-chair)
- Wasima Anwari
- Joey Hoang
- Penny Montague

Keynote Speakers

Michael Jordan

Inaugural Grace Wahba Lecture

Michael is the Pehong Chen Distinguished Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California, Berkeley.

Martin Hairer

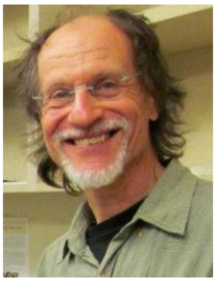
Wald Lecture

Martin is a Professor of Pure Mathematics at the Department of Mathematics of Imperial College London. His main areas of interest are probability theory and analysis, with a particular focus on the analysis of stochastic PDEs.

Heping Zhang

Neyman Lecture

Heping is a Professor of Statistics and Data Science at Yale School of Public Health. He has published over 300 research articles and monographs in theory and applications of statistical methods and biomedical research.



Hans-Georg Müller

Rietz Lecture

Hans-Georg is a Professor in the Department of Statistics at the University of California, Davis. His areas of interest include statistical methodology and modelling, nonparametric statistics, biostatistics, and data analysis.

Keynote Speakers

Rodrigo Bañuelos

Medallion Lecture

Rodrigo is a Professor of Mathematics at Purdue University. His research is at the interface of probability, harmonic analysis, and spectral theory; three areas that are rich with beautiful problems that are very difficult to solve.

Rina Foygel Barber

Medallion Lecture

Rina is a Professor in the Department of Statistics at the University of Chicago. Her research interests are in developing and analysing estimation, inference, and optimization tools for structured high-dimensional data problems.

Vlada Limic

Medallion Lecture

Vlada works at the University of Strasbourg. She conducts research in probability and stochastic processes, studying properties of various stochastic models.

Roman Vershynin

Medallion Lecture

Roman is a Professor of Mathematics at the University of California, Irvine and an Associate Director of the ACO Center (Algorithms, Combinatorics and Optimization). His research spans high-dimensional probability and mathematical data science, with many particular interests.

Keynote Speakers

Russell Lyons

IMS/BS Schramm Lecture

Russell is a Professor at Indiana University. His research interests include probability theory on graphs, geometric group theory, combinatorics, statistical mechanics, ergodic theory, and harmonic analysis.

Krzysztof Burdzy

IMS Presidential Address

Krzysztof (Chris) is currently serving as the President of the Institute of Mathematical Statistics until July 2022. He is also a Professor of Mathematics and Adjunct Professor of Statistics at the University of Washington.

Bernard Silverman

Conference Dinner Speech

Bernard is a Professor of Modern Slavery Statistics at the University of Nottingham. He served as Chief Scientific Adviser to the Home Office (2010-17), and Master of St Peter's College at the University of Oxford (2003-09).

What You Need to Know

Wi-Fi/Internet Access

Wi-Fi Name: UoLConference. No password required.

Disabled Access

There is ramped/sloped or stepped access at the walkway area between North and South Block entrance. Please note there is not a buzzer/intercom.

Food and Drink

Tea, coffee and refreshments will be provided during the morning and afternoon tea breaks at the **Grand Lobby 1st Floor** within Senate House. If you have any allergies, please notify a staff member.

There is also a café on site where you can purchase additional items. This is located on the Ground Floor of Senate House, South Block.

Senate House (Reception/Helpdesk)

The helpdesk opening hours are as follows:

Monday to Friday: 9am – 5.45pm

Helpdesk telephone: 020 7862 8500

Conference Registration Desk Opening Hours:

Monday 27 June: 7:30 -- 17:00

Tuesday 28 June: 8:00 -- 17:00

Wednesday 29 June: 8:00 -- 17:00

Thursday 30 June: 8:30 -- 16:00

Policies

Mobile phones

Please ensure your mobile phones are on silent or turned off during each session to avoid disruption.

Smoking

No smoking is allowed on the premises.

Photography/video

Please refrain from any flash photography to avoid disruption for the speakers and other attendees.

Reception and Conference Dinner

Welcome Reception – 27 June – Marshall Building, LSE

The London School of Economics and Political Science will be hosting a reception on 27 June of the 2022 IMS Annual Meeting. We welcome all our attendees to join us in the Great Hall, Marshall Building at LSE to network with other attendees over canapes and drinks.

Please ensure you have your invitation on the day as this is an invite-only event and our security staff will need to check you in.

Venue Address:

The Great Hall, Marshall Building, 44 Lincoln's Inn Fields, London, WC2A 2ES

Time:

6.30 – 9pm

Dress Code:

Casual

Dinner – Wednesday 29 June – Old Building, LSE

The London School of Economics and Political Science will be hosting a conference dinner on 29 June of the 2022 IMS Annual Meeting. This event is limited to attendees who have purchased a ticket only.

We will kick off with a reception in the Senior Common Room, followed by a sit-down dinner in the Staff Dining room (located next to the Senior Common Room).

Please ensure you arrive on time and have your dinner ticket with you on the day so our stewards can confirm your place.

Conference dinner speech will be delivered by Professor Sir Bernard Silverman.

Venue Address:

The Senior Common Room and The Staff Dining Room, 5th Floor, The Old Building, Houghton St, London WC2A 2AE.

Time:

Reception: 7pm

Dinner: 7.45pm

Dress Code:

Smart casual

Programme at a Glance

2022 IMS Annual Meeting Outline Timetable

Time	27 June	28 June	29 June	30 June
08:30 – 09:15	Wald Lecture I	Wald Lecture II	Neyman Lecture	
09:20 – 10:05	Rietz Lecture	Medallion Lectures (Banuelos, Barber)	Medallion Lectures (Limic, Vershynin)	
09:00 – 10:00				CTS16 – 22, CTS24
10:05 – 10:30	Tea break			
10:30 – 12:30	IP3, IP6, IP9, IP13 IS1, IS2, IS4 RSS Session CTS3	IP4, IP5, IP7 IS9, IS12, IS17, IS22, IS24 CTS4	IP14, IP16, IP19, IP20 IS5, IS6, IS21 CTS5, CTS7	IP22, IP23, IP26, IP29 IS8, IS10, IS13, IS25 CTS1
12:30 – 13:30	Lunch			
13:30 – 15:30	IP1, IP10 IS3, IS11, IS26, IS28 CTS2, CTS23, TC6	IP8, IP11, IP12 IS14, IS19, IS23, IS27, IS29 CTS6	IP15, IP17, IP18, IP21 IS16, IS18, IS20, IS30	IP24, IP25, IP27, IP28 IS7, IS15, IS31 Wahba Session TC17
15:30 – 16:00	Tea break			
16:00 – 17:30	TC1 – 5, TC7, TC8 Brown Session	TC9, TC10, TC12 – 16 New Researchers Session		
16:00 – 17:40			CTS8 – 15	
16:00 – 16:45				IMS/BS Schramm Lecture
16:50 – 17:35				Wahba Lecture
17:30 – 19:00		Award Presentation & Presidential Address		
18:30 – 20:00	Conference Reception			
19:00 – 21:00			Conference Dinner	

The duration of talks (including Q&A):

Named & Medallion lectures: 45 mins

Invited talks (IP or IS): 30 mins

Topic contributed session talks (TC): 22.5 mins

Contributed talks (CTS): 20 mins

Contents

Monday 27-06-2022	08:30-09:15	Room: Beveridge Hall	Chair: Thomas Mikosch	Wald Lecture I

Universality and Crossover in 1+1 Dimensions (Part I)
Speaker: **Martin Hairer, Imperial College London**
The focus of these two lectures will be on the large scale behaviour of 1+1-dimensional stochastic processes, namely on models in one space dimension that evolve in time. In this situation, there are two well-known renormalisation group fixed points, namely the Edwards-Wilkinson (EW) model and the KPZ fixed point, recently characterised by Matetski, Quastel and Remenik. It has long been conjectured that there is exactly one crossover process between the two, namely the KPZ equation. We will show some recent results supporting this conjecture, in particular a proof of the fact that the directed mean curvature flow in a noisy environment converges to the KPZ equation in a suitable regime.

Monday 27-06-2022	09:20-10:05	Room: Beveridge Hall	Chair: Rajen Shah	Rietz Lecture

Statistical Tools for Random Objects in Metric Spaces
Speaker: **Hans-Georg Müller, University of California, Davis**
Random objects, i.e., random variables that take values in a separable metric space, pose many challenges for statistical analysis, since vector operations are not available in general metric spaces. Examples of random objects include distributions, covariance matrices and covariance surfaces, networks and other objects. The increasing ubiquity of samples of random objects has stimulated the development of metric statistics, a collection of statistical tools to characterize, infer and relate samples of such random objects. For the geometric exploration of random objects we introduce depth profiles; the depth profile for any given object is the distribution of distances to all other objects. These distributions are amenable to further statistical analysis, leading to notions of transport ranks and centrality. Theory concerns the convergence of the empirical estimates to the population targets. For random objects in geodesic spaces, we present transport regression, based on a transport algebra, where predictors and responses are transports. We discuss theory, comparisons with Fréchet regression and Wasserstein regression and various data illustrations. Parts of this presentation are based on joint work with Yaqing Chen, Paromita Dubey and Changbo Zhu.

Tuesday 28-06-2022	08:30-09:15	Room: Beveridge Hall	Chair: Thomas Mikosch	Wald Lecture II

Universality and Crossover in 1+1 Dimensions (Part II)
Speaker: **Martin Hairer, Imperial College London**
The second lecture will focus on a third universality class which we call the Brownian Castle (BC). This is a Markov process on a space of discontinuous functions on the line which has an explicit and relatively simple description in terms of the Brownian Web. We conjecture that the BC is less stable than either EW or KPZ, and it is again natural to ask whether it admits crossovers processes connecting it to either of these models. We will see that such processes do indeed exist, at least as far as the connection to EW is concerned. In very stark contrast to the crossover from EW to KPZ however, there is an infinite-dimensional family of such processes that all admit local descriptions and are all distinct.

These lectures are based on joint works (partly still in progress) with A. Gerasimovics, K. Matetski, G. Cannizzaro, and R. Sun

Tuesday 28-06-2022	09:20-10:05	Room: G22 & G26	Chair: Tailen Hsing	Medallion Lecture I

A Doob h-process and its Applications to Singular Integrals on Z^d
Speaker: **Rodrigo Bañuelos, Purdue University**
The speaker will discuss sharp ℓ^p inequalities for a collection of discrete singular integrals on the d-dimensional lattice Z^d, $d \geq 1$, that arise from a Doob h-process where the function h is the periodic Poisson kernel. This is a modification of the Gundy–Varopoulos construction widely used to obtain sharp or near sharp estimates for singular integrals and Fourier multipliers in different geometric settings, including Euclidean spaces, manifolds, Lie groups and Wiener space. When $d = 1$ the new construction leads to the discrete Hilbert transform introduced by David Hilbert at the beginning of the 20th century. This case shows that the classical Hilbert transform on the real line and the discrete Hilbert transform on the integers have the same p–norms, $1 < p < \infty$, which had been a long-standing conjecture initiated in part by an erroneous proof of E. C. Titchmarsh in 1926. Similar problems for other discrete Calderón-Zygmund singular integral motivate the construction for $d > 1$.

Tuesday 28-06-2022	09:20-10:05	Room: Beveridge Hall	Chair: Yingying Fan	Medallion Lecture II

Distribution-free Prediction: Exchangeability and Beyond
Speaker: **Rina Foygel Barber, University of Chicago**
Conformal prediction is a recently developed method that provides predictive inference for the output of any estimation algorithm, without requiring assumptions on the data distribution or on the underlying algorithm. In this talk, I will describe two recent extensions to the conformal prediction framework. First, I will present the jackknife+ and CV+ methods, which are modifications of

cross-validation (and leave-one-out cross-validation, also known as the jackknife) that ensure distribution-free predictive coverage. These methods offer a compromise between the computational burden of full conformal prediction, and the sample size loss of split conformal prediction. As for conformal prediction, jackknife+ and CV+ assume exchangeability of the data - in particular, the training and test data must be drawn from the same distribution, and the model fitting algorithm must treat data points symmetrically. In some applications, both of these conditions may be too restrictive — we may suspect that our data is not exchangeable due to phenomena such as distribution drift, and moreover we may also wish to fit models using regression methods that do not treat data points symmetrically in order to correct for this potential drift. The second extension is a new framework for non-exchangeable conformal methods (including non-exchangeable versions of conformal prediction, split conformal prediction, and the jackknife+), where both of these assumptions are relaxed. This work is joint with Emmanuel Candès, Aaditya Ramdas, and Ryan Tibshirani.

Wednesday 29-06-2022	08:30-09:15	Room: Beveridge Hall	Chair: Piotr Fryzlewicz	Neyman Lecture

Genes, Brain, and Us

Speaker: **Heping Zhang, Yale University**

Many human conditions, including cognition, are complex and depend on both genetic and environmental factors. After the completion of the Human Genome Project, genome-wide association studies have associated genetic markers such as single-nucleotide polymorphisms with many human conditions and diseases. Despite the progress, it remains difficult to identify genes and environmental factors for complex diseases - the so-called geneticist's nightmare. Furthermore, although the impact of these discoveries on human health is not shock and awe, "drugs with support from human genetic studies for related effects succeed from phase I trials to final approval twice as often as those without such evidence." Therefore, it is important and promising, while challenging, to identify genetic variants for complex human health-related conditions. Many of us devoted a tremendous amount of effort, or even our entire careers, to develop statistical theory and methods to meet this challenge. In this talk, I will highlight some of the work that many of my students assisted me in over the past several years. The first area is the identification of super-variants. A super-variant is a set of alleles in multiple loci of human genome although unlike the loci in a gene, contributing loci to a super-variant can be anywhere in the genome. The concept of super-variant follows a common practice in genetic studies by the means of collapsing a set of variants, specifically single nucleotide polymorphisms. The novelty and challenge lie in how to find, replicate, interpret, and eventually make use of the super-variants. Our work has been mainly based on the use of tree- and forest-based methods, and a data analytic flow that we proposed in 2007, which in retrospect resembles the spirit of "deep learning" that Hinton coined in 2006. The second area is our progress in conducting statistical inference for high dimensional and structured data objects. Such data objects not only more and more commonly appear in imaging genetic studies, but also in other areas of data science including artificial intelligence. They do not belong to a Euclidean space for which most of the statistical theory and methods such as the distribution function are developed. I will introduce some concepts and important properties of ball covariance and divergence, which, respectively, act as a measure of dependence between two random objects in two possibly different Banach spaces, and as a measure of difference between two probability measures in separable Banach spaces.

Wednesday 29-06-2022	09:20-10:05	Room: G22 & G26	Chair: Adam Jakubowski	Medallion Lecture III

Multiplicative Coalescent Related Processes

Speaker: **Vlada Limic, Strasbourg University**

The talk will give an overview of some recent, either completed or ongoing, research on scaling limits of near-critical random graphs. The graphs in these studies are either close to homogenous, or they are class-wise homogenous. After recalling the classical results in this area, my plan is to describe main results in the recent joint papers with Vitalii Konarovskyi, and finally to outline ongoing research projects with Josué Corujo, and with David Clancy and Vitalii Konarovskyi. Keywords: near-critical random graphs, multiplicative coalescent, breadth-first walks, stochastic block model, interacting multiplicative coalescent, scaling limits.

Wednesday 29-06-2022	09:20-10:05	Room: Beveridge Hall	Chair: Andreas Kyprianou	Medallion Lecture IV

Privacy, Probability, and Synthetic Data

Speaker: **Roman Vershynin, University of California, Irvine**

In a world where artificial intelligence and data science become omnipresent, data sharing is increasingly locking horns with data-privacy concerns. Among the main data privacy concepts that have emerged are anonymization and differential privacy. Today, another solution is gaining traction: synthetic data. The goal of synthetic data is to create an as-realistic-as-possible dataset, one that not only maintains the nuances of the original data, but does so without risk of exposing sensitive information. The combination of differential privacy with synthetic data has been suggested as a best-of-both-worlds solution. However, the road to privacy is paved with NP-hard problems. This talk outlines three probabilistic approaches toward creating synthetic data that come with provable privacy and utility guarantees and doing so computationally efficiently. This is joint work with March Boedihardjo and Thomas Strohmer.

Thursday 30-06-2022	16:00-16:45	Room: Beveridge Hall	Chair: Adam Jakubowski	IMS/BS Schramm Lecture

Monotonicity for Continuous-Time Random Walks
Speaker: **Russell Lyons, Indiana University**
Variable-speed, continuous-time random walk on a graph is given by an assignment of nonnegative rates to its edges. There are independent Poisson processes associated to the edges with the given rates. When a walker is at a vertex, it jumps to a neighbor at the time of the next event that occurs for the corresponding incident edges. In the case of a Cayley graph of a finitely generated group, we are particularly interested in the setting where the edge rates depend only on the corresponding generators. Our lecture is concerned with monotonicity in the rates for various fundamental properties of random walks. We will survey results, counterexamples, and open questions. We will give general ideas of proofs, but avoid technicalities. Most of the talk will be devoted to two questions on Cayley graphs: On infinite graphs, we ask about the limiting linear rate of escape, i.e., the limit of the distance divided by the time. Does this increase when the rates are increased? On finite graphs, we ask about the convergence to the stationary (uniform) distribution. Does this happen faster when the rates are increased? It turns out that both questions have surprising answers. This is joint work with Graham White.

Thursday 30-06-2022	16:50-17:35	Room: Beveridge Hall	Chair: Douglas Nychka	Wahba Lecture

On the Blending of Statistical Machine Learning and Microeconomics
Speaker: **Michael Jordan, University of California, Berkeley**
Statistical decisions are often given meaning in the context of other decisions, particularly when there are scarce resources to be shared. Managing such sharing is one of the classical goals of microeconomics, and it is given new relevance in the modern setting of large, human-focused datasets, and in data-analytic contexts such as classifiers and recommendation systems. I will discuss several recent projects that aim to explore the interface between machine learning and microeconomics, including leader/follower dynamics in strategic classification, the robust learning of optimal auctions, a Lyapunov theory for matching markets with transfers, and the use of contract theory as a way to design mechanisms for statistical inference.

Monday 27-06-2022	10:30-12:30	Room: Chancellors Hall	Chair: Ben Swallow	RSS Session

A Dynamical Systems Approach to Understand, Design, and Explore Cell Fate Transitions
Speaker: **Ruben Perez-Carrasco, Imperial College London**
Traditional descriptions of cellular states involve static pictures of gene expression. Nevertheless, during embryo development or in synthetic circuits cells are under constant transformation, challenging this static view, and requiring tools that provide a quantitative description of the timing and nature of these transients. In this talk I will discuss how stochastic dynamical systems can be applied to address this problem, demonstrating its application though a bistable switch in charge of patterning a tissue, unraveling general dynamical principles of pattern formation. I will show further how these principles can be explored with synthetic biology tools in different bifurcation scenarios.

Inferring Kinetic Parameters of Oscillatory Gene Regulation from Single Cell Time-series Data
Speaker: **Joshua Burton, University of Manchester**
Gene expression dynamics, such as stochastic oscillations and aperiodic fluctuations, have been associated with cell fate changes in multiple contexts, including development and cancer. Single cell live imaging of protein expression with endogenous reporters is widely used to observe such gene expression dynamics. However, the experimental investigation of regulatory mechanisms underlying the observed dynamics is challenging, since these mechanisms include complex interactions of multiple processes, including transcription, translation and protein degradation. Here, we present a Bayesian method to infer kinetic parameters of oscillatory gene expression regulation using an auto-negative feedback motif with delay. Specifically, we use a delay-adapted nonlinear Kalman filter within a Metropolis-adjusted Langevin algorithm to identify posterior probability distributions. Our method can be applied to time-series data on gene expression from single cells and is able to infer multiple parameters simultaneously. We apply it to published data on murine neural progenitor cells and show that it outperforms alternative methods. We further analyse how parameter uncertainty depends on the duration and time resolution of an imaging experiment, to make experimental design recommendations. This work demonstrates the utility of parameter inference on time course data from single cells and enables new studies on cell fate changes and population heterogeneity. This is joint work with Cerys S. Manning, Magnus Rattray, Nancy Papalopulu and Jochen Kursawe.

Hidden Markov Modelling for Digital Circadian and Sleep Health
Speaker: **Bärbel Finkenstädt Rand, University of Warwick**
Telemonitoring of Circadian rhythmicity in physical activity (PA), sleep, and body temperature, could identify individuals at increased risk of poor health, including cancer and cardiovascular diseases, and could be used to support chronomodulated therapies and personalized prevention. Among the numerous application fields of HMMs, there is a growing interest within the context of e-Health in order to gain insight into an individual's health status based on relevant biomarker data such as PA which can be easily and objectively measured in a non-obtrusive way under normal living conditions using accelerometry or actigraphy by means of wearable sensing devices. We will show novel HMM approaches with circadian-clock driven transition probabilities which give rise to model derived and interpretable "circadian parameters" for monitoring and quantifying a subject's circadian rhythm, as well as quality and quantity of sleep. Results of applications to individuals at risk, such as nightshiftworkers and cancer patients, will be shown and an outlook towards the development of a platform for statistical telemonitoring will be discussed.

How Living Cells do Statistics? Lessons from Fisher, Shannon, and Rényi
Speaker: **Michal Komorowski, Institute of Fundamental Technological Research Polish Academy of Sciences**
An engineer designing a communication system would use few distinct signaling components while ensuring that the output of each component is highly accurate. Yet, the signaling systems of living cells are quite different. A single stimulus usually activates several signaling effectors (outputs). Besides, cellular responses are marked with substantial cell-to-cell heterogeneity rising the question of how accurate cellular signaling actually is. In the talk, I will explore the analogy between statistical inference and cellular sensing trying to reconcile the cross-wired signaling architecture and cell-to-cell heterogeneity with the engineer's expectations.

Monday 27-06-2022	16:00-17:30	Room: Chancellors Hall	Chair: Tracy Ke	Brown Session

Statistical Inference for High-Dimensional Generalized Linear Models with Binary Outcomes
Speaker: **Rong Ma, Standford University**
We develop a unified statistical inference framework for high-dimensional binary generalized linear models (GLMs) with general link functions and unknown design distributions. A two-step weighted bias-correction method is proposed for constructing confidence intervals and simultaneous hypothesis tests for individual components of the regression vector. Minimax lower bound for the expected length is established and the proposed confidence intervals are shown to be rate-optimal up to a logarithmic factor. The theoretical analysis provides important insights on the adaptivity of optimal confidence intervals with respect to the sparsity of the regression vector. The numerical performance of the proposed procedure is demonstrated through simulation studies and an analysis of a single cell RNA-seq data set, which yields interesting biological insights that integrate well into the current literature on the cellular immune response mechanisms as characterized by single-cell transcriptomics. New lower bound techniques are introduced and they can be of independent interest to solve other inference problems in high-dimensional binary GLMs. This is a joint work

with T. Tony Cai and Zijian Guo.

Assumption-Lean Analysis of Cluster Randomized Trials in Infectious Diseases for Intent-to-Treat Effects and Network Effects

Speaker: **Chan Park, University of Wisconsin-Madison**

In infectious diseases, cluster randomized trials (CRTs) are a popular experimental design to study the effect of interventions where an entire cluster of individuals, usually households or villages, are randomized to treatment or control. When analyzing data from CRTs in infectious disease settings, investigators primarily use parametric methods, usually a mixed-effect model to adjust for pre-treatment covariates and intra-correlations within clusters, and focus on the overall intent-to-treat (ITT) effect, i.e. the population average effect of the cluster-level intervention on the outcome. While simple, if the parametric models are mis-specified, the results may be misleading. Additionally, studying only the overall ITT effect may hide heterogeneity in the ITT effect in the observed covariates. Also, individuals may not comply with the cluster-level intervention, potentially inducing meaningful spillover effects. For example, in CRTs of vaccine studies, some may actually not get vaccinated for various reasons (e.g. immunocompromised, severe side effects). However, their vaccinated peers may protect the unvaccinated individuals in the form of herd immunity. In causal inference, this protection is a type of spillover effect. The main theme of our work is to propose “assumption-lean” methods to analyze these two types of effects, the ITT effects and the network effects induced from noncompliance. That is, we lean towards making fewer assumptions, even if the effects are not point-identifiable, and the maintained assumptions are standard or generally plausible in CRTs.To study the ITT effects in an assumption-lean manner, we propose a modest extension of a nonparametric, regression-esque method that (i) work in CRTs, (ii) are invariant to affine transformations of the outcome, and (iii) have desirable asymptotic properties even when both the cluster size and the number of clusters are growing. To study the network effects induced by noncompliance, Kang and Keele (2018) showed that point-identification of these effects is generally infeasible in a CRT without strong assumptions. Instead, we follow an assumption-lean approach where we propose a new method to obtain sharp bounds of these effects. At a high level, our new method combines linear programming (LP) and risk minimization from supervised machine learning (ML) where a trained classifier from risk minimization shrinks the LP bounds for the network effects. Also, compared to existing approaches on bounds under, our bounds (i) use flexible ML classifiers to potentially make the bounds narrower and (ii) irrespective of classifiers’ quality, our bounds will always cover the desired effect, with a good classifier leading to shorter bounds. Practically, this means that investigators can potentially get shorter bounds by not only getting good data from a CRT, but also by choosing better classification algorithms from ML. We conclude by reanalyzing a CRT studying the effect of face masks and hand sanitizers on transmission of 2008 interpandemic influenza in Hong Kong. This talk is based on joint work with Dr. Hyunseung Kang.

Exact Clustering in Tensor Block Model: Statistical Optimality and Computational Limit

Speaker: **Rungang Han, Duke University**

High-order clustering aims to identify heterogeneous substructures in multi-way datasets that arise commonly in neuroimaging, genomics, social network studies, etc. The non-convex and discontinuous nature of this problem pose significant challenges in both statistics and computation. In this paper, we propose a tensor block model and the computationally efficient methods, *high-order Lloyd algorithm* (HLloyd), and *high-order spectral clustering* (HSC), for high-order clustering. The convergence guarantees and statistical optimality are established for the proposed procedure under a mild sub-Gaussian noise assumption. Under the Gaussian tensor block model, we completely characterize the statistical-computational trade-off for achieving high-order exact clustering based on three different signal-to-noise ratio regimes. The analysis relies on new techniques of high-order spectral perturbation analysis and a “singular-value-gap-free” error bound in tensor estimation, which are substantially different from the matrix spectral analyses in the literature. Finally, we show the merits of the proposed procedures via extensive experiments on both synthetic and real datasets.

Tuesday 28-06-2022	16:00-17:30	Chancellors Hall	Chair: Raaz Dwivedi	New Researchers Session

Robust Inference Using Posterior Bootstrap

Speaker: **Emilia Pompe, University of Oxford**

Bayesian inference is known to provide misleading uncertainty estimation when the considered model is misspecified. Therefore, it is of interest to design other methods, that are more robust to misspecification, while having the advantages of Bayesian inference. This talk will explore various alternatives to standard Bayesian inference under model misspecification, based on extensions of the Weighted Likelihood Bootstrap (Newton Raftery, 1994). We will talk about Posterior Bootstrap, which is an extension of Weighted Likelihood Bootstrap allowing the user to properly incorporate the prior information. We will see how Edgeworth expansions can be used to understand the impact of the prior and guide the choice of hyperparameters. Finally, we will talk about extending Posterior Bootstrap to hierarchical models.

Variational Inference versus MCMC

Speaker: **Yian Ma, University of California, San Diego**

I will introduce some recent progress towards understanding the scalability of variational inference and how it compares to the Markov chain Monte Carlo (MCMC) methods. I will focus on fact-checking the folklore that "variational inference is fast but biased, MCMC is unbiased but slow". I will follow an optimization perspective on the infinite dimensional probability space, where variational inference projects the probabilities onto a finite dimensional parameter space while MCMC leverages stochastic sample paths.

Comparison of Markov Chains via Weak Poincaré Inequalities with Application to Pseudo-marginal MCMC
Speaker: **Andi Wang, University of Bristol**
I will discuss the application of functional inequalities known as weak Poincaré inequalities to bound convergence of Markov chains to equilibrium. I will show that this enables the straightforward and transparent derivation of subgeometric convergence bounds. We will apply these to study pseudo-marginal MCMC methods for intractable likelihoods, which are subgeometric in many practical settings. We are then able to provide new insights into the practical use of pseudo-marginal algorithms, such as analysing the effect of averaging in Approximate Bayesian Computation (ABC) and to study the case of lognormal weights relevant to Particle Marginal Metropolis–Hastings (PMMH). Joint work with Christophe Andrieu, Anthony Lee and Sam Power.

Thursday 30-06-2022	13:30-15:30	Room: Chancellors Hall	Chair: Ming Yuan	Wahba Session

Grace Wahba's Work on Nonparametric Function Estimation
Speaker: **Wing Hung Wong, Stanford University**
Grace Wahba's work had a profound impact on both statistics and machine learning. In this talk I will review two aspects of her pioneering contributions to nonparametric function estimation, namely her development of regularization methods in function fitting and her discovery of the relation between smoothing splines and Bayesian learning in certain Gaussian Process model.

Decision Trees in Practice: Lower Bounds and a New Shrinkage Algorithm
Speaker: **Bin Yu, University of California, Berkeley**
Decision trees are important both as interpretable models, amenable to high-stakes decision-making, and as building blocks of ensemble methods such as random forests and gradient boosting. To help understand their performance in practice, in the first part of the talk, we investigate why there is a prediction performance gap between them and powerful but uninterpretable machine learning methods including random forests, gradient boosting, and deep learning. We partially explain this performance gap by proving sharp squared error generalization lower bounds for any decision tree fitted to a sparse additive generative model. In the second part of the talk, to improve the prediction performance of decision trees, we introduce Hierarchical Shrinkage (HS), a post-hoc algorithm that does not modify the tree structure, and instead regularizes the tree by shrinking the prediction over each node towards the sample means of its ancestors. The post-hoc shrinkage in HS is extremely fast and compatible with any tree growing algorithm, and can be used synergistically with other regularization techniques. Extensive experiments over a wide variety of real-world datasets show that HS substantially increases the predictive performance of decision trees, even when used in conjunction with other regularization techniques. Moreover, we find that applying HS to each tree in an RF often improves accuracy, as well as its interpretability by simplifying and stabilizing its decision boundaries and SHAP values. We further explain the success of HS in improving prediction performance by showing its equivalence to ridge regression on a (supervised) basis constructed of decision stumps associated with the internal nodes of a tree.

Their statistical properties, however, are not well understood. The most cited prior works have focused on deriving pointwise consistency guarantees for CART in a classical nonparametric regression setting. We take a different approach, and advocate studying the generalization performance of decision trees with respect to different generative regression models. This allows us to elicit their inductive bias, that is, the assumptions the algorithms make (or do not make) to generalize to new data, thereby guiding practitioners on when and how to apply these methods. In this paper, we focus on sparse additive generative models, which have both low statistical complexity and some nonparametric flexibility. We prove a sharp squared error generalization lower bound for a large class of decision tree algorithms fitted to sparse additive models with C1 component functions. This bound is surprisingly much worse than the minimax rate for estimating such sparse additive models. The inefficiency is due not to greediness, but to the loss in power for detecting global structure when we average responses solely over each leaf, an observation that suggests opportunities to improve tree-based algorithms, for example, by hierarchical shrinkage. To prove these bounds, we develop new technical machinery, establishing a novel connection between decision tree estimation and rate-distortion theory, a sub-field of information theory.

Non-stationary Spatial Data: Think Globally Act Locally
Speaker: **Douglas Nychka, Colorado School of Mines**
Large spatial data sets are now ubiquitous in environmental science. Fine spatial sampling or many observations across large domains provides a wealth of information and can often address new scientific questions. The richness and scale of large datasets, however, often reveal heterogeneity in spatial processes that add more complexity to a statistical analysis. Current methods in spatial statistics inherit the foundational work in nonparametric regression and splines that was pioneered by Grace Wahba and others. This talk will also trace some of the threads of this research to environmental statistics. The approach in this work is to estimate spatially varying covariance parameters in a local manner but then encode these into a sparse Markov random field model for a global representation. This strategy makes it possible to estimate and then simulate (unconditional) non-stationary Gaussian processes. This approach is illustrated for the emulation of surface temperature fields from an ensemble of climate model experiments and inherits the ideas from Grace Wahba on applying splines to climate change.

Grace Wahba's Contribution to Statistical Machine Learning and Optimization

Speaker: **Hao Helen Zhang, University of Arizona**

Grace Wahba has made monumental contributions to statistical machine learning and data science. Her reproducing kernel Hilbert space (RKHS) theory plays a central role in modern supervised learning for model building, tuning, and prediction. This talk will highlight her fundamental works in regression and classification, including support vector machines, kernel learning, and sparse modeling for high dimensional data, as well as her state-of-art optimization algorithms for analyzing heterogeneous and complex data.

Monday 27-06-2022	10:30-12:30	Invited Sessions

IP3: Room: G03 Stochastic Partial Differential Equations Chair: Benjamin Gess

Optimal Regularity in Time and Space for Stochastic Porous Medium Equations
Speaker: **Hendrik Weber, University of Bath**
We prove optimal regularity estimates in Sobolev spaces in time and space for solutions to stochastic porous medium equations. The noise term considered here is multiplicative, white in time and coloured in space. The coefficients are assumed to be Hölder continuous and the cases of smooth coefficients of at most linear growth as well as u are covered by our assumptions. The regularity obtained is consistent with the optimal regularity derived for the deterministic porous medium equation in [Gess 2020] and [Gess, Sauer, Tadmor 2020] and the presence of the temporal white noise.

The proof relies on a significant adaptation of velocity averaging techniques from their usual L^1 context to the natural L^2 setting of the stochastic case. We introduce a new mixed kinetic/mild representation of solutions to quasilinear SPDE and use L^2 based a priori bounds to treat the stochastic term.

This is joint work with S. Bruno (Bath) and B. Gess (Bielefeld and Leipzig).

Statistical Inference for Linear Anisotropic SPDEs from Multiple Local Measurements
Speaker: **Randolf Altmeyer, University of Cambridge**
The problem of parameter estimation for a general second order linear stochastic partial differential equation (SPDE) is considered. One trajectory of the solution to the SPDE is observed continuously in time and averaged in space over a small window at multiple locations. Estimators for the diffusivity, transport and reaction coefficients are constructed. These estimators are shown to be minimax rate optimal by proving an explicit lower bound in the asymptotic regime where the spatial window shrinks to zero and with a growing number of observations. Interestingly, the rate of convergence depends on the differential order in which the respective coefficient appears, with the fastest rate achieved for the diffusivity coefficient and the slowest rate for the reaction terms. The proof of the lower bound relies on an explicit analysis of the reproducing kernel Hilbert space for a sum of independent Ornstein-Uhlenbeck processes, and may be of independent interest.

A semigroup approach for quasilinear rough PDEs
Speaker: **Alexandra Neamtu, University of Konstanz**
We investigate parabolic quasilinear evolution equations driven by a γ-Hölder rough path for $\gamma \in (\frac{1}{3}, \frac{1}{2})$. We explore the mild formulation that combines functional analysis techniques and controlled rough paths theory. In this framework we also connect rough paths and pathwise mild solutions. We apply our results to the stochastic Landau-Lifshitz-Gilbert and Shigesada-Kawasaki-Teramoto equations. This talk is based on a joint work with Antoine Hocquet (TU Berlin).

IP6: Room: G04 Liouville Quantum Gravity and Planar Maps Chair: Janne Junnila and Jason Miller

Liouville conformal field theory and the Virasoro algebra
Speaker: **Guillaume Baverez, Humboldt-Universität zu Berlin**
I will show that the Virasoro algebra is represented on the Hilbert space of Liouville theory as generators of Markov processes valued in the space of distributions on the circle. As a special case, the Hamiltonian of the theory is recovered as the zero mode of the Virasoro algebra. This result gives a probabilistic interpretation of a famous result from representation theory known as the Sugawara construction. If time permits, I will present some applications of this point of view to the study of conformal blocks and scattering matrix of Liouville CFT. This is joint with Guillarmou, Kupiainen, Rhodes & Vargas.

A Tour to Imaginary Chaos
Speaker: **Janne Junnila, University of Helsinki**
I will give a brief introduction to the basic theory of imaginary multiplicative chaos. Imaginary chaos distributions are random generalized functions that can be viewed as analytic continuations of a real Gaussian multiplicative chaos measure along its intermittency parameter gamma to purely imaginary values. They also appear naturally in applications, for example in the scaling limit of the spin field of the critical XOR-Ising model. The talk will be based on joint works with Juhan Aru, Antoine Jego, Eero Saksman and Christian Webb.

Tightness for Approximations to the Chemical Distance Metric for Simple CLEs
Speaker: **Jason Miller, University of Cambridge**
We consider a conformal loop ensemble (CLE) with simple loops ($\kappa \in (8/3, 4)$) in a simply connected domain $D \subseteq \mathbf{C}$ whose boundary is itself a type of CLE_κ loop. Let Υ be the carpet of Γ, i.e., the set of points in D not surrounded by a loop of Γ. We prove that certain approximations to the chemical distance metric in Υ are tight. We also show that any subsequential limit defines a geodesic metric on Υ which is Hölder continuous with respect to the Euclidean metric. We conjecture that the subsequential limit is unique, conformally covariant, and describes the scaling limit of the chemical distance metric for discrete loop models which converge to CLE_κ for $\kappa \in (8/3, 4)$ such as the critical Ising model.

IP9: Room: G05 KPZ Universality Chair: Daniel Remenik

The Critical 2*d* Stochastic Heat Flow
Speaker: **Nikolaos Zygouras, University of Warwick**
We consider directed polymers in random environment in the critical dimension two, focusing on the intermediate disorder regime when the model undergoes a phase transition. We prove that, at the critical temperature the diffusively rescaled random field of partition functions has a unique scaling limit ; a universal process of random measures on R^2 with logarithmic correlations, which we call the Critical 2*d* Stochastic Heat Flow. This is the natural candidate for the long sought solution of the critical 2*d* Stochastic Heat Equation with multiplicative space-time white noise. Based on a joint work with Francesco Caravenna and Rongfeng Sun

Universality in Random Growth Processes
Speaker: **Sourav Sarkar, University of Cambridge**
Universality in disordered systems has always played a central role in the direction of research in Probability and Mathematical Physics, a classical example being the Gaussian universality class (the central limit theorem). In this talk, I will describe a different universality class for random growth models, called the KPZ universality class. Since Kardar, Parisi and Zhang introduced the KPZ equation in their seminal paper in 1986, the equation has made appearances everywhere from bacterial growth, fire front, coffee stain to the top edge of a randomized game of Tetris; and this field has become a subject of intense research interest in Mathematics and Physics for the last 15 to 20 years. The random growth processes that are expected to have the same scaling and asymptotic fluctuations as the KPZ equation and converge to the universal limiting object called the KPZ fixed point, are said to lie in the KPZ universality class, though this KPZ universality conjecture has been rigorously proved for only a handful of models till now. Here, I will talk about some recent results on universal geometric properties of the KPZ fixed point and the underlying landscape and show that the KPZ equation and exclusion processes converge to the KPZ fixed point under the 1:2:3 scaling, establishing the KPZ universality conjecture for these models, which were long-standing open problems in this field. The talk is based on joint works with Jeremy Quastel, Balint Virag and Duncan Dauvergne.

One-point Distribution of the Geodesic in Directed Last Passage Percolation
Speaker: **Zhipeng Liu, University of Kansas**
In the recent twenty years, there have been huge developments in understanding the universal law behind a family of 2d random growth models, the so-called Kardar-Parisi-Zhang (KPZ) universality class. Especially, limiting distributions of the height functions are identified for a number of models in this class. On the other hand, different from the height functions, the geodesics of these models are much less understood. There were studies on the qualitative properties of the geodesics in the KPZ universality class very recently, but the precise limiting distributions of the geodesic locations remained unknown. In this talk, we will discuss our recent results on the one-point distribution of the geodesic of a representative model in the KPZ universality class, the directed last passage percolation with iid exponential weights. We will give an explicit formula of the one-point distribution of the geodesic location joint with the last passage times, and its limit when the parameters go to infinity under the KPZ scaling. The limiting distribution is believed to be universal for all the models in the KPZ universality class. We will further discuss some applications of our formulas.

Stationary Measure for the Open KPZ Equation
Speaker: **Alisa Knizel, University of Chicago**
The Kardar-Parisi-Zhang (KPZ) equation is the stochastic partial differential equation that models interface growth. In the talk I will present the construction of a stationary measure for the KPZ equation on a bounded interval with general inhomogeneous Neumann boundary conditions. Along the way, we will encounter classical orthogonal polynomials, the asymmetric simple exclusion process, and precise asymptotics of q-Gamma functions. This construction is a joint work with Ivan Corwin.

IP13: Room: G07 Exact Monte Carlo for Stationary Analysis Chair: Gareth Roberts

Epsilon-Strong Simulation of Fractional Brownian Motion and Related Stochastic Differential Equations
Speaker: **Jing Dong, Columbia University**
In this talk, we introduce a class of stochastic simulation procedures that aim at approximating a wide range of continuous random objects and associated functionals of the sample path with strong error guarantees, and the estimation errors can be sequentially refined. These procedures can be easily combined with other techniques such as Multi-level Monte Carlo for efficient estimation of expectations with the added benefit of a straightforward rate of convergence analysis. In the specific context of fractional Brownian motion (fBM) and fBM driven stochastic differential equations, we develop algorithms that achieve near-optimal complexity. The development also leads to an enhanced convergence rate analysis of wavelet constructions of fBM.

Ensemble Rejection Sampling
Speaker: **George Deligiannidis, University of Oxford**
We introduce Ensemble Rejection Sampling, a scheme for exact simulation from the posterior distribution of the latent states of a class of non-linear non-Gaussian state-space models. Ensemble Rejection Sampling relies on a proposal for the high-dimensional state sequence built using ensembles of state samples. Although this algorithm can be interpreted as a rejection sampling scheme acting on an extended space, we show under regularity conditions that the expected computational cost to obtain an exact sample

increases cubically with the length of the state sequence instead of exponentially for standard rejection sampling. We demonstrate this methodology by sampling exactly state sequences according to the posterior distribution of a stochastic volatility model and a non-linear autoregressive process. We also present an application to rare event simulation. This is joint work with A. Doucet and S. Rubenthaler.

Statistical Disaggregation — A Monte Carlo Approach for Imputation under Constraints
Speaker: **Hongsheng Dai, University of Essex**
Statistical disaggregation has become more and more important for smart energy systems. A typical example of such disaggregation problems is to learn energy consumption for a higher resolution level (data recorded at higher frequency) based on data at a lower resolution (data recorded at lower frequency). Constrained models are often used in such problems and they are often very useful compared to their unconstrained counterparts in terms of reducing uncertainty and leading to an improvement of the overall performance. However, these constrained models usually are not expressible as ordinary distributions due to their intractable density functions which makes it hard to conduct further analysis. A novel constrained Monte Carlo sampling algorithm based on Langevin diffusions and rejection sampling will be presented to solve the problem of sampling from constrained models. This new method is applied to a statistical disaggregation problem for an electricity consumption dataset.

Regenerative Non-reversible MCMC and the Restore Algorithm
Speaker: **Gareth Roberts, University of Warwick**
Non-reversible MCMC schemes are now of great interest due to their strong scaling properties in both the dimension of the simulation problem and also (within statistical applications) in the size of the data set. However almost all existing methods rely on skew detailed balance constructions which can be restrictive. This talk will describe a flexible new non-reversible MCMC methodology called the Restore algorithm. The algorithm does not require any augmentation of the state space but possesses an embedded regenerative structure which facilitates the proof of good theoretical properties as well as the construction of efficient associated coupling from the past algorithms. This is joint work with Murray Pollock and Andi Wang.

IS1: Room: G16	**A Statistical View on Neural Networks**	**Chair: Michael Kohler**

A Theoretical Comparison of Deep Learning and Bayes with Deep Gaussian Process Priors
Speaker: **Johannes Schmidt-Hieber, University of Twente**
Deep neural networks have received a lot of attention recently and considerable progress has been made to build an underlying mathematical foundation. In a first part of the talk we summarise some statistical convergence results. Deep Gaussian process priors can be viewed as continuous analogues of Bayesian neural networks and this raises the question whether there is a closer link with deep learning. In the second part of the talk, we show that the posterior for a suitable deep Gaussian process prior can achieve fast posterior contraction rates and discuss the connection with deep learning. This is joint work with Gianluca Finocchio.

Analysis of Convolutional Neural Network Image Classifiers in a Rotationally Symmetric Model
Speaker: **Benjamin Walter, TU Darmstadt**
Convolutional neural network image classifiers are defined and the rate of convergence of the misclassification risk of the estimates towards the optimal misclassification risk is analyzed. Here we consider images as random variables with values in some functional space, where we only observe discrete samples as function values on some finite grid. Under suitable structural and smoothness assumptions on the functional a posteriori probability, which includes some kind of symmetry against rotation of subparts of the input image, it is shown that least squares plug-in classifiers based on convolutional neural networks are able to circumvent the curse of dimensionality in binary image classification if we neglect a resolution-dependent error term.

Over-parametrized deep neural networks minimizing the empirical risk do not generalize well
Speaker: **Adam Krzyzak, Concordia University**
Recently it was shown in several papers that backpropagation is able to find the global minimum of the empirical risk on the training data using over-parametrized deep neural networks. In this presentation, a similar result is shown for deep neural networks with the sigmoidal squasher activation function in a regression setting, and a lower bound is presented which proves that these networks do not generalize well on a new data in a sense that networks which minimize the empirical risk do not achieve the optimal minimax rate of convergence for estimation of smooth regression functions. Some situations where over-parametrized neural networks with one hidden layer achieve a good rate of convergence will also be discussed.

A Statistical Analysis of an Image Classification Problem
Speaker: **Sophie Langer, University of Twente**
In this talk we consider a simple supervised classification problem for object recognition on grayscale images. There are two possible perspectives to solve this problem. Firstly, one can interpret object recognition as a high-dimensional classification problem, where every pixel is a variable. The task is then to map these pixel values to the conditional class probabilities or the labels. Increasing the dimension makes the problem considerably harder, leading to slow convergence rates due to the curse of dimensionality. Another perspective is to view images as two-dimensional objects. Increasing the number of pixels leads to higher resolution and therefore better performance is expected for large images. Following the second route, we present a new image deformation model, for which

we propose and analyze two different classifiers. The first method estimates the image deformation by support alignment. Under a minimal separation condition, it is shown that perfect classification is possible. The second method fits a CNN to the data. We derive a rate for the misclassification error depending on the sample size and the number of pixels d^2. Under suitable conditions, this rate is of order $1/\sqrt{d}$. Because of the setting we have chosen, not only are our methods not affected by the curse of high dimension, but they actually improve with increasing dimension.

IS2: Room: G22 Optimal Transport Methods for Statistical Data Analysis Chair: Jonathan Niles-Week

Optimal Transport Distributional Regression
Speaker: **Victor Panaretos, EPFL**
We present a framework for performing regression when both covariate and response are probability distributions on a compact interval. Our regression model is based on the theory of optimal transportation, and links the conditional Fréchet mean of the response to the covariate via an optimal transport map. We define a Fréchet-least-squares estimator of this regression map, and establish its consistency and rate of convergence to the true map, under both full and partial observations of the regression pairs. Computation of the estimator is shown to reduce to a standard convex optimization problem. Based on joint work with Laya Ghodrati (EPFL).

Permuted and Unlinked Monotone Regression in Multidimension: an Approach based on Mixture Modelling and Optimal Transport
Speaker: **Bodhisattva Sen, Columbia University**
Suppose that we have a regression problem with response variable Y and predictor X, both in the d-dimensional Euclidean space, for d1. In permuted or unlinked regression we have access to separate unordered data on X and Y, as opposed to data on (X,Y)-pairs in usual regression. So far in the literature the case d=1 has received attention, see e.g., the recent papers by Rigollet and Weed [Information Inference, 8, 619–717] and Balabdaoui et al. [J. Mach. Learn. Res., 22(172), 1–60]. In this paper, we consider the general multivariate setting with d1. We show that the notion of cyclical monotonicity of the regression function is sufficient for identification and estimation in the permuted/unlinked regression model. We develop a computationally efficient and easy-to-use algorithm for denoising based on the Kiefer-Wolfowitz [Ann. Math. Statist., 27, 887–906] nonparametric maximum likelihood estimator and techniques from the theory of optimal transport. We provide explicit upper bounds on the associated mean squared denoising error for Gaussian noise. As in previous work on the case d=1, the permuted/unlinked setting involves slow (logarithmic) rates of convergence rooting in the underlying deconvolution problem.

Sample Complexity of Entropic Optimal Transport
Speaker: **Philippe Rigollet, Massachusetts Institute of Technology**
We study the sample complexity of entropic optimal transport, in high dimensions using computationally efficient plug-in estimators. We derive parametric rates for estimating various quantities of interest, including the entropic regression function which is a natural analog to the optimal transport map. Furthermore, we propose a practical model for transfer learning based on entropic optimal transport, and prove parametric oracle inequalities for a broad class of such problems. This is joint work with Austin J. Stromme (MIT).

The Sketched Wasserstein Distance for Mixture Distributions
Speaker: **Florentina Bunea, Cornell University**
The Sketched Wasserstein Distance (SWD) is a new probability distance, dedicated to mixture distributions, discrete or continuous. It is constructed by lifting any metric d defined on the space of the mixture components to a metric on the space of mixture distributions. A representation theorem shows that the SWD between any two mixture distributions equals the Wasserstein distance between their respective mixture weight vectors, the sketches, identified with discrete distributions supported on the space of mixture components. Dual-based SWD estimates, constructed on the basis of p-dimensional observations from discrete K-mixtures, are computationally and statistically attractive. They can be minimax-rate optimal, with rate that mimics that of estimation from much lower, K-dimensional, observations. Furthermore, new results on ASN estimation of potentially sparse mixture weight vectors translate into root n distributional limits of SWD estimators for discrete mixtures. A comparison with estimators of the standard Wasserstein distance, extensive simulation studies and a data analysis provide strong support to the applicability of the new Sketched Wasserstein Distance between mixture distributions.

IS4: Room: G26 Second Generation Change-Point Methods: Complex Models and Fast Computation Chair: Holger Dette

Organiser: **Alexander Aue, UC Davis**

A Fast and Efficient Change-point Detection Framework based on Approximate k-Nearest Neighbor Graphs
Speaker: **Hao Chen, University of California, Davis**
Change-point analysis is thriving in this big data era to address problems arising in many fields where massive data sequences are collected to study complicated phenomena over time. It plays an important role in processing these data by segmenting a long sequence into homogeneous parts for follow-up studies. The task requires the method to be able to process large datasets quickly and deal with various types of changes for high-dimensional data. We propose a new approach making use of approximate k-nearest

neighbor information from the observations, and derive an analytic formula to control the type I error. The time complexity of our proposed method is $O(dn(\log n + k \log d) + nk^2)$ for an n-length sequence of d-dimensional data. The test statistic we consider incorporates a useful pattern for moderate- to high- dimensional data so that the proposed method could detect various types of changes in the sequence. The new approach is also asymptotic distribution free, facilitating its usage for a broader community. We apply our method to fMRI datasets and Neuropixels datasets to illustrate its effectiveness.

Trend Filtering with Adaptive Bayesian Change-point Analysis for Count Time Series
Speaker: **Toryn Schafer, Cornell University**
Model development for sequential count-valued data characterized by small counts and non-stationarities is essential for broader applicability and appropriate inference in the scientific community. Specifically, we introduce global-local shrinkage priors into a Bayesian dynamic generalized linear model to adaptively estimate both change-points and a smooth trend for count time series. We utilize a parsimonious state-space approach to identify a dynamic signal with local parameters to track smoothness of the local mean at each time-step. This setup provides a flexible framework to detect unspecified change-points in complex series, such as those with large interruptions in local trends. We detail the extension of our approach to time-varying parameter estimation within dynamic Negative Binomial regression analysis to identify structural breaks. Finally, we compare our algorithm against several alternatives to demonstrate its efficacy in diverse simulation scenarios and empirical examples in energy and biology.

Sparse Change Detection in High-dimensional Regression
Speaker: **Fengnan Gao, Fudan University**
We introduce a new method for estimating the location of sparse changes in high-dimensional linear regression coefficients, without assuming that those coefficients are individually sparse. The procedures works by constructing different sketches (projections) of the design matrix at each time point, where consecutive projection matrices differ in sign in exactly one column. The sequence of sketched design matrices is then compared against a single sketched response vector to form a sequence of test statistics whose behavior shows a surprising link to the well-known CUSUM statistics of univariate change-point analysis. Strong theoretical guarantees are derived for the estimation accuracy of the procedure, which is computationally attractive, and simulations confirm that our methods perform well in a broad class of settings.

Are Deviations in a Gradually Varying Mean Relevant? A Testing Approach Based on Sup-Norm Estimators
Speaker: **Holger Dette, Ruhr-Universität Bochum**
Classical change point analysis aims at (1) detecting abrupt changes in the mean of a possibly non-stationary time series and at (2) identifying regions where the mean exhibits a piecewise constant behaviour. In many applications however, it is more reasonable to assume that the mean changes gradually in a smooth way. Those gradual changes may either be non-relevant (i.e., small), or relevant for a specific problem at hand, and the present paper presents statistical methodology to detect the latter. More precisely, we consider a locally stationary process with a time varying trend and propose a test for the null hypothesis that the maximum absolute deviation of the trend from a given benchmark (such as the value of the trend at the beginning of the observation period) is smaller than a given threshold. A test for this type of hypotheses is developed using an appropriate estimator for the maximum deviation. We derive the limiting distribution of a standardized version of this estimator, where the standardization depends on the Lebesgue measure of the set of extremal points of the difference between trend and benchmark. A refined procedure based on an estimate of this set is developed and its consistency is proved.

Monday 27-06-2022	13:30-15:30	Invited Sessions

IP1: Room: G03	Geometric Probability	Chair: Joseph Yukich

Fluctuations of Random Convex Interfaces
Speaker: **Pierre Calka, University of Rouen Normandy**
We consider the convex hull of a point set constituted with independent and uniformly distributed points in a smooth convex body K of R^d. We show that the rescaled maximal radial fluctuation as well as maximal facet area follow asymptotically a Gumbel extreme value distribution as the size of the input goes to infinity. These results rely in particular on the study of a so-called typical facet of the random polytope. In particular, the rates of convergence are similar to those observed for a large variety of random interfaces in probability theory (random cluster models, oriented random walks...). This is joint work with J. E. Yukich.

The Modularity of Random Graphs on the Hyperbolic Plane
Speaker: **Nikolaos Fountoulakis, University of Birmingham**
We consider a model that was introduced by Krioukov et al. in 2010, which views a complex network as an expression of hidden popularity hierarchies (i.e., nodes higher up in the hierarchy have more global reach), encapsulated by an underlying hyperbolic space. For certain parameters, this model was proved to have typical features that are observed in complex networks such as power law degree distribution, bounded average degree, and positive clustering coefficient. In this talk, we investigate the modularity of this random graph. This is a parameter which has been introduced in the context of complex networks in an attempt to quantify how close a network is to an ideal modular network in which the nodes form small interconnected communities that are joined together with relatively few edges. This is joint work with Jordan Chellig (Birmingham) and Fiona Skerman (Uppsala)

Curve Envelopes and Their Applications
Speaker: **Raúl Jiménez, University of Carlos III, Madrid**
We review different ways of surrounding curves by using data coming from a functional sample. The most known way to do it is with the k nearest neighbor curves. The band delimited by the data used for surrounding a given curve is termed its envelope. Some theory and applications of curve enveloping are discussed. This includes functional time series forecasting, reconstruction of partially observed functional data, curve classification, clustering and outlier detection for functional data.

Coverage and Connectivity in Stochastic Geometry
Speaker: **Mathew Penrose, University of Bath**
Consider a random uniform sample of size n over a smooth d-dimensional compact manifold A with boundary, embedded in R^m, with $m \geq d \geq 2$. The coverage threshold T_n is the smallest r such that the union Z of geodetic balls of radius r centred on the sample points covers A. The connectivity threshold K_n is twice the smallest r required for Z to be connected. These thresholds are random variables determined by the sample, and are of interest, for example, in wireless communications, set estimation, and topological data analysis. We discuss new results on the large-n limiting distributions of T_n and K_n. When A has unit volume, with v denoting the volume of the unit ball in R^d and $|dA|$ the perimiter of A, these take the form of weak convergence of $nvT_n^d - (2-2/d)\log n - a_d \log(\log n)$ to a Gumbel-type random variable with cumulative distribution function

$$F(x) = \exp(-b_d e^{-x} + c_d |dA|)e^{-x/2}),$$

for suitable constants a_d, c_d with $b_2 = 1$, $b_d = 0$ for $d > 2$. The corresponding result for K_n takes the same form with different constants a_d, c_d. Some of this work is joint with Xiaochuan Yang (Brunel).

IP10: Room: G05	Topology of Random Objects	Chair: Takashi Owada

Bootstrapping Persistent Betti Numbers and Other Stabilizing Statistics
Speaker: **Johannes Krebs, KU Eichstaett**
We investigate multivariate bootstrap procedures for general stabilizing statistics, with specific application to topological data analysis. The work relates to other general results in the area of stabilizing statistics, including central limit theorems for geometric and topological functionals of Poisson and binomial processes in the critical regime. Here the involved asymptotic covariance structure is difficult to determine in practice, motivating the use of a bootstrap approach. A smoothed bootstrap procedure is shown to give consistent estimation in these settings. Specific statistics considered include the persistent Betti numbers and the Euler characteristic of Čech and Vietoris-Rips complexes over point sets in Euclidean space.

Homological Connectivity in Random Čech Complexes and Poisson Approximation
Speaker: **Omer Bobrowski, Technion-Israel Institute of Technology**
A well-known phenomenon in random graphs is the phase-transition for connectivity, proved first by ErdHos Rényi in 1959. In this talk we will discuss a high-dimensional analogue of this phenomenon which we refer to as "homological connectivity". Loosely speaking, homological connectivity is the point where the homology of a (random) filtration stops changing. The model we study is the Čech complex generated over a spatial Poisson point process. We will show that there is a sequence of sharp phase transitions (for

different degrees of homology). In addition, we will show that in each critical window, the obstructions to homological connectivity have a functional Poisson process limit.

The Bulk and the Extremes of Minimal Spanning Acycles and Persistence Diagrams of Random Complexes
Speaker: **Sayan Mukherjee, Duke University**
Frieze showed that the expected weight of the minimum spanning tree (MST) of the uniformly weighted graph converges to $\zeta(3)$. Recently, this result was extended to a uniformly weighted simplicial complex, where the role of the MST is played by its higher-dimensional analogue – the Minimum Spanning Acycle (MSA). In this work, we go beyond and look at the histogram of the weights in this random MSA – both in the bulk and in the extremes. In particular, we focus on the 'incomplete' setting, where one has access only to a fraction of the potential face weights. Our first result is that the empirical distribution of the MSA weights asymptotically converges to a measure based on the shadow – the complement of graph components in higher dimensions. As far as we know, this result is the first to explore the connection between the MSA weights and the shadow. Our second result is that the extremal weights converge to an inhomogeneous Poisson point process. A interesting consequence of our two results is that we can also state the distribution of the death times in the persistence diagram corresponding to the above weighted complex, a result of interest in applied topology.

Large Deviation Principle for Geometric and Topological Functionals and Associated Point Processes
Speaker: **Christian Hirsch, Aarhus University**
We prove a large deviation principle for the point process associated to k-element connected components in $\mathbb{R}^d$ with respect to the connectivity radii $r_n \to \infty$. The random points are generated from a homogeneous Poisson point process, so that $(r_n)_{n\geq 1}$ satisfies $n^k r_n^{d(k1)} \to \infty$ and $nr_n^d \to 0$ as $n \to \infty$ (i.e., sparse regime). The rate function for the obtained large deviation principle can be represented as relative entropy. As an application, we deduce large deviation principles for various functionals and point processes appearing in stochastic geometry and topology. As concrete examples of topological invariants, we consider persistent Betti numbers of geometric complexes and the number of Morse critical points of the min-type distance function. This talk is based on joint work with Takashi Owada.

IS3: Room: G07 Interaction Discovery in Genetics Chair: Hongzhe Li

Statistical Models for Discovery Interactions and Their Computation
Speaker: **Jun Liu, Harvard University**
In this talk we will review some of our recently developed methods for detecting various interactions, such as protein-protein interactions, gene-gene interactions, and amino-acid mutation interactions. We will start with a simple Bayesian network method for mining protein-protein interactions from unstructured text (such as pubmed abstracts), then move on to models for mining sophisticated associations in general text data. Next, we consider the problem of detecting gene-gene interactions that may affect a certain trait, such as the risk of type-1 diabetes, or the expressions of a large set of genes. Our simple stepwise logistic regression method can efficiently detect interaction effects and has an interesting connection with the general multiple index models. Lastly, we suggest that some interactions may be more effectively discovered via co-evolution and co-expression information and investigate the statistical modeling implications of such intuitions.

Statistical and Machine Learning Approaches for the Identification of Virus-host Interactions
Speaker: **Fengzhu Sun, University of Southern California**
Viruses play important roles in controlling bacterial population size, altering host metabolism, and have broader impacts on the functions of microbial communities, such as human gut, soil, and ocean microbiomes. However, the investigations of viruses and their functions were vastly underdeveloped. Metagenomic studies provide enormous resources for the identifications of novel viruses and their hosts. We developed several statistical and machine learning methods, VirHostMatcher, VirHostMatcherNet, and ContigNet for the identification of prokaryotic hosts of viruses or viral contigs. Applications of these tools to metagenomics data identified a large number of novel virus-host interactions.

Microbial Co-variation Discovery using Mixture Margin Copula Models
Speaker: **Rebecca Deek, University of Pennsylvania**
The human microbiome is a complex and dynamic emergent system, characterized by the interactions that arise between its individual parts, known as microbial taxa. Due to its role in human-host health, it is important to understand these relationships and how they change with time and disease status. Emergent properties of interest include bivariate microbial co-variations, which can describe the resulting ecosystem, but the excessive zeros seen in microbial sequencing make it challenging to learn such dependency structures. Motivated by this, we propose generative fixed- and random-effects copulas, with mixed zero-beta margins, to model bivariate microbial relative abundance data in cross-sectional and longitudinal studies, respectively. Copulas allow for specification of the dependency structure separately from that of the margins and the dependence parameter has biologically relevant interpretation, capturing the taxon-taxon co-variations. In addition, both the zero component and non-zero proportions of the data contribute to the estimate of the dependence parameter. Our work shows that a two-stage estimation procedure provides valid inference of the model parameters. We also develop a two-stage likelihood ratio test that is more powerful than existing tests based on sample correlation.

We illustrate the performance of the proposed methods using simulations and several real microbiome data sets.

Discovering High-order Interaction with Signed Iterative Random Forests
Speaker: **Sumanta Basu, Cornell University**
The recent explosion of high-dimensional genomic datasets, paired with powerful supervised learning algorithms, is beginning to elucidate molecular interactions that drive development and function. However, state-of-the-art prediction algorithms are typically black-boxes, offering limited insight into the complex associations they learn. We address this challenge by building on the iterative Random Forest (iRF) algorithm to explicitly map responses as a function of the learned feature interactions. Our method, signed iRF (siRF), describes “subsets” of rules that frequently occur on iRF decision paths. We refer to these “rule subsets” as signed interactions. Signed interactions share not only the same set of interacting features but also exhibit similar thresholding behavior, and thus describe a stable relationship between interacting features and responses. We describe stable and predictive importance metrics (SPIMs) to rank signed interactions in terms of their stability, predictive accuracy, and strength of interaction. We evaluate our proposed approach in biologically inspired simulations and a case study predicting enhancer activity from TF binding.

IS11: Room: Chancellors Hall	Statistical Methods in Reinforcement Learning	Chair: Chengchun Shi

Recent Machine Learning Developments for Multiple Outcomes in Precision Health
Speaker: **Michael Kosorok, University of North Carolina at Chapel Hill**
Precision health is the science of data-driven decision support for improving health at the individual and population levels. This includes precision medicine and precision public health and circumscribes all health-related challenges which can benefit from a precision operations approach. This framework strives to develop study design, data collection and analysis tools to discover empirically valid solutions to optimize outcomes in both the short and long term. This includes utilizing multiple competing outcomes. In this presentation, we will discuss two approaches to doing this. In the first approach, we utilize patient preferences about the relative importance of the two outcomes. In the second approach, we utilize expert opinion when the experts may be subject to error. The second example involves an interesting new type of inverse reinforcement learning. The ideas will be illustrated with applications in mental health.

Policy Learning with Competing Agents
Speaker: **Stefan Wager, Stanford University**
Decision makers often aim to learn a treatment assignment policy under a capacity constraint on the number of agents that they can treat. When agents can respond strategically to such policies, competition arises, complicating the estimation of the effect of the policy. In this paper, we study capacity-constrained treatment assignment in the presence of such interference. We consider a dynamic model where heterogeneous agents myopically best respond to the previous treatment assignment policy. When the number of agents is large but finite, we show that the threshold for receiving treatment under a given policy converges to the policy’s mean-field equilibrium threshold. Based on this result, we develop a consistent estimator for the policy effect and demonstrate in simulations that it can be used for learning optimal capacity-constrained policies in the presence of strategic behavior.

Demystifying (Deep) Reinforcement Learning: The Pessimist, The Optimist, and Their Provable Efficiency
Speaker: **Zhuoran Yang, Yale University**
Coupled with powerful function approximators such as deep neural networks, reinforcement learning (RL) achieves tremendous empirical successes. However, its theoretical understanding lags behind. In particular, it remains unclear how to provably attain the optimal policy with a finite regret or sample complexity. In this talk, we will present the two sides of the same coin, which demonstrates an intriguing duality between pessimism and optimism.- In the offline setting, we aim to learn the optimal policy based on a dataset collected a priori. Due to a lack of active interactions with the environment, we suffer from insufficient coverage of the dataset. To maximally exploit the dataset, we propose a pessimistic least-squares value iteration algorithm, which achieves a minimax-optimal sample complexity.- In the online setting, we aim to learn the optimal policy by actively interacting with an environment. To strike a balance between exploration and exploitation, we propose an optimistic least-squares value iteration algorithm, which achieves a $\sqrt{T}$ regret in the presence of linear, kernel, and neural function approximators.

A Reinforcement Learning Framework for A/B Testing
Speaker: **Chengchun Shi, London School of Economics and Political Science**
A/B testing, or online experiment is a standard business strategy to compare a new product with an old one in pharmaceutical, technological, and traditional industries. Major challenges arise in online experiments of two-sided marketplace platforms (e.g., Uber) where there is only one unit that receives a sequence of treatments over time. In those experiments, the treatment at a given time impacts current outcome as well as future outcomes. The aim of this paper is to introduce a reinforcement learning framework for carrying A/B testing in these experiments, while characterizing the long-term treatment effects. Our proposed testing procedure allows for sequential monitoring and online updating. It is generally applicable to a variety of treatment designs in different industries. In addition, we systematically investigate the theoretical properties (e.g., size and power) of our testing procedure. Finally, we apply our framework to both simulated data and a real-world data example obtained from a technological company to illustrate its advantage over the current practice.

IS26: Room: G22 Recent Developments in High-Dimensional Time Series Chair: Haeran Cho & Matteo Barigozzi

Community Network Auto-Regression for High-Dimensional Time Series
Speaker: **Elynn Y. Chen, UC Berkeley**
Modeling responses on the nodes of a large-scale network is an important task that arises commonly in practice. This paper proposes a community network vector autoregressive (CNAR) model, which utilizes the network structure to characterize the dependence and intra-community homogeneity of the high dimensional time series. The CNAR model greatly increases the flexibility and generality of the network vector autoregressive (Zhu et al, 2017, NAR) model by allowing heterogeneous network effects across different network communities. In addition, the non-community-related latent factors are included to account for unknown cross-sectional dependence. The number of network communities can diverge as the network expands, which leads to estimating a diverging number of model parameters. We obtain a set of stationary conditions and develop an efficient two-step weighted least-squares estimator. The consistency and asymptotic normality properties of the estimators are established. The theoretical results show that the two-step estimator improves the one-step estimator by an order of magnitude when the error admits a factor structure. The advantages of the CNAR model are further illustrated on a variety of synthetic and real datasets.

This is joint work with Jianqing Fan and Xuening Zhu.

CP Factor Model for Dynamic Tensors
Speaker: **Rong Chen, Rutgers University**
Observations in various applications are frequently represented as a time series of multidimensional arrays, called tensor time series, preserving the inherent multidimensional structure. In this paper, we present a factor model approach, in a form similar to tensor CP decomposition, to the analysis of high-dimensional dynamic tensor time series. As the loading vectors are uniquely defined but not necessarily orthogonal, it is significantly different from the existing tensor factor models based on Tucker-type tensor decomposition. The model structure allows for a set of uncorrelated one-dimensional latent dynamic factor processes, making it much more convenient to study the underlying dynamics of the time series. A new high order projection estimator is proposed for such a factor model, utilizing the special structure and the idea of the higher order orthogonal iteration procedures commonly used in Tucker-type tensor factor model and general tensor CP decomposition procedures. Theoretical investigation provides statistical error bounds for the proposed methods, which shows the significant advantage of utilizing the special model structure. Simulation study is conducted to further demonstrate the finite sample properties of the estimators. Real data application is used to illustrate the model and its interpretations.

Rank and Factor Loadings Estimation in Time Series Tensor Factor Model by Pre-averaging
Speaker: **Clifford Lam, London School of Economics and Political Science**
Tensor time series data appears naturally in a lot of fields, including finance and economics. Asa major dimension reduction tool, similar to its factor model counterpart, the idiosyncractic components of a tensor time series factor model can exhibit serial correlations, especially in financial and economic applications. This rules out a lot of state-of-the-art methods that assume white idiosyncractic components, or even independent/Gaussian data. While the traditional higher order orthogonal iteration (HOOI) is proved to be convergent to a set of factor loading matrices, the closeness of them to the true underlying factor loading matrices are in general not established, or only under some strict circumstances like having i.i.d. Gaussian noises (Zhang and Xia, 2018). Under the presence of serial and cross-correlations in the idiosyncractic components and time series variables with only bounded fourth order moments, we propose a pre-averaging method that accumulates information from tensor fibres for better estimating all the factor loading spaces. The estimated directions corresponding to the strongest factors are then used for projecting the data for a potentially improved re-estimation of the factor loading spaces themselves, with theoretical guarantees and rate of convergence spelt out. We also propose a new rank estimation method which utilizes correlation information from the projected data, in the same spirit as Fan et al. (2020) for factor models with independent data. Extensive simulation results reveal competitive performance of our rank and factor loading estimators relative toother state-of-the-art or traditional alternatives. A set of financial data is also analysed. This is joint work with Weilin Chen.

Spectral Inference of High Dimensional Time Series
Speaker: **Danna Zhang, University of California, San Diego**
High dimensional non-Gaussian time series data are increasingly encountered in a wide range of applications. We consider the problem of spectral inference of high dimensional time series using the framework of functional dependence measure. In particular, we establish a distributional theory on high dimensional spectra estimates by Gaussian approximation, which can be applied to address various testing problems for time series. We also develop two different resampling methods to implement spectral inference in practice and show the theoretical validity in the high dimensional setting.

IS28: Room: G26 Scalable Particle Filter Algorithms and its Applications Chair: Ning Ning

Diffusion Schrodinger Bridges: Generative modeling, Inference and Applications to Filtering
Speaker: **Arnaud Doucet, University of Oxford**
Denoising diffusion models are novel powerful generative models that provide state-of-the-art performance in numerous domains but are computationally expensive at generation time. We propose a principled acceleration technique based on Schrodinger bridges.

Furthermore, we show how these techniques can be extended to perform inference. In a filtering context, this provides a novel class of particle filters that completely bypasses the use of importance sampling.

An Iterated Block Particle Filter for Inference on Coupled Dynamic Systems with Shared and Unit-specific Parameters
Speaker: **Edward Ionides, University of Michigan, Ann Arbor**
We consider inference for a collection of partially observed, stochastic, interacting, nonlinear dynamic processes. Each process is called a unit, and our primary motivation arises in biological metapopulation systems where a unit is a spatially distinct subpopulation. Block particle filters are an effective tool for simulation-based likelihood evaluation for these systems, which are strongly dependent through time on a single unit and relatively weakly coupled between units. Iterated filtering algorithms can facilitate likelihood maximization for simulation-based filters. We introduce a new iterated block particle filter algorithm applicable to parameters that are either unit-specific or shared between units. We demonstrate this algorithm to carry out inference on a coupled epidemiological model for spatio-temporal measles case report data in twenty towns.

Intermediate Distributions and Complexity Bounds for SMC
Speaker: **Anthony Lee, University of Bristol**
It is now fairly common to use Sequential Monte Carlo (SMC) algorithms for normalizing constant estimation of high-dimensional, complex distributions without any particular structure. In order for the algorithm to give reasonable accuracy, it is well known empirically that one must introduce appropriate intermediate distributions that allow the particle system to “gradually evolve” from a simple initial distribution to the complex target distribution, and one must also specify an appropriate number of particles to control the error. Since both the number of intermediate distributions and the number of particles affect the computational cost of the algorithm, it is crucial to study and attempt to minimize the computational cost of the algorithm subject to constraints on the error. We present three strategies that have been used to specify intermediate distributions and provide bounds on the computational complexity of normalizing constant estimation with well-tuned sequences, which involves obtaining bounds on the length of sequences of intermediate distributions. Although the results for SMC algorithms involve some fairly strong assumptions on the Markov kernels involved, they are to the best of our knowledge the only general results available thus far. This primarily theoretical analysis also suggests where further research is required to tune the approach.

Tuesday 28-06-2022 10:30-12:30 Invited Sessions

IP4: Room: G03 Numerics and Stochastics Chair: Evelyn Buckwar

Splitting Methods in Approximate Bayesian Computation for Partially Observed Diffusion Processes
Speaker: **Evelyn Buckwar, Johannes Kepler University Linz**
Approximate Bayesian Computation (ABC) has become one of the major tools of likelihood-free statistical inference in complex mathematical models. Simultaneously, stochastic differential equations (SDEs) have developed as an established tool for modelling time dependent, real world phenomena with underlying random effects. When applying ABC to stochastic models, two major difficulties arise. First, the derivation of effective summary statistics and proper distances is particularly challenging, since simulations from the stochastic process under the same parameter configuration result in different trajectories. Second, exact simulation schemes to generate trajectories from the stochastic model are rarely available, requiring the derivation of suitable numerical methods for the synthetic data generation. In this talk we consider SDEs having an invariant density and apply measure-preserving splitting schemes for the synthetic data generation. We illustrate the results of the parameter estimation with the corresponding ABC algorithm with simulated data.This talk is based on joined work with Massimiliano Tamborrino (University of Warwick) and Irene Tubikanec (Johannes Kepler University Linz).

Randomized Operator Splitting Schemes for Abstract Evolution Equations
Speaker: **Monika Eisenmann, Lund University**
Evolution equations are an important building block for modeling processes in physics, biology and social sciences. Moreover, optimization problems can be reformulated into evolution equations. Their applications benefit from randomized optimization methods like the stochastic gradient descent method. Such a stochastic optimizer corresponds to a randomized operator splitting scheme in the evolution equation setting. While deterministic operator splitting methods are already a powerful tool in the approximation of evolution equations, we extend the approach to a randomized version. Such a randomization can, for example, decrease the cost of a single approximation step compared to a standard time-integrator. In this talk, we propose a randomized operator splitting scheme in an abstract setting and exemplify the theory with a randomized domain decomposition scheme.

The Exponential Lie Series for Itô Integrals
Speaker: **Anke Wiese, Heriot-Watt University**
We consider stochastic differential systems driven by a multi-dimensional Wiener process (more generally by continuous semimartingales) and governed by non-commuting vector fields. The flow map describes the transport of the initial condition to the solution of the differential equation at a future time. We prove that the logarithm of the flow map is a Lie series. This is an important property for the development of strong Lie group integration schemes that ensure approximate solutions themselves lie in any homogeneous manifold on which the solution evolves.

Simulation of McKean Vlasov SDEs with Super-linear Drift in Measure and Space
Speaker: **Xingyuan Chen, University of Edinburgh**
We present a particle system-based split-step implicit Euler type scheme (SSM) for the simulation of McKean–Vlasov Stochastic Differential Equations (MV-SDEs) with drifts of super-linear growth in the measure and spatial component, with Lipschitz diffusion coefficient. The super-linear growth in the measure component stems from convolution operations with super-linear growth functions allowing in particular application to the granular media equation with multi-well confining potentials. The scheme attains, in stepsize, a classical (path-space) root mean-square error rate of $1/2-\varepsilon$ for $\varepsilon>0$ and rate $1/2$ in the non-path-space mean-square error metric. Several numerical examples illustrate that the standard and the taming scheme fail while the SSM performs as expected and preserves geometric properties.

IP5: Room: G04 Percolation Chair: Vincent Tassion

Almost Sharp Sharpness for Boolean Percolation
Speaker: **Barbara Dembin, ETH Zurich**
We consider a Poisson point process on $\mathbb{R}^d$ with intensity λ for $d \geq 2$. On each point, we independently center a ball whose radius is distributed according to some power-law distribution. When the distribution has a finite d-moment, there exists a non-trivial phase transition in λ associated to the existence of an infinite connected component of balls. We aim here to prove that the subcritical regime behaves well in some sense. We will see that the problematic case is when the distribution has fat tails and that in that case we cannot prove sharp threshold in λ. We prove that the subcritical regime is sharp for all but a countable number of power-law distributions. Joint work with Vincent Tassion.

Height Function Delocalisation on Cubic Planar Graphs
Speaker: **Piet Lammers, IHES**
Delocalisation plays an important role in statistical physics. This talk will discuss the delocalisation transition in the context of height functions, which are integer-valued functions on the square lattice or similar two-dimensional graphs. By drawing a link with a phase coexistence result for site percolation on planar graphs, we prove delocalisation for a broad class of height functions on planar

graphs of degree three. The proof also uses a new technique for symmetry breaking. The analysis includes several popular models such as the discrete Gaussian model, the solid-on-solid model, and the uniformly random K-Lipschitz function. Inclusion of the first model also implies the BKT phase transition in the Villain model on the triangular lattice. The talk is based on arXiv:2012.09687

Sharpness of the Phase Transition for Level-set Percolation of Strongly Correlated Gaussian Fields
Speaker: **Stephen Muirhead, University of Melbourne**
We study the phase transition in the connectivity of the excursion sets of strongly correlated Gaussian fields. We establish 'sharpness' of the transition for a wide class of fields, discrete and continuous, whose correlations decay algebraically with exponent $\alpha \in (0,d)$, including the Gaussian free field on $Z^d, d \geq 3 (\alpha = d-2)$, the Gaussian membrane model on $Z^d, d \geq 5 (\alpha = d-4)$, among other examples. This result is new for all models in dimension $d \geq 3$ except the Gaussian free field, for which sharpness was proven in a recent breakthrough by Duminil-Copin, Goswami, Rodriguez and Severo; even then, our proof is technically simpler and yields new near-critical information on the percolation density.

Recent Progress on Strongly Correlated Percolation Models
Speaker: **Pierre-Francois Rodriguez, Imperial College London**
The talk will report on recent progress concerning the near-critical behaviour of certain percolation models in three dimensions. A case in point is the phase transition associated to two related percolation problems involving the Gaussian free field (GFF) in 3D. In one of these cases, they determine a unique "fixed point" associated to the transition, comprising a set of exponents which are proved to obey Fisher's scaling law. This is one of several relations classically conjectured by physicists to hold on the grounds of a corresponding scaling ansatz.

IP7: Room: G05	**Random Matrices**	**Chair: Florence Merlevede**

Localization and Delocalization in ErdHos–Rényi graphs
Speaker: **Johannes Alt, University of Geneva**
We consider the ErdHos–Rényi graph on N vertices with edge probability d/N. It is well known that the structure of this graph changes drastically when d is of order log N. Below this threshold it develops inhomogeneities which lead to the emergence of localized eigenvectors, while the majority of the eigenvectors remains delocalized. In this talk, I will present the phase diagram depicting these localized and delocalized phases in the eigenvalue-d-plane and our recent progress in establishing it rigorously. This is based on joint works with Raphael Ducatez and Antti Knowles.

Quantitative Estimates on Random Matrices using Free Probability Tools
Speaker: **Marwa Banna, New York University Abu Dhabi**
In this talk, I highlight the connection between free probability theory and random matrices and show how free probability tools can be used to obtain different results on random matrices such as: regularity properties of limiting spectral distributions and obtaining quantitative bounds for such limit theorems.

Extensions of Approximate Message Passing Algorithms
Speaker: **Zhou Fan, Yale University**
Approximate Message Passing (AMP) algorithms are a general family of iterative algorithms defined by a random matrix and a sequence of non-linear functions, whose iterates admit a Gaussian State Evolution characterization in the large-dimensional limit. In this talk, I will briefly survey some of the theory and applications of these algorithms, and then present two new results: First, I will describe an extension of AMP algorithms to rotationally invariant matrices, where both the algorithm and its State Evolution are determined by the free cumulants of the spectral law. Second, I will describe a simplified approach to universality of such algorithms using matrix-tensor networks, and an extension of AMP algorithms to generalized Wigner matrices having potentially heavy-tailed entries and heterogeneous variances using this approach.

This is joint work with Tianhao Wang and Xinyi Zhong.

Properties of Large Lotka-Volterra Systems of Coupled Differential Equations with Random Interactions
Speaker: **Jamal Najim, Universite Gustave Eiel, Paris**
Lotka-Volterra (LV) systems of coupled differential equations are widely used in mathematical biology and theoretical ecology to model populations with interactions. In the case of a large system, the random interactions form a large random matrix the statistical properties of which reflect (to some extent) some properties of the underlying biological system.

In this talk, we will present some properties of such large LV systems (existence of a unique equilibrium, its positivity, and its stability properties) for some standard random matrix models (circular, elliptic, sparse).

Based on joint works with Akjouj, Clenet, El Ferchichi, Massol.

IS9: Room: G07	**Statistical Analysis of Large Social Networks**	**Chair: Tracy Ke**

A Self-normalizing Cycle Count Statistic
Speaker: **Jiashun Jin, Carnegie Mellon University**

Consider a large binary data matrix. We introduce a self-normalizing cycle count (SCC) statistic, and show that it is asymptotically N(0,1) in many settings. We show that SCC can be useful for solving several recent problems in network analysis, including network global testing, estimating the number of communities, and goodness-of-fit. We present several optimality results on SCC and support our results with more than a handful of real data examples.

Population-level Balance in Signed Networks
Speaker: **Ji Zhu, University Michigan, Ann Arbor**
Statistical network models are useful for understanding the underlying formation mechanism and characteristics of complex networks. However, statistical models for signed networks have been largely unexplored. In signed networks, there exist both positive (e.g., like, trust) and negative (e.g., dislike, distrust) edges, which are commonly seen in real-world scenarios. The positive and negative edges in signed networks lead to unique structural patterns, which pose challenges for statistical modeling. In this paper, we introduce a statistically principled latent space approach for modeling signed networks and accommodating the well-known balance theory, i.e., "the enemy of my enemy is my friend" and "the friend of my friend is my friend". The proposed approach treats both edges and their signs as random variables, and characterizes the balance theory with a novel and natural notion of population-level balance. This approach guides us towards building a class of balanced inner-product models, and towards developing scalable algorithms via projected gradient descent to estimate the latent variables. We also establish non-asymptotic error rates for the estimates, which are further verified through simulation studies. In addition, we apply the proposed approach to an international relation network, which provides an informative and interpretable model-based visualization of countries during World War II.

Resampling Methods for Networks
Speaker: **Liza Levina, University Michigan, Ann Arbor**
With network data becoming ubiquitous in many applications, many models and algorithms for network analysis have been proposed, yet methods for providing uncertainty estimates are much less common. Bootstrap and other resampling procedures have been an effective general tool for estimating uncertainty from i.i.d. samples, but resampling network data is substantially more complicated, since we typically only observe one network. This talk will provide an overview of several recent resampling methods we have developed for networks.

Optimal Estimation of the Number of Network Communities
Speaker: **Tracy Ke, Harvard University**
In network analysis, how to estimate the number of communities K is a fundamental problem. We consider a broad setting where we allow severe degree heterogeneity and a wide range of sparsity levels, and propose Stepwise Goodness of Fit (StGoF) as a new approach. This is a stepwise algorithm, where for , we alternately use a community detection step and a goodness of fit (GoF) step. We adapt SCORE Jin for community detection, and propose a new GoF metric. We show that at step m, the GoF metric diverges to in probability for all $m < K$ and converges to $N(0,1)$ if $m = K$. This gives rise to a consistent estimate for K. Also, we discover the right way to define the signal-to-noise ratio (SNR) for our problem and show that consistent estimates for K do not exist if , and StGoF is uniformly consistent for K if . Therefore, StGoF achieves the optimal phase transition.

IS12: Room: G22	High-Dimensional Statistical Learning	Chair: Peter Bühlmann

Inducement of Sparsity
Speaker: **Heather Battey, Imperial College London**
Parameter orthogonalisation (Cox and Reid, 1987) is presented as inducement of population-level sparsity. The latter is taken as a unifying theme for the talk, in which sparsity-inducing parameterisations or data transformations are sought. Three recent examples are framed in this light: sparse parameterisations of covariance models; systematic construction of factorisable transformations for the elimination of nuisance parameters; and inference in high-dimensional regression. The solution strategy for the problem of exact or approximate sparsity inducement appears to be context specific and may entail, for instance, solving one or more partial differential equation, or specifying a parameterised path through transformation or parameterisation space.

The Price of Unfairness in Linear Bandits with Biased Feedback
Speaker: **Christophe Giraud, Université Paris-Saclay**
Artificial intelligence is increasingly used in a wide range of decision making scenarios with higher and higher stakes. At the same time, recent work has highlighted that these algorithms can be dangerously biased, and that their results often need to be corrected to avoid leading to unfair decisions. In this talk, we will discuss the problem of sequential decision making with biased linear bandit feedback. At each round, a player selects an action described by a covariate and by a sensitive attribute. She receives a reward corresponding to the covariates of the action that she has chosen, but only observe a biased evaluation of this reward, where the bias depends on the sensitive attribute. The worst case regret is shown to be much larger than in classical linear bandit problems. It is specified in terms of an explicit geometrical constant, which characterizes the difficulty of bias estimation. The gap-depend rate reveals the importance of the between group gap for the difficulty of the problem. Interestingly, these results reveal a transition between a regime where the problem is as difficult as its unbiased counterpart, and a regime where it can be much harder.

Robust Regression with Covariate Filtering

Speaker: **Po-Ling Loh, University of Cambridge**

We study the problem of linear regression where both covariates and responses are potentially (i) heavy-tailed and (ii) adversarially contaminated. Several computationally efficient estimators have been proposed for the simpler setting where the covariates are sub-Gaussian and uncontaminated; however, these estimators may fail when the covariates are either heavy-tailed or contain outliers. In this work, we show how to modify the Huber regression, least trimmed squares, and least absolute deviation estimators to obtain estimators which are simultaneously computationally and statistically efficient in the stronger contamination model. Our approach is quite simple, and consists of applying a filtering algorithm to the covariates, and then applying the classical robust regression estimators to the remaining data. We show that the Huber regression estimator achieves near-optimal error rates in this setting, whereas the least trimmed squares and least absolute deviation estimators can be made to achieve near-optimal error after applying a post-processing step.

IS17: Room: G26 Recent Progress in Randomization Inference Chair: Qingyuan Zhao

Covariate-adaptive Randomization Inference in Matched Designs

Speaker: **Samuel Pimentel, University of California, Berkeley**

It is common to conduct inference in matched observational studies by proceeding as though treatment assignments within matched sets are assigned uniformly at random and using this distribution as the basis for inference. This approach ignores observed discrepancies in matched sets that may be consequential for the distribution of treatment, which are succinctly captured by within-set differences in the propensity score. We address this problem via covariate-adaptive randomization inference, which modifies the permutation probabilities to vary with estimated propensity score discrepancies and avoids requirements to exclude matched pairs or model an outcome variable. We show that the test achieves type I error control arbitrarily close to the nominal level when large samples are available for propensity score estimation. We characterize the large-sample behavior of the new randomization test for a difference-in- means estimator of a constant additive effect. We also show that existing methods of sensitivity analysis generalize effectively to covariate-adaptive randomization inference. Finally, we evaluate the empirical value of covariate-adaptive randomization procedures via comparisons to traditional uniform inference using simulations and a study of health outcomes.

Randomization Inference beyond the Sharp Null: Bounded Null Hypotheses and Quantiles of Individual Treatment Effects

Speaker: **Xinran Li, University of Illinois at Urbana-Champaign**

Randomization (a.k.a. permutation) inference is typically interpreted as testing Fisher's "sharp" null hypothesis that all effects are exactly zero. This hypothesis is often criticized as uninteresting and implausible. We show, however, that many randomization tests are also valid for a "bounded" null hypothesis under which effects are all negative (or positive) for all units but otherwise heterogeneous. The bounded null is closely related to important concepts such as monotonicity and Pareto efficiency. Inverting tests of this hypothesis yields confidence intervals for the maximum (or minimum) individual treatment effect. We then extend randomization tests to infer other quantiles of individual effects, which can be used to infer the proportion of units with effects larger (or smaller) than any threshold. The proposed confidence intervals for all quantiles of individual effects are simultaneously valid, in the sense that no correction due to multiple analyses is needed. In sum, we provide a broader justification for Fisher randomization tests, and develop exact nonpaametric inference for quantiles of heterogeneous individual effects. We illustrate our methods with simulations and applications, where we find that Stephenson rank statistics often provide the most informative results.

Sensitivity Analysis of Individual Treatment Effects: A Robust Conformal Inference Approach

Speaker: **Emmanuel Candès, Stanford University**

We propose a model-free framework for sensitivity analysis of individual treatment effects (ITEs), building upon ideas from conformal inference. For any unit, our procedure reports the Gamma-value, a number which quantifies the minimum strength of confounding needed to explain away the evidence for ITE. Our approach rests on the reliable predictive inference of counterfactuals and ITEs in situations where the training data is confounded. Under a common sensitivity model, we characterize the shift between the distribution of the observations and that of the counterfactuals. We first develop a general method for predictive inference of test samples from a shifted distribution; we then leverage this to construct covariate-dependent prediction sets for counterfactuals. No matter the value of the shift, these prediction sets (resp. approximately) achieve marginal coverage if the propensity score is known exactly (resp. estimated). We describe a distinct procedure also attaining coverage, however, conditional on the training data. In the latter case, the prediction intervals cannot in general be tightened. We analyze real datasets with our methods. This is joint work with Zhimei Ren and Ying Jin.

Conformal Prediction, Hypothesis Testing, and Rank-Sum Test

Speaker: **Jing Lei, Carnegie Mellon University**

Conformal prediction has been developed as a tool for distribution-free prediction interval construction. In this work we present some methods that use the idea of conformal prediction for some nonparametric hypothesis testing problems, including independence test and a special case of conditional independence test. These methods inherit the flexibility and distribution-free validity from conformal prediction, and are suitable for modern, complex data sets.

IS22: Room: Chancellors Hall Tensors in Statistics I Chair: Anru Zhang

Shrinking towards Separability: Covariance Estimation for Matrix-variate Data
Speaker: **Peter Hoff, Duke University**
A separable covariance model for a random matrix provides a parsimonious description of the covariances among the rows and among the columns of the matrix. However, in many applications the assumption of exact separability is unlikely to be met, and data analysis with a separable model may overlook or misrepresent important patterns in the data. As an alternative, we propose a compromise between separable and unstructured covariance estimation, by adaptively shrinking the sample covariance matrix towards separability. We show that the resulting estimator has better risk than either the separable or unstructured estimators when the population covariance is not very close to being separable, and similar risk to the separable estimator when it is.

Modelling Matrix Time Series via a Tensor CP-Decomposition
Speaker: **Jinyuan Chang, Southwestern University of Economics and Finance, Chengdu**
We propose to model matrix time series based on a tensor CP-decomposition. Instead of using an iterative algorithm which is the standard practice for estimating CP-decompositions, we propose a new and one-pass estimation procedure based on a generalized eigenanalysis constructed from the serial dependence structure of the underlying process. A key idea of the new procedure is to project a generalized eigenequation defined in terms of rank-reduced matrices to a lower-dimensional one with full-ranked matrices, to avoid the intricacy of the former of which the number of eigenvalues can be zero, finite and infinity. The asymptotic theory has been established under a general setting without the stationarity. It shows, for example, that all the component coefficient vectors in the CP-decomposition are estimated consistently with the different error rates, depending on the relative sizes between the dimensions of time series and the sample size. The proposed model and the estimation method are further illustrated with both simulated and real data; showing effective dimension-reduction in modelling and forecasting matrix time series.

High-order Joint Embedding for Multi-Level Link Prediction
Speaker: **Annie Qu, University of California, Irvine**
Link prediction infers potential links from observed networks, and is one of the essential problems in network analyses. In contrast to traditional graph representation modeling which only predicts two-way pairwise relations, we propose a novel tensor-based joint network embedding approach on simultaneously encoding pairwise links and hyperlinks onto a latent space, which captures the dependency between pairwise and multi-way links in inferring potential unobserved hyperlinks. The major advantage of the proposed embedding procedure is that it incorporates both the pairwise relationships and subgroup-wise structure among nodes to capture richer network information. In addition, the proposed method introduces a hierarchical dependency among links to infer potential hyperlinks, and leads to better link prediction. In theory we establish the estimation consistency for the proposed embedding approach, and provide a faster convergence rate compared to link prediction utilizing pairwise links or hyperlinks only. Numerical studies on both simulation settings and Facebook ego-networks indicate that the proposed method improves both hyperlink and pairwise link prediction accuracy compared to existing link prediction algorithms.

Multi-linear tensor autoregressive models
Speaker: **Han Xiao, Rutgers University**
Contemporary time series analysis has seen more and more tensor type data from many fields. In the first part of the talk, we propose a multi-linear autoregressive model for tensor-valued time series. Comparing with the traditional VAR approach, the tensor autoregressive model preserves the tensor structure and has advantages in terms of interpretability, dimension reduction and computation. We propose to use the alternating algorithms to obtain the LSE and MLE. The performance of the models and methods is demonstrated by theoretical studies and simulated and real examples. In the second part of the talk, we consider the tensor autoregressive model under co-integration. We investigate the MLE under a separable error covariance structure and provide asymptotic results for the co-integration vectors, as well as the projection onto the co-integration space.

IS24: Room: G11 Structured High-Dimensional Inference Chair: Alexandre Tsybakov

Statistical Guarantees for Generative Models without Domination
Speaker: **Arnak Dalalyan, CREST/ ENSAE / IP Paris**
The talk is based on the paper https://arxiv.org/abs/2010.09237 in which we introduce a convenient framework for studying (adversarial) generative models from a statistical perspective. It consists in modeling the generative device as a smooth transformation of the unit hypercube of a dimension that is much smaller than that of the ambient space. Naturally, the quality of the generative model is measured by means of an integral probability metric. In the particular case of integral probability metric defined through a smoothness class, we establish a risk bound quantifying the role of various parameters. In particular, it clearly shows the impact of dimension reduction on the error of the generative model. The obtained risk bounds are proved to be minimax rate optimal.

Assigning Topics to Documents by Successive Projections
Speaker: **Olga Klopp, ESSEC, Paris**
Topic models provide a useful tool to organize and understand the structure of large corpora of text documents, in particular, to discover hidden thematic structure. A common approach to topic modeling is to associate each topic with a probability distribution

on the dictionary of words and to consider each document as a mixture of topics. Since the number of topics is typically substantially smaller than the size of the corpus and of the dictionary, the methods of topic modeling can lead to a dramatic dimension reduction. In this talk I will consider the problem of estimating topics distribution for each document in a given corpus, that is, the clustering aspect of the problem. I will introduce a new algorithm, the Successive Projection Overlapping Clustering (SPOC) algorithm inspired by the Successive Projection Algorithm for separable matrix factorization. This algorithm is simple to implement and computationally fast. We establish theoretical guarantees on the performance of the SPOC algorithm, in particular, near matching minimax upper and lower bounds on its estimation risk. We also propose a new method that estimates the number of topics.

Variable Selection, Monotone Likelihood Ratio and Group Sparsity
Speaker: **Mohamed Ndaoud, University of South California**
In the pivotal variable selection problem, we derive the ex- act non-asymptotic minimax selector over the class of all s-sparse vectors, which is also the Bayes selector with respect to the uniform prior. While this optimal selector is, in general, not realizable in polynomial time, we show that its tractable counterpart (the scan selector) attains the minimax expected Hamming risk to within factor 2, and is also exact minimax with respect to the probability of wrong recovery. As a consequence, we establish explicit lower bounds under the monotone likelihood ratio property and we obtain a tight characterization of the minimax risk in terms of the best separable selector risk. We apply these general results to derive necessary and sufficient conditions of exact and almost full recovery in the location model with light tail distributions and in the problem of group variable selection under Gaussian noise.

Optimal Discriminant Analysis in High-Dimensional Latent Factor Models
Speaker: **Marten Wegkamp, Cornell University**
In high-dimensional classification problems, a commonly used approach is to first project the high-dimensional features into a lower dimensional space, and base the classification on the resulting lower dimensional projections. In this talk, we formulate a latent-variable model with a hidden low-dimensional structure to justify this two-step procedure and to guide which projection to choose. We propose a computationally efficient classifier that takes certain principal components (PCs) of the observed features as projections, with the number of retained PCs selected in a data-driven way. A general theory is established for analyzing such two-step classifiers based on any low-dimensional projections. We derive explicit rates of convergence of the excess risk of the proposed PC-based classifier. The obtained rates are further shown to be optimal up to logarithmic factors in the minimax sense. Our theory allows, but does not require, the lower-dimension to grow with the sample size and is also valid even when the feature dimension exceeds the sample size. This is joint work with Xin Bing.

Tuesday 28-06-2022	13:30-15:30	Invited Sessions

IP8: Room: G03 Growth Processes **Chair: Alexandre Stauffer**

The Brownian Castle, its Universality Class and Beyond
Speaker: **Giuseppe Cannizzaro, University of Warwick**
In the context of randomly fluctuating interfaces in (1+1)-dimensions two Universality Classes have generally been considered, the Kardar-Parisi-Zhang and the Edwards-Wilkinson. Starting from a modification of the classical Ballistic Deposition model we will show that this picture is not exhaustive and another Universality Class has to be taken into account. We will describe how it arises, briefly discuss its connections to KPZ and introduce a new stochastic process, the Brownian Castle, deeply connected to the Brownian Web, which should capture the large-scale behaviour of models within this Class. Time allowing, we will discuss the crossover regimes between the Brownian Castle and Edwards-Wilkinson. This talk is based on joint works with Martin Hairer and Rongfeng Sun.

Infection Spread in a Sea of Random Walks
Speaker: **Duncan Dauvergne, University of Toronto**
We consider a class of interacting particle systems where particles perform independent random walks on and spread an infection according to a susceptible-infected-recovered model. I will discuss a new method for understanding this model and some variants. A highlight of this method is that if recovery rate is low, then the infection survives forever with positive probability, and spreads outwards linearly leaving a herd immunity region in its wake. This is based on joint work with Allan Sly.

Non-equilibrium Multi-scale Analysis and Coexistence in Competing First Passage Percolation
Speaker: **Thomas Finn, University of Bath**
We consider first passage percolation in a hostile environment, a random competition model on Z^d wherein two growth processes FPP_1 and FPP_λ compete for the occupancy of sites as follows. From time 0, FPP_1 spreads at rate 1 from the origin while FPP_λ is dormant in seeds that are placed according to a product of Bernoulli measures of parameter p. When FPP_1 or FPP_λ attempts to occupy a seed, the occupancy is suppressed and the seed is activated. Activated seeds are occupied by FPP_λ that then spread at rate λ. Once a site is occupied by either process, it remains occupied by that process henceforth. We discuss known results for the model and the recent establishment of a coexistence regime, in which FPP_1 and FPP_λ simultaneously occupy infinite connected regions with positive probability. Our proof relies on the introduction of multi-scale analysis with non-equilibrium feedback, a novel renormalisation scheme that can handle non-equilibrium and non-monotone random processes. Counterintuitively, adding seeds may benefit FPP_1 by delaying the activation of other seeds. We believe the techniques we develop are robust and can be adapted to other models lacking equilibrium dynamics and monotonicity. This is based on joint work with Alexandre Stauffer.

Planar aggregation with subcritical fluctuations and the Hastings-Levitov models?
Speaker: **Vittoria Silvestri, University of Rome La Sapienza**
The ALE (Aggregate Loewner Evolution) models describe growing random clusters on the complex plane, built by iterated composition of random conformal maps. A striking feature of these models is that they can be used to define natural off-lattice analogues of several fundamental discrete models, such as Eden or Diffusion Limited Aggregation, by tuning the correlation between the defining maps appropriately. In this talk I will discuss shape theorems and fluctuations of ALE clusters, which include Hastings-Levitov clusters as particular cases, in the subcritical regime. Based on joint work with James Norris (Cambridge) and Amanda Turner (Lancaster). .

IP11: Room: G04 Infinitely Divisible Laws and Processes **Chair: Jan Rosinski**

Berry–Essen Theorem for Functionals of Certain Infinitely Divisible Processes
Speaker: **Andreas Basse-O'Connor, Aarhus University**
In this talk, we derive Berry–Esseen bounds for non-linear functionals of certain Infinitely Divisible Processes. More precisely, we consider the convergence rate in the Central Limit Theorem for functionals of heavy-tailed moving averages, including the linear fractional stable noise, stable fractional ARIMA processes, and stable Ornstein-Uhlenbeck processes. Our rates are obtained for the Wasserstein and Kolmogorov distances and depend strongly on the interplay between the process's memory, controlled by parameter a, and its tail-index, controlled by a parameter b. For example, we obtain the classical $n^{-1/2}$ convergence rate when the tails are not too heavy and the memory is not too strong, more precisely, when a*b>3 or a*b>4 in the Wasserstein and Kolmogorov distance, respectively. Our quantitative bounds rely on a new second-order Poincare inequality on the Poisson space, which we derive through Stein's method and Malliavin calculus. This inequality improves and generalizes a result by Last, Peccati, and Schulte. The talk is based on joint work with M. Podolskij (University of Luxembourg) and C. Thäle (Ruhr University Bochum).

How to Characterize the Mean Value of an Infinitely Divisible Process
Speaker: **Witold Bednorz, University of Warsaw**
There are new results regarding selection processes that solve the Talagrand hypothesis. I comment on this approach and then show

how the result can be generalized to characterize the mean value of any infinitely divisible process.

IP12: Room: G05 Optimal Transport in Machine Learning and Artificial Intelligence Chair: Marco Cuturi

Variational Inference via Wasserstein Gradient Flows
Speaker: **Philippe Rigollet, Massachusetts Institute of Technology**
Bayesian methodology typically generates a high-dimensional posterior distribution that is known only up to normalizing constants, making the computation even of simple summary statistics such as mean and covariance a major computational hurdle. Along with Monte Carlo Markov Chains (MCMC), Variational Inference (VI) has emerged as a central computational approach to large-scale Bayesian inference. Rather than sampling from the true posterior, VI aims at producing a simple but good approximation of the target posterior for which summary statistics are easy to compute. However, unlike MCMC theory, which is well-developed and builds on now-classical probabilistic ideas, VI is still poorly understood and dominated by heuristics. In this work, we propose a principled method for VI that builds upon the theory of gradient flows on the Bures-Wasserstein space of Gaussian measures. Akin to MCMC, it comes with theoretical guarantees when the target measure is strongly log-concave. This joint work with Francis Bach, Silvère Bonnabel, Sinho Chewi, and Marc Lambert.

Riemannian Score-Based Generative Modelling
Speaker: **Valentin De Bortoli, ENS Ulm, CNRS, Paris**
Score-based generative models (SGMs) are a powerful class of generative models that exhibit remarkable empirical performance. Score-based generative modelling (SGM) consists of a "noising" stage, whereby a diffusion is used to gradually add Gaussian noise to data, and a generative model, which entails a "denoising" process defined by approximating the time-reversal of the diffusion. Existing SGMs assume that data is supported on a Euclidean space, i.e. a manifold with flat geometry. In many domains such as robotics, geoscience or protein modelling, data is often naturally described by distributions living on Riemannian manifolds and current SGM techniques are not appropriate. We introduce here Riemannian Score-based Generative Models (RSGMs), a class of generative models extending SGMs to compact Riemannian manifolds. We demonstrate our approach on a variety of manifolds, and in particular with earth and climate science spherical data.

IS14: Room: G11 Robust Statistics in High Dimensions Chair: Po-Ling Loh

Differentially Private Inference via Noisy Optimization
Speaker: **Marco Avella Medina, Columbia University**
We propose a general optimization-based framework for computing differentially private M-estimators and a new method for the construction of differentially private confidence regions. Firstly, we show that robust statistics can be used in conjunction with noisy gradient descent and noisy Newton methods in order to obtain optimal private estimators with global linear or quadratic convergence, respectively. We establish global convergence guarantees, under both local strong convexity and self-concordance, showing that our private estimators converge with high probability to a neighborhood of the non-private M-estimators. The radius of this neighborhood is nearly optimal in the sense it corresponds to the statistical minimax cost of differential privacy up to a logarithmic term. Secondly, we tackle the problem of parametric inference by constructing differentially private estimators of the asymptotic variance of our private M-estimators. This naturally leads to the use of approximate pivotal statistics for the construction of confidence regions and hypothesis testing. We demonstrate the effectiveness of a bias correction that leads to enhanced small-sample empirical performance in simulations.

This is joint work with Casey Bradshaw and Po-Ling Loh.

Adversarial Robustness via Optimal Transport
Speaker: **Muni Sreenivas Pydi, University of Wisconsin-Madison**
Adversarial robustness refers to the robustness of models on adversarially perturbed data. In this talk, we explore the many interesting connections between adversarial robustness and optimal transport theory. We focus on a binary classification setup with 0-1 loss and give two characterizations of the best achievable error in the presence of a data-perturbing adversary: (1) as a function of an optimal transport cost between the probability distributions of the two classes, and (2) as the Bayes error of a minimax hypothesis test involving Wasserstein uncertainty sets. The first characterization leads to a recipe for finding the optimal robust decision rule, and the second characterization leads to the existence of a Nash equilibrium in a zero-sum game between the model and the adversary.

Robust Estimation and Inference for Joint Quantile and Expected Shortfall Regression
Speaker: **Wenxin Zhou, University of California, San Diego**
Expected Shortfall (ES), as a financial term, refers to the average return on a risky asset conditional on the return below a certain quantile of its distribution. The latter is also known as the Value-at-Risk (VaR). In their Fundamental Review of the Trading Book (Basel Committee, 2016, 2019), the Basel Committee on Banking Supervision confirmed the replacement of VaR with ES as the standard risk measure in banking and insurance. From a statistical perspective, we consider a linear regression framework that simultaneously models the quantile and the ES of a response variable given a set of covariates. The existing approach is based on

minimizing a joint loss function, which is not only discontinuous but also non-convex. This inevitably limits its applicability for analyzing large-scale data. Motivated by the idea of using Neyman-orthogonal scores to reduce sensitivity with respect to nuisance parameters, we propose a computationally efficient two-step procedure and its robust variant for joint quantile and ES regression. Under increasing-dimensional settings, we establish explicit non-asymptotic bounds on estimation and Gaussian approximation errors, which lay the foundation for statistical inference of ES regression. In high-dimensional sparse settings, we study the theoretical properties of regularized two-step ES regression estimator as well as its robust counterpart.

IS19: Room: G26	Analysis of Multilayer Networks	Chair: Marianna Pensky

Coevolving Latent Space Network with Attractors Models for Polarization
Speaker: **Eric Kolaczyk, University of McGill**
We develop a broadly applicable class of coevolving latent space network with attractors (CLSNA) models, where nodes represent individual social actors assumed to lie in an unknown latent space, edges represent the presence of a specified interaction between actors, and attractors are added in the latent level to capture the notion of attractive and repulsive forces. We apply the CLSNA models to understand the dynamics of partisan polarization on social media, where we expect US Republicans and Democrats to increasingly interact with their own party and disengage with the opposing party. Using longitudinal social networks from the social media platforms Twitter and Reddit, we investigate the relative contributions of positive (attractive) and negative (repulsive) forces among political elites and the public, respectively. Our goals are to disentangle the positive and negative forces within and between parties and explore if and how they change over time. Our analysis confirms the existence of partisan polarization in social media interactions among both political elites and the public. Moreover, while positive partisanship is the driving force of interactions across the full periods of study for both the public and Democratic elites, negative partisanship has come to dominate Republican elites' interactions since the run-up to the 2016 presidential election. This is joint work with Xiaojing Zhu, Cantay Caliskan, Dino P. Christenson, Kostas Spiliopoulos, and Dylan Walker.

Global and Individualized Community Detection in Inhomogeneous Multilayer Networks
Speaker: **Zongming Ma, University of Pennsylvania**
In network applications, it has become increasingly common to obtain datasets in the form of multiple networks observed on the same set of subjects, where each network is obtained in a related but different experiment condition or application scenario. Such datasets can be modeled by multilayer networks where each layer is a separate network itself while different layers are associated and share some common information. The present paper studies community detection in a stylized yet informative inhomogeneous multilayer network model. In our model, layers are generated by different stochastic block models, the community structures of which are (random) perturbations of a common global structure while the connecting probabilities in different layers are not related. Focusing on the symmetric two block case, we establish minimax rates for both global estimation of the common structure and individualized estimation of layer-wise community structures. Both minimax rates have sharp exponents. In addition, we provide an efficient algorithm that is simultaneously asymptotic minimax optimal for both estimation tasks under mild conditions. The optimal rates depend on the parity of the number of most informative layers, a phenomenon that is caused by inhomogeneity across layers. The method is extended to handle multiple and potentially asymmetric community cases. We demonstrate its effectiveness on both simulated examples and a real multi-modal single-cell dataset.

Discovering underlying dynamics in time series of networks
Speaker: **Carey Priebe, Johns Hopkins University**
We study a model for time series of networks where the generating mechanism for the underlying latent positions exhibit low-dimensional manifold structure under a suitable notion of distance. This distance can then be approximated by a measure of separation between the networks themselves. When a Euclidean realization of this distance exists, we can find consistent Euclidean representations for the underlying network distributions at any given time. This permits the visualization of network evolution over time, which yields a flow in Euclidean space. Moreover, it transforms important network inference questions, such as change-point and anomaly detection, into a more classically familiar setting and reveals underlying temporal dynamics of the network. We illustrate our methodology with simulated and real data examples, and in the latter case, we are able to precisely formulate and identify a change point, corresponding to a massive shift in pandemic policy, in a communication network of a large organization.

Clustering of Diverse Multiplex Networks
Speaker: **Marianna Pensky, University of Central Florida**
The talk introduces the DIverse MultiPLEx (DIMPLE) network model where all layers of the network have the same collection of nodes and are equipped with the Stochastic Block Models (SBM). In addition, all layers can be partitioned into groups with the same community structures, although the layers in the same group may have different matrices of block connection probabilities. The DIMPLE model generalizes a multitude of papers that study multilayer networks with the same community structures in all layers (which include the tensor block model and the checker-board model as particular cases), as well as the Mixture Multilayer Stochastic Block Model (MMLSBM), where the layers in the same group have identical matrices of block connection probabilities. Since the techniques from either of the above mentioned groups cannot be applied to the DIMPLE model, we introduce novel clustering algorithms

IS23: Room: G22 Tensors in Statistics II Chair: Han Xiao

Statistics with Orthogonally Decomposable Tensors
Speaker: **Arnab Auddy, Columbia University**
With the advent of more and more complex data generating mechanisms, it becomes necessary to model higher order interactions among the observed variables. In this talk, we will see how orthogonally decomposable tensors provide a unified framework for many such problems. While this is a natural extension of matrix SVD to tensors, they automatically provide much better identifiability properties. Moreover, a small perturbation affects each singular vector in isolation, and hence their recovery does not depend on the gap between consecutive singular values. In addition to the attractive statistical properties, these methods present us with intriguing computational considerations. To this end, we will discuss some statistical vs computational tradeoffs and describe methods of principal component estimation that have near optimal rates.

Guaranteed Functional Tensor Singular Value Decomposition with Applications in Longitudinal Microbiome Data Analysis
Speaker: **Anru Zhang, Duke University**
In this talk, we introduce the functional tensor singular value decomposition (FTSVD), a novel dimension reduction framework for tensors with one functional mode and several tabular modes. The problem is motivated by high-order longitudinal data analysis. Our model assumes the observed data to be a random realization of an approximate CP low-rank functional tensor measured on a discrete-time grid. Incorporating tensor algebra and the theory of Reproducing Kernel Hilbert Space (RKHS), we propose a novel RKHS-based constrained power iteration with spectral initialization. Our method can successfully estimate both singular vectors and functions of the low-rank structure in the observed data. Under mild assumptions, we establish the non-asymptotic contractive error bounds for the proposed algorithm. We also apply the proposed procedure to perform dimension reduction to assist in the analysis of high-dimensional longitudinal microbiome data. The new method can extract the key components in the trajectories of bacterial abundance, identify representative bacterial taxa for these key trajectories, and group subjects based on the change in bacteria abundance over time. The new method is flexible to handle microbiome measurements at irregular time points for different subjects.

Tensor PCA in High Dimensional CP Models
Speaker: **Cun-Hui Zhang, Rutgers University**
The CP decomposition for high dimensional non-orthogonal spike tensors is an important problem with broad applications across many disciplines. However, previous works with theoretical guarantee typically assume restrictive incoherence conditions on the basis vectors for the CP components. We propose new computationally efficient composite PCA and concurrent orthogonalization algorithms for tensor CP decomposition with theoretical guarantees under mild incoherence conditions. The composite PCA applies the principal component or singular value decompositions twice, first to a matrix unfolding of the tensor data to obtain singular vectors and then to the matrix folding of the singular vectors obtained in the first step. It can be used as an initialization for any iterative optimization schemes for the tensor CP decomposition. The concurrent orthogonalization algorithm iteratively estimates the basis vector in each mode of the tensor by simultaneously applying projections to the orthogonal complements of the spaces generated by others CP components in other modes. It is designed to improve the alternating least squares estimator and other forms of the high order orthogonal iteration for tensors with low or moderately high CP ranks. Our theoretical investigation provides estimation accuracy and convergence rates for the two proposed algorithms. Our implementations on synthetic data demonstrate significant practical superiority of our approach over existing methods.

IS27: Room: G07 Analyzing Stochastic Gradient Methods: Noise, Nonconvexity and Dependency Chair: Vivak Patel

Online SGD under Infinite Noise Variance
Speaker: **Murat A. Erdogdu, University of Toronto**
We study stochastic convex optimization under infinite noise variance. Specifically, when the stochastic gradient is unbiased and has uniformly bounded α-th moment, for some $\alpha \in (1,2]$, we quantify the convergence rates of stochastic gradient and mirror descent algorithms. We complement our convergence results with information-theoretic lower bounds showing that no other algorithm using only stochastic first-order oracles can achieve better rates than mirror descent.

Consistency of Stochastic Gradient Descent
Speaker: **Vivak Patel, University of Wisconsin-Madison**
Stochastic Gradient Descent (SGD) is a foundational estimation method in data science. Despite its wide usage for nonconvex estimation problems, SGD's theory on such problems has lagged behind, raising the question of whether SGD is actually converging to a desirable solution for such problems. In this talk, we present novel results that address this concern for very general, realistic problems.

Necessary and Sufficient Geometries for Gradient Methods
Speaker: **John Duchi, Stanford University**
We study the impact of the constraint set and gradient geometry on the convergence of online and stochastic methods for convex optimization, providing a characterization of the geometries for which stochastic gradient and adaptive gradient methods are (mini-

max) optimal. To do so, we make strong connections with classical results on Gaussian sequence estimation; among these, we show that when the constraint set is quadratically convex, diagonally pre-conditioned stochastic gradient methods are minimax optimal, in analogy to the classical result that linear estimators are optimal for Gaussian sequence estimation over quadratically convex sets. We further provide a converse that shows that when the constraints are not quadratically convex—for example, any Lp-ball for p<2—the methods are far from optimal. These results complement our understanding of the applications of different gradient methods, suggesting when various methods are likely to succeed or fail.

IS29: Room: Chancellors Hall	New Developments on Foundations of Statistical Inference in Data Science	Chair: Minge Xie

Testing in Bayesian and Frequentist Paradigms and a Possible Compromise
Speaker: **Vladimir Vovk, University of Royal Holloway**
Bayes factors are widely used in Bayesian statistics for hypothesis testing, whereas p-values are often considered to be the hallmark of hypothesis testing in frequentist statistics (although it can also be argued that there is nothing frequentist in p-values). In this talk I will introduce a middle ground, e-values, also known as betting scores, and discuss their advantages and limitations.

Exact Multivariate Inference in Estimating Equations - an Approach by Data Depth and Confidence Distribution
Speaker: **Regina Liu, Rutgers University**
We introduce a new method for uncertainty quantification of estimators using directly the finite-sample distribution of a set of estimating equations. Specifically, we use an inversion method to build an exact Confidence Distribution (CD) for target parameters. In comparison to a point or interval estimator, a CD is a distribution estimator that carries much richer information for the parameters. We construct a CD from an estimating function without any approximation. More precisely, we incorporate data depth to develop a powerful nonparametric multivariate inferential tool, called depth-CD, to deal with multivariate estimating equations under the finite sample setting. To implement this inferential procedure, we also develop a numerical computation method, called CD Monte Carlo algorithm, to generate independent Monte Carlo samples for constructing confidence regions. The procedure is general and has many desirable theoretical properties. It is particularly useful in finite sample settings where the underlying distribution of the estimating equations is asymmetric (e.g., non-normal and skewed) or irregular (e.g., non-smooth or non-differential). The proposed approach is illustrated for the finite sample inference problems in both rank regression and censored quantile regression models. Our numerical studies indicate that confidence regions derived from the proposed approach can achieve the desired coverage probabilities even when the sample size is small, and they also outperform those derived from large sample theory. (This is Joint work with Xinyu Sun and Min-ge Xie).

Beyond Wald–Neyman–Pearson
Speaker: **Peter Grünwald, CWI and Leiden University**
A standard practice in statistical hypothesis testing is to mention the p-value alongside the accept/reject decision. We show the advantages of mentioning an e-value instead. With p-values, we cannot use an extreme observation (e.g. $p \ll \alpha$) for getting better frequentist decisions. With e-values we can, since they provide Type-I risk control in a generalized Neyman–Pearson setting with the decision task (a general loss function) determined post-hoc, after observation of the data — thereby providing a handle on 'roving αs'. When Type-II risks are taken into consideration, the only admissible decision rules in the post-hoc setting turn out to be e-value-based. We also propose to replace confidence intervals and distributions by the *e-posterior*, which provides valid post-hoc frequentist uncertainty assessments irrespective of prior correctness: if the prior is chosen badly, e-intervals get wide rather than wrong, suggesting e-posterior credible intervals as a safer alternative for Bayes credible intervals. The resulting *quasi-conditional paradigm* addresses foundational and practical issues in statistical inference.

Generalized Fiducial Inference on Differentiable Manifolds
Speaker: **Jan Hannig, University of North Carolina at Chapel Hill**
We consider the problem of defining a general fiducial density on an implicitly-defined differentiable manifold. Our proposed density extends the Generalized Fiducial Distribution (GFD) of Hannig et al. (2016). The resulting GFD formula is obtained by projecting the Jacobian differential in the ambient space onto the tangent space of the manifold. To circumvent the need for an intractable marginal integral calculation, we use two Monte Carlo algorithms that can efficiently explore a constrained parameter space and adapt them for use with the Constrained GFD. To demonstrate the new GFD formula we consider a number of simple examples. We also apply this methodology to the density estimation problem using splines and an estimation of Gaussian processes. Finally, we discuss how the manifold point of view could contribute to the philosophical understanding of fiducial distribution. This is joint work with A. Murph and J. P. Williams.

Wednesday 29-06-2022	10:30-12:30	Invited Sessions

IP14: Room: G11 Financial Mathematics Chair: Caroline Hillairet

Some Principles of Cooperative Pricing in Insurance
Speaker: **Delia Coculescu, University of Zurich**
We propose pricing rules for insurance contracts in presence of insolvency risk of the insurer. These are determined by an equilibrium argumentation, using a cooperative game approach. Based on work with Ph. Arzner, F. Delbaen and K.-T. Eisele.

Regression-Based Simulation Schemes for a Class of Anticipated BSDEs and XVA Application
Speaker: **Wissal Sabbagh, Ensae Paris, CREST**
Motivated by the equations of cross valuation adjustments (XVAs) in the realistic case where capital is deemed fungible as a source of funding for variation margin, we study two numerical schemes for a class of anticipated BSDEs where the coefficient entails a conditional elicitable risk measure of the martingale part of the solution: an implicit scheme solved by Picard iteration and an explicit scheme. We establish the time-consistency of the schemes and provide elements regarding the analysis of the spatial regression error. A neural net regression implementation is benchmarked numerically in a high-dimensional and hybrid market/default risks XVA use-case.

IP16: Room: G04 Rough path theory, signatures and applications Chair: Thomas Cass

Regularity Structures and Machine Learning
Speaker: **Ilya Chevyrev, University of Edinburgh**
In many machine learning tasks, it is crucial to extract low-dimensional and descriptive features from a data set. In this talk, I present a method to extract features from multi-dimensional space-time signals which is motivated by the success of signatures in machine learning together with the success of regularity structures in the analysis of SPDEs. I will present a flexible definition of a model feature vector along with numerical experiments in which we combine these features with basic supervised linear regression to predict solutions to parabolic and dispersive PDEs with a given forcing and boundary conditions. Interestingly, in the dispersive case, the prediction power relies heavily on whether the boundary conditions are appropriately included in the model. The talk is based on the following joint work with Andris Gerasimovics and Hendrik Weber.

Framing RNN as a Kernel Method: A Neural ODE Approach
Speaker: **Adeline Fermanian, Mines Paris-PSL**
Building on the interpretation of a recurrent neural network (RNN) as a continuous-time neural differential equation, we show, under appropriate conditions, that the solution of a RNN can be viewed as a linear function of a specific feature set of the input sequence, known as the signature. This connection allows us to frame a RNN as a kernel method in a suitable reproducing kernel Hilbert space. As a consequence, we obtain theoretical guarantees on generalization and stability for a large class of recurrent networks. Our results are illustrated on simulated datasets.

Non-degeneracy of Stochastic Line Integrals
Speaker: **Xi Geng, University of Melbourne**
In this talk, we discuss quantitative criteria for the existence of density for stochastic line integrals along solutions to hypoelliptic differential equations driven by fractional Brownian motion. A main motivation and application is the signature uniqueness problem for these equations. In relation to stochastic line integrals, if time permits I will also show that they can be used to generate a natural class of elliptic differential equations with C_b^∞ vector fields, whose solutions do not have Gaussian tail in the regime of Hurst parameter $H < 1/2$. This talk is based on joint work with S. Wang and ongoing joint work with H. Boedihardjo.

Higher Order Kernel Mean Embeddings to Capture Filtrations of Stochastic Processes
Speaker: **Maud Lemercier, University of Oxford**
Stochastic processes are random variables with values in some space of paths. However, reducing a stochastic process to a path-valued random variable ignores its filtration, i.e. the flow of information carried by the process through time. By conditioning the process on its filtration, we introduce a family of higher order kernel mean embeddings (KMEs) that generalizes the notion of KME and captures additional information related to the filtration. We derive empirical estimators for the associated higher order maximum mean discrepancies (MMDs) and prove their consistency. We construct a filtration-sensitive kernel two-sample test able to pick up information that gets missed by the standard MMD test. In addition, we define a family of universal kernels on stochastic processes for solving real-world calibration and optimal stopping problems in quantitative finance via classical kernel-based regression methods.

IP19: Room: G05 Stochastic processes, extremes and risk Chair: Clement Dombry

High dimensional inference for stochastic differential equations
Speaker: **Mark Podolskij, Universite du Luxembourg**

In this talk we study the properties of the Lasso estimator of the drift component in the diffusion setting. More specifically, we consider a multivariate parametric diffusion model X observed continuously over the interval $[0,T]$ and investigate drift estimation under sparsity constraints. We allow the dimensions of the model and the parameter space to be large. We obtain an oracle inequality for the Lasso estimator and derive an error bound for the L^2-distance using concentration inequalities for linear functionals of diffusion processes. The probabilistic part is based upon elements of empirical processes theory and, in particular, on the chaining method.

Path Diagrams for Multivariate Stochastic Processes in Continuous Time
Speaker: **Lea Schenk, Karlsruhe Institute of Technology**
Real-world processes, such as neuropsychological processes in the brain, can often be modelled with multivariate stochastic processes in continuous time. However, relations between different components of the process can be complex. In order to encapsulate these relations, we introduce a graphical model, the so-called path diagram, extending the work of Eichler (2007, Journal of Econometrics) for discrete-time processes to the continuous-time setting. For a formal definition of our graphical model, we first define Granger causality and contemporaneous correlation in an appropriate way. Given a multivariate process $\mathbf{X}_V$ and a graph $G=(V,E)$, we then define the path diagram as follows. Vertices a and b are connected by a directed edge if $\mathbf{X}_a$ is Granger causal for $\mathbf{X}_b$ with respect to $\mathbf{X}_V$, and by an undirected edge whenever $\mathbf{X}_a$ and $\mathbf{X}_b$ are contemporaneously correlated with respect to $\mathbf{X}_V$. We further present sufficient conditions for the path diagram to satisfy the local and block recursive Markov property, as well as the global AMP Markov property and the global Granger-causal Markov property. As an explicit example, we investigate the path diagram for multivariate Ornstein–Uhlenbeck processes.

Extremal Processes of Spatial Branching Structures
Speaker: **Bastien Mallein, Université Sorbonne Paris Nord**
A branching random walk is a particle system on the real line in which at each integer time, every particle produces a random number of offspring centred around its current position. This process can be thought of as the analogue of random walk in the space of random point measures. It is then possible to define a branching convolution operator "$\star$" on the space of laws of point measures, where $P\star Q$ is the law of the point measure obtained by generating a point measure of law P and replacing each of its atoms by a shifted independent copy of the point measure Q. For large families of laws L, we show that laws satisfying the equality $E=L\star E$ are randomly shifted, decorated Poisson point processes. Based on a recent work of Chen, Garban and Shekhar, the laws satisfying $E=E\star L$ can also be characterized as randomly shifted decorated Poisson point processes. We show that these two results can be used in conjunction to obtain the asymptotic behaviour of the relative positions of largest particles in a number of spatial branching processes.

Palm Theory for Extremes of Stationary Time Series
Speaker: **Hrvoje Planinić, University of Zagreb**
We consider high-level exceedances of stationary time series whose finite-dimensional distributions are regularly varying and whose exceedances group into iid clusters. It is well known that the common (asymptotic) distribution of these extremal clusters (i.e., the distribution of a "typical" cluster) can be determined from the so-called tail process of the underlying time series. In this talk we will present a new, and arguably more probabilistic, view on this relationship by relying on the language of Palm theory. The key thing is to observe that the tail process represents a cluster which contains a typical high-level exceedance and as such is biased towards clusters with more exceedances.

IP20: Room: G07	Levy Processes: Recent Advances in Theory and Applications	Chair: Aleks Mijatovic

When is the Convex Hull of a Lévy Process Smooth?
Speaker: **Jorge Gonzlez Cazares, University of Warwick**
We characterise, in terms of their transition laws, the class of one-dimensional Lévy processes whose convex hull has a continuously differentiable boundary. We show that this phenomenon is exhibited by a broad class of infinite variation Lévy processes and depends subtly on the behaviour of the Lévy measure at zero. We introduce a class of strongly eroded Lévy processes, whose Dini derivatives vanish at every local minimum of the trajectory for all perturbations with a linear drift, and prove that these are precisely the processes with smooth convex hulls. We study how the smoothness of the convex hull can break and construct examples exhibiting a variety of smooth/non-smooth behaviours. Finally, we conjecture that an infinite variation Lévy process is either strongly eroded or abrupt, a claim implied by Vigon's point-hitting conjecture.

Creeping of Lévy Processes through Curves
Speaker: **Loïc Chaumont, Université d'Angers**
A Lévy process is said to creep through a curve if, at its first passage time across this curve, the process reaches it with positive probability. We first study this property for bivariate subordinators. Given the graph $\{(t,f(t)):t\geq 0\}$ of any continuous, non increasing function f such that $f(0)>0$, we give an expression of the probability that a bivariate subordinator $(Y,Z$ issued from 0 creeps through this graph in terDavid Bang, University of Warwickms of its renewal function and the drifts of the components Y and Z. We apply this result to the creeping probability of any real Lévy process through the graph of any continuous, non increasing function at a time where the process also reaches its past supremum. This probability involves the density of the renewal function of the bivariate upward ladder process as well as its drift coefficients. This is a joint work with Thomas Pellas.

How Smooth can the Convex Hull of a Lévy Path be?
Speaker: **David Bang, University of Warwick**
We describe the modulus of continuity of the derivative C' of the convex minorant of a Lévy path on a set of times, where C' is continuous and non-constant. Since the convex minorant is piecewise linear, C' has non-trivial behaviour either at the hitting time τ_s of slope $s \in \mathbb{R}$ or at time 0. Times 0 and τ_s each require a different approach, revealing an interesting diagonal connection: upper (resp. lower) functions at time 0 have a similar structure to lower (resp. upper) functions at time τ_s. We prove that the modulus of continuity depends only on the fine structure of the small jumps of the Lévy process and is independent of the time horizon. In the case of lower functions at time 0 and upper functions at time τ_s, we find sharp results essentially characterising the modulus of continuity of C' for Lévy processes attracted to a stable process. As a consequence of the relation between the lower fluctuations of the convex minorant and the path of the Lévy process, we obtain novel results for the growth rate at 0 for the meanders of wide classes of Lévy processes.

A Meyer-Ito Formula for Stable Processes via Fractional Calculus
Speaker: **Gerónimo Uribe Bravo, Universidad Nacional Autonoma de Mexico**
We aim at presenting a natural application of fractional calculus to (one-dimensional and asymmetric) stable processes by inverting their infinitesimal generator on an adequate space. We hence recover the potential in the transient case and the recurrent potential otherwise. Multiple consequences include a generalization of the celebrated Tanaka formula for Brownian motion into the stable setting, a Meyer-Ito theorem with a non-zero local time term, and some concrete semimartingale decompositions for power functions applied to stable processes.

IS5: Room: Chancellors Hall	**Local Differential Privacy**	**Chair: Angelika Rohde**

Near Instance-Optimality in Differential Privacy
Speaker: **Hilal Asi, Stanford University**
We develop two notions of instance optimality in differential privacy, inspired by classical statistical theory: one by defining a local minimax risk and the other by considering unbiased mechanisms and analogizing the Cramer–Rao bound, and we show that the local modulus of continuity of the estimand of interest completely determines these quantities. We also develop a complementary collection of mechanisms, which we term the inverse sensitivity mechanisms, which are instance optimal (or nearly instance optimal) for a large class of estimands. Moreover, these mechanisms uniformly outperform the smooth sensitivity framework on each instance for several function classes of interest, including real-valued continuous functions. We carefully present two instantiations of the mechanisms for median and robust regression estimation with corresponding experiments.

The Power of Private Likelihood-ratio Tests for Goodness-of-fit in Frequency Tables
Speaker: **Stefano Favaro, University of Torino and Collegio Carlo Alberto**
Privacy-protecting data analysis investigates statistical methods under privacy constraints. This is a rising challenge in modern statistics, as the achievement of confidentiality guarantees, which typically occurs through suitable perturbations of the data, may determine a loss in the statistical utility of the data. In this paper, we consider privacy-protecting tests for goodness-of-fit in frequency tables, this being arguably the most common form of releasing data. Under the popular framework of (ε,δ)-differential privacy for perturbed data, we introduce a private likelihood-ratio test for goodness-of-fit and we study its large sample properties, showing the importance of taking the perturbation into account to avoid a loss in the statistical significance of the test. Our main contribution provides a quantitative characterization of the trade-off between confidentiality, measured via differential privacy parameters ε and δ, and utility, measured via the power of the test. In particular, we establish a precise Bahadur–Rao type large deviation expansion for the power of the private likelihood-ratio test, which leads to: i) identify a critical quantity, as a function of the sample size and (ε,δ), which determines a loss in the power of the private likelihood-ratio test; ii) quantify the sample cost of (ε,δ)-differential privacy in the private likelihood-ratio test, namely the additional sample size that is required to recover the power of the likelihood-ratio test in the absence of perturbation. Such a result relies on a novel multidimensional large deviation principle for sum of i.i.d. random vectors, which is of independent interest.

Network Change-point Detection under Local Differential Privacy
Speaker: **Mengchu Li, University of Warwick**
Network data are ubiquitous in our daily life, containing rich but often sensitive information. We are concerned with localising change points in dynamic networks, under local differential privacy (LDP) constraints. We investigate the fundamental limits in consistently localising change points under both node and edge privacy constraints, demonstrating interesting phase transition in terms of the signal-to-noise ratio condition, accompanied by polynomial-time algorithms. The private signal-to-noise ratio conditions quantify the costs of the privacy for change point localisation problems and exhibit a different scaling in the sparsity parameter compared to the non-private counterparts. Our algorithms are shown to be optimal under the edge LDP constraint up to log factors. Under node LDP constraint, a gap exists between our upper bound and lower bound and we leave it as an interesting open problem, echoing the challenges in high-dimensional statistical inference under LDP constraints.

IS6: Room: G22	Bayesian Computation	Chair: Arnaud Doucet

Monotonic Alpha-divergence Minimisation for Variational Inference
Speaker: **Kamélia Daudel, University of Oxford**
Variational Inference methods have garnered a lot of attention in Bayesian Statistics due to their applicability to high-dimensional Machine Learning problems. Yet, the theoretical results and empirical performances of Variational Inference methods are often impacted by two factors : one, an inappropriate choice of the objective function appearing in the optimisation problem and two, a search space that is too restrictive to match the target at the end of the optimisation procedure. In this talk, we explore how we can remedy these two issues in order to build improved Variational Inference methods. More specifically, we suggest selecting the alpha-divergence as a more general class of objective functions and we propose novel ways to carry out mixture models optimisation for Variational Inference purposes. The specificity of our approach is that we derive numerically advantageous algorithms that ensure a systematic decrease in the alpha-divergence at each step. In addition, our framework allows us to unravel important connections with gradient-based schemes from the optimisation literature as well as an integrated EM algorithm from the importance sampling literature. The talk is based on joint work with Randal Douc and François Roueff.

Sequential Monte Carlo Algorithms for Agent-based Models of Disease Transmission
Speaker: **Jeremy Heng, ESSEC Business School**
Agent-based models of disease transmission involve stochastic rules that specify how a number of individuals would infect one another, recover or be removed from the population. Common yet stringent assumptions stipulate interchangeability of agents and that all pairwise contact are equally likely. Under these assumptions, the population can be summarized by counting the number of susceptible and infected individuals, which greatly facilitates statistical inference. We consider the task of inference without such simplifying assumptions, in which case, the population cannot be summarized by low-dimensional counts. We design improved particle filters, where each particle corresponds to a specific configuration of the population of agents, that take either the next or all future observations into account when proposing population configurations. Using simulated data sets, we illustrate that orders of magnitude improvements are possible over bootstrap particle filters. We also provide theoretical support for the approximations employed to make the algorithms practical.

High-dimensional MCMC Analysis: from Diffusions to Dirichlet Forms
Speaker: **Ning Ning, University of Michigan, Ann Arbor**
Markov chain Monte Carlo (MCMC) algorithms have played a significant role in statistics, physics, machine learning and others, and they are the only known general and efficient approach for some high-dimensional problems. The random walk Metropolis (RWM) algorithm as the most classical MCMC algorithm, has had a great influence on the development and practice of science and engineering. The behavior of the RWM algorithm in high-dimensional problems is typically investigated through a weak convergence result of diffusions. In this talk, I will introduce the Mosco convergence of Dirichlet forms in analyzing the RWM algorithm on graphs, whose target distribution is the Gibbs measure that includes any probability measure satisfying a Markov property.

Accelerated Sampling on Discrete Spaces with Non-Reversible Markov Jump Processes
Speaker: **Samuel Power, University of Bristol**
In Bayesian inference problems and elsewhere, Markov Chain Monte Carlo (MCMC) algorithms are an indispensable tool for sampling from complex probability distributions. On continuous state-spaces, there has been a great deal of successful work on how to construct efficient MCMC dynamics which can converge quickly, under very general circumstances. Much of this success has stemmed from identifying continuous-time dynamical processes (ODEs, SDEs, PDMPs) which admit the desired invariant measure, and then discretising those processes to form tractable discrete-time chains. This approach has apparently seen less use in the setting of discrete spaces. In this work, we aim to bridge this gap by identifying “canonical” Markov processes (both reversible and non-reversible) on structured discrete spaces which admit a given invariant measure, and then use them to derive new algorithms for efficient sampling on discrete spaces. The algorithms are parameter-free (no tuning is required) and can be simulated directly in continuous time, easily and without discretisation error. We provide theoretical results supporting the use of non-reversible dynamics, and a range of numerical experiments demonstrate the practical benefits of our algorithms.

IS21: Room: G26	Modern Approaches to Missing Data	Chair: Richard Samworth

Optimal Nonparametric Testing of Missing Completely At Random, and its Connections to Compatibility
Speaker: **Thomas Berrett, University of Warwick**
Given a set of incomplete observations, we study the nonparametric problem of testing whether data are Missing Completely At Random (MCAR). Our first contribution is to characterise precisely the set of alternatives that can be distinguished from the MCAR null hypothesis. This reveals interesting and novel links to the theory of FrÝuxin Chen, University of Pennsylvaniaechet classes (in particular, compatible distributions) and linear programming, and we leverage tools developed in these fields to propose MCAR tests that are consistent against all detectable alternatives. Moreover, we define a natural measure of ease of detectability (an incompatibility index), and exploit ideas from max-flow min-cut theory to prove that our tests achieve the optimal minimax separation rate according to this measure in certain cases.

Inference for Heteroskedastic PCA with Missing Data
Speaker: **Yuxin Chen, University of Pennsylvania**
This work studies how to construct confidence regions for principal component analysis (PCA) in high dimension, a problem that has been vastly under-explored. While computing measures of uncertainty for nonlinear/nonconvex estimators is in general difficult in high dimension, the challenge is further compounded by the prevalent presence of missing data and heteroskedastic noise. We propose a suite of solutions to perform valid inference on the principal subspace based on two estimators: a vanilla SVD-based approach, and a more refined iterative scheme called HeteroPCA. We develop non-asymptotic distributional guarantees for both estimators, and demonstrate how these can be invoked to compute both confidence regions for the principal subspace and entrywise confidence intervals for the spiked covariance matrix. Particularly worth highlighting is the inference procedure built on top of HeteroPCA, which is not only valid but also statistically efficient for broader scenarios (e.g., it covers a wider range of missing rates and signal-to-noise ratios). Our solutions are fully data-driven and adaptive to heteroskedastic random noise, without requiring prior knowledge about the noise levels and noise distributions.

Supervised Learning with Missing Values
Speaker: **Julie Josse, Inria**
An abundant literature addresses missing data in an inferential framework: estimating parameters and their variance from incomplete tables. Here, I will review recent works on supervised-learning settings: predicting a target when missing values appear in both training and testing data. First, we study the seemingly-simple case where the target to predict is a linear function of the fully-observed data and show that show that multilayer perceptrons with ReLU activation functions can be consistent but highly complex. Based on an Neumann series approximation of the optimal predictor, we propose a new principle architecture, called NeuMiss networks. Their originality and strength come from the use of a new type of non-linearity: the multiplication by the mask. We provide an upper bound of the Bayes risk of NeuMiss networks, and we show that they have a good predictive accuracy. Then, we go beyond the linear regression setting and show how imputation can be coupled with powerful learners such as random forest to achieve consistency.

High-dimensional Changepoint Estimation with Heterogeneous Missingness
Speaker: **Tengyao Wang, London School of Economics and Political Science**
We propose a new method for change-point estimation in partially-observed, high-dimensional time series that undergo a simultaneous change in mean in a sparse subset of coordinates. Our first methodological contribution is to introduce a "MissCUSUM" transformation, that captures the interaction between the signal strength and the level of missingness in each coordinate. In order to borrow strength across the coordinates, we project these MissCUSUM statistics along a direction found as the solution to an optimisation problem. The change-point can then be estimated as the location of the peak of the projected series. In a model that allows different missingness probabilities in different component series, we identify that the key interaction between the missingness and the signal is an observation-probability-weighted sum of squares of the signal change in each coordinate. More specifically, we prove that the angle between the estimated and oracle projection directions, as well as the change-point location error, are controlled with high probability by the sum of two terms, both involving this weighted sum of squares, and representing the error incurred due to noise and due to missingness respectively. The striking effectiveness of our methodology is further demonstrated both on simulated data, and on an oceanographic data set.

Wednesday 29-06-2022	13:30-15:30	Invited Sessions

IP15: Room: G03 Insurance Mathematics Chair: Mogens Steffensen

Climate Risk Management and Insurance
Speaker: **Enrico Bis, Imperial College London**
In this talk I will review some recent work on i) physical risks and climate sensitive valuation of insured assets and ii) parametric insurance design for climate resilient technology adoption.

When Insurance Gets Exciting
Speaker: **Roger Laeven, Amsterdam School of Economics**
In this talk, we consider multivariate population processes in which general, not necessarily Markovian, mutually exciting jump processes dictate the stochastic arrivals. Such models provide a natural contender for modeling a variety of contagious phenomena, including modern technology risks in actuarial science. We establish results to determine the corresponding time-dependent joint probability distribution, allowing for general intensity decay functions, general intensity jumps, and general sojourn times. We obtain an exact, full characterization of the time-dependent joint transform of the population process and its underlying intensity process in terms of a fixed-point representation and corresponding convergence results. We also derive the asymptotic tail behavior of the population process and its underlying intensity process in the setting of heavy-tailed intensity jumps. By exploiting the results we establish, arbitrary joint spatial-temporal moments and other distributional properties can now be readily evaluated using standard transform differentiation and inversion techniques, and we illustrate this in a few examples. This is based on joint work with Raviar Karim and Michel Mandjes.)

Equilibrium Investment with Random Risk Aversion
Speaker: **Mogens Steensen, University of Copenhagen**
We solve the problem of an investor who maximizes utility but faces random preferences. We propose a problem formulation based on expected certainty equivalents. We tackle the time-consistency issues arising from that formulation by applying the equilibrium theory approach. To this end, we provide the proper definitions and prove a rigorous verification theorem. We complete the calculations for the cases of power and exponential utility. For power utility, we illustrate in a numerical example, that the equilibrium stock proportion is independent of wealth, but decreasing in time, which we also supplement by a theoretical discussion. For exponential utility, the usual constant absolute risk aversion is replaced by its expectation.

IP17: Room: G04 Stochastic models in fluid dynamics Chair: Oana Lang

High Order Numerical Schemes in Time for Stochastic Models in Ocean Dynamics
Speaker: **Camilla Fiorini, Conservatoire National des Arts et Metiers, M2N Laboratory**
In this work we consider the surface quasi-geostrophic (SQG) system under location uncertainty (LU) and propose a Milstein-type scheme for these equations, which is then used in a multi-step method. The SQG system considered here consists of one stochastic partial differential equation, which models the stochastic transport of the buoyancy, and a linear operator linking the velocity and the buoyancy. In the LU setting, the Euler–Maruyama scheme converges with weak order 1 and strong order 0.5. Our aim is to develop higher order schemes in time, based on a Milstein-type scheme in a multi-step framework. First we compared different kinds of Milstein schemes. The scheme with the best performance is then included in the two-step scheme. Finally, we show how our two-step scheme decreases the error in comparison to other multi-step schemes.

LDP and CLT for Scaling Limit of SPDEs with Transport Noise
Speaker: **Lucio Galeati, University of Bonn**
This talk is based on a series of works, in collaboration with F. Flandoli and D. Luo, in which we explore a particular scaling limit involving the presence of transport-type noise on (linear and nonlinear) PDEs. As a paradigmatic example, I will focus on stochastic 2D Euler in vorticity form; under the scaling limit, weak solutions converge to those of deterministic 2D Navier-Stokes, thus inverting the classical paradigm where solutions to Euler are obtained by a vanishing viscosity limit of Navier-Stokes. I will present some recent results concerning large deviations and gaussian fluctuations underlying this convergence.

Numerical Data Assimilation for Stochastic Advection by Lie Transport Models
Speaker: **Wei Pan, Imperial College London**
In this talk, I will give an overview of calibration and Bayesian inference techniques for partially observed stochastic advection by Lie transport (SALT) models. Two such examples are 2D SALT Euler and SALT thermal quasi-geostrophic equations, both implemented using a discretisation of order 10^6. The methods incorporate a mixture of model reduction, particle filter, nudging, tempering and jittering. The work is a part of the Stochastic Transport in Upper Ocean Dynamics (STUOD) project.

On Maximal Solutions for a Stochastic Shallow Water Model
Speaker: **Oana Lang, Imperial College London**
In this talk I will present the analytical properties of a stochastic shallow water model which has been derived using the Location

Uncertainty (Mémin, 2014) approach. We prove that there exists a unique maximal strong solution and a unique global weak solution, using a method which relies on a Cauchy approximating sequence. Our method is very convenient as although the strong solution exists in a higher order Sobolev space, we show that it suffices to prove the Cauchy property in L^2. This is different from the approaches currently used in the literature which rely mainly on tightness arguments. This is joint work with Dan Crisan (Imperial College London) and Etienne Memin (Inria) and it is part of the Stochastic Transport in Upper Ocean Dynamics (STUOD) project.

IP18: Room: G05 Stochastics Interacting Systems Chair: Daniel Valesin

Slow Mixing and Cut-off in Spin Plaquette Models
Speaker: **Paul Chleboun, University of Warwick**
I will discuss the dynamics of a certain class of continuous time Markov process called spin plaquette models. These are finite range spin systems which evolve according to Glauber dynamics. They have recently attracted a attention in the physics literature because they are expected to exhibit glassy behavior, despite the absence of any disorder (like spin glasses) or hard constraints in the dynamics (like kinetically constrained models).

I will focus on the rate of convergence to the stationary distribution (relaxation and mixing time) for the square plaquette model. In certain regimes the total variation mixing time is shown to exhibit a cut off phenomena.

This is based on joint work with A. Smith (uOttawa).

Contact process on the random hyperbolic graph
Speaker: **Bruno Schapira, Aix-Marseille Universite**
AbstractWe consider the contact process on the random hyperbolic graph, both in finite and infinite volume. We will discuss metastability type results, and explicit computation of some critical exponents related to the non-extinction probability as the infection parameter approaches its critical value, which is here zero. This is joint work with A. Linker, D. Mitsche and D. Valesin.

Algebraic Approach to Stochastic Duality for Markov Processes
Speaker: **Chiara Franceschini, Modena University**
In this talk I will overview the concept of duality for Markov processes and, in particular, some special classes of interacting particle systems. I will explain how such relations arise naturally from an algebraic description of the models and provide some classical examples as well as more recent results. I conclude with some applications in the context of scaling limit for interacting particle systems.

Dynamical Noise Sensitivity for the Voter Model
Speaker: **Rangel Baldasso, Leiden University**
We study noise sensitivity of the consensus opinion of the voter model on finite graphs with respect to noise affecting the initial opinions and noise affecting the dynamics. We prove that the final opinion is stable with respect to small perturbations of the initial configuration and is sensitive to perturbations of the dynamics governing the evolution of the process. This talk is based on a joint work with G. Amir, O. Angel, and R. Peretz.

IP21: Room: G16 Regularization by noise Chair: Oleg Butkovsky

Regularisation of Rough and Almost Critical SDEs
Speaker: **Mate Gerencser, TU Wien**
The extent to which a random perturbation can regularise a differential equation depends on the scaling of noise, leading to a natural notion of criticality. Counterexamples are available in the supercritical case, but positive results are often far from the critical threshold. In this work we identify a scale of vector fields for which we show the strongest form of well-posedness (strong existence and path-by-path uniqueness) in the complete subcritical regime for any fractional Brownian noise. In a separate direction, we study the regularising effects with drifts in classical Holder spaces but with multiplicative noise, in the context of for example rough and Young differential equations. Joint works with Konstantinos Dareiotis (Leeds) and with Lucio Galeati (Bonn).

Three Applications of Stochastic Sewing to Regularization by Noise
Speaker: **Nicolas Perkowski, Free University Berlin**
Khoa Lê's stochastic sewing lemma combines martingale and rough path ideas and this leads to strong bounds for (limits of) stochastic Riemann type sums. Since its introduction in 2020 the stochastic sewing lemma quickly became an invaluable tool in stochastic analysis. I will discuss three applications to regularization by noise: the construction of weak solutions to Lévy-driven SDEs with singular drift, the regularization of diffusion coefficients in fractional SDEs, and the solution of averaged inviscous Burgers equations with low regularity initial data. This is based on joint works with Florian Bechtold, Helena Kremp and Toyomu Matsuda.

Uniqueness for Nonlinear Fokker–Planck Equations and for McKean–Vlasov SDEs: the Degenerate Case
Speaker: **Michael Rockner, University of Bielefeld**
This work is concerned with the existence and uniqueness of generalized (mild or distributional) solutions to (possibly degenerate)

Fokker—Planck equations (possibly degenerate)$\rho_t - \delta\beta(\rho) + \mathrm{div}(Db(\rho)\rho) = 0$ in $(0,\infty) \times R^d$, $\rho(0,x) = \rho_o(x)$.Under suitable assumptions on $\beta : R \to R, b : R \to R$ and $D : R^d \to R^d, d \geq 1$, this equation generates a unique flow $\rho(t) = S(t)\rho_0 : [0,\infty) \to L^1(R^d)$ as a mild solution in the sense of nonlinear semigroup theory. This flow is also unique in the class of $L^\infty((0,T)R^d) \cup L^1((0,T) \times R^d)$, $\forall T > 0$, Schwartz distributional solutions on $(0,\infty) \times R^d$. Moreover, for $\rho_0 \in L^1(R^d) \cup H^{1(R^d)}$, $t \to S(t) = \rho_0$ is differentiable from the right on $[0,\infty)$ in $H^{1(R^d)}$-norm. As a main application, the weak uniqueness of the corresponding McKean—Vlasov SDEs is proven.

Non-explosion by Stratonovich Noise for ODEs
Speaker: **Mario Maurelli, Università degli Studi di Milano**
For an ODE where the drift has superlinear growth, explosion may occur in finite time. Here we show that, in dimension two or higher, the addition of a suitable multiplicative noise of Stratonovich type prevents explosion almost surely. We also show the existence of an invariant measure and the geometric ergodicity for the corresponding SDE. The results and the proofs extend previous results and proofs on non-explosion by noise for ODEs.

IS16: Room: G22 New Developments in High-Dimensional Learning and Nonparametric Inference Chair: Yingying Fan

Optimal rates for Robust Deep Learning
Speaker: **Jianqing Fan, Princeton University**
This talk is on the stability of deep ReLU neural networks for nonparametric regression under the assumption that the noise has only a finite p-th moment. We unveil how the optimal rate of convergence depends on p, the degree of smoothness and the intrinsic dimension in a class of nonparametric regression functions with hierarchical composition structure when both the adaptive Huber loss and deep ReLU neural networks are used. This optimal rate of convergence cannot be obtained by the ordinary least squares but can be achieved by the Huber loss with a properly chosen parameter that adapts to the sample size, smoothness, and moment parameters. A concentration inequality for the adaptive Huber ReLU neural network estimators with allowable optimization errors is also derived. To establish a matching lower bound within the class of neural network estimators using the Huber loss, we employ a different strategy from the traditional route: constructing a deep ReLU network estimator that has a better empirical loss than the true function and the difference between these two functions furnishes a low bound. This step is related to the Huberization bias, yet more critically to the approximability of deep ReLU networks. As a result, we also contribute some new results on the approximation theory of deep ReLU neural networks. (Joint work with Yihong Gu and Wenxin Zhou)

Optimal Subgroup Selection
Speaker: **Richard Samworth, University of Cambridge**
In clinical trials and other applications, we often see regions of the feature space that appear to exhibit interesting behaviour, but it is unclear whether these observed phenomena are reflected at the population level. Focusing on a regression setting, we consider the subgroup selection challenge of identifying a region of the feature space on which the regression function exceeds a pre-determined threshold. We formulate the problem as one of constrained optimisation, where we seek a low-complexity, data-dependent selection set on which, with a guaranteed probability, the regression function is uniformly at least as large as the threshold; subject to this constraint, we would like the region to contain as much mass under the marginal feature distribution as possible. This leads to a natural notion of regret, and our main contribution is to determine the minimax optimal rate for this regret in both the sample size and the Type I error probability. The rate involves a delicate interplay between parameters that control the smoothness of the regression function, as well as exponents that quantify the extent to which the optimal selection set at the population level can be approximated by families of well-behaved subsets. Finally, we expand the scope of our previous results by illustrating how they may be generalised to a treatment and control setting, where interest lies in the heterogeneous treatment effect.

Rank-constrained Least-squares: Prediction and Inference
Speaker: **Ziwei Zhu, University of Michigan**
In this work, we focus on the high-dimensional trace regression model with a low-rank coefficient matrix. We establish a nearly optimal in-sample prediction risk bound for the rank-constrained least-squares estimator under no assumptions on the design matrix. Lying at the heart of the proof is a covering number bound for the family of projection operators corresponding to the subspaces spanned by the design. By leveraging this complexity result, we perform a power analysis for a permutation test on the existence of a low-rank signal under the high-dimensional trace regression model. We show that the permutation test based on the rank-constrained least-squares estimator achieves non-trivial power with no assumptions on the minimum (restricted) eigenvalue of the covariance matrix of the design. Finally, we use alternating minimization to approximately solve the rank-constrained least-squares problem to evaluate its empirical in-sample prediction risk and power of the resulting permutation test in our numerical study.

Asymptotic Properties of High-Dimensional Random Forests
Speaker: **Yingying Fan, University of Southern California**
As a flexible nonparametric learning tool, random forests algorithm has been widely applied to various real applications with appealing empirical performance, even in the presence of high-dimensional feature space. Unveiling the underlying mechanisms has led to some important recent theoretical results on the consistency of the random forests algorithm and its variants. However, to

our knowledge, all existing works concerning random forests consistency in high dimensional setting were established for various modified random forests models where the splitting rules are independent of the response. In light of this, in this paper we derive the consistency rates for the random forests algorithm associated with the sample CART splitting criterion, which is the one used in the original version of the algorithm (Breiman2001), in a general high-dimensional nonparametric regression setting through a bias-variance decomposition analysis. Our new theoretical results show that random forests can indeed adapt to high dimensionality and allow for discontinuous regression function. Our bias analysis characterizes explicitly how the random forests bias depends on the sample size, tree height, and column subsampling parameter. Some limitations on our current results are also discussed.

IS18: Room: G07	**Quantum Computing and Statistics**	**Chair: Yazhen Wang**

Bell's Theorem as a No-go Result in Classical Distributed Monte-Carlo Simulation
Speaker: **Richard Gill, Leiden University**
It has long been realized that the mathematical core of Bell's theorem is essentially a classical probabilistic proof that a certain distributed computing task is impossible: namely, the Monte Carlo simulation of certain iconic quantum correlations. I will present a new and simple proof of the theorem using Fourier methods (time series analysis) which should appeal to probabilists and statisticians. I call it Gull's theorem since it was sketched in a conference talk many years ago by astrophysicist Steve Gull, but never published. Indeed, there was a gap in the proof.

The connection with the topic of this session is the following: though a useful quantum computer is perhaps still a dream, many believe that a useful quantum internet is very close indeed. The first application will be: creating shared secret random cryptographic keys which, due to the laws of physics, cannot possibly be known to any other agent. So-called loophole-free Bell experiments have already been used for this purpose.

Like other proofs of Bell's theorem, the proof concerns a thought experiment, and the thought experiment could also in principle be carried out in the lab. This connects to the concept of functional Bell inequalities, whose application in the quantum research lab has not yet been explored. This is again a task for classical statisticians to explore.

Preprint: Gull's theorem revisited https://arxiv.org/abs/2012.00719

Statistical Aspects of the Quantum Supremacy Demonstration
Speaker: **Yosef Rinott, Hebrew University of Jerusalem**
In quantum computing, a demonstration of quantum supremacy consists of presenting a task, possibly of no practical value, whose computation is feasible on a quantum device, but cannot be performed by classical computers in any feasible amount of time. Google's (2019) notable claim of quantum supremacy consists of demonstrating the ability of a quantum circuit to generate, albeit with considerable noise, bitstrings from a distribution that is considered hard to simulate on classical computers. Verifying that the generated data is indeed from the claimed distribution and assessing the circuit's noise level and its fidelity is a statistical undertaking. In this talk, I will try to provide the minimal required background on quantum computing, and then to explain the statistical problems involved in the supremacy demonstration including modeling and estimation. I will not address the question of whether indeed supremacy was demonstrated.

Quantum Algorithms for Nonconvex Optimization: Escaping from Saddle Points and Beyond
Speaker: **Tongyang Li, Peking University**
The theories of optimization answer foundational questions in statistics, machine learning, etc., and lead to new algorithms for practical applications. In this talk, I will introduce quantum algorithms for escaping from saddle points. In particular, given a function $f: R^n \to R$, our quantum algorithm outputs an ε-approximate second-order stationary point using $\tilde{O}(\log n/\varepsilon^{1.75})$ queries to the quantum evaluation oracle (i.e., the zeroth-order oracle). Compared to the classical algorithm by Jin et al. with $\tilde{O}(\log^6(n)/\varepsilon^{1.75})$ queries to the gradient oracle (i.e., the first-order oracle), our quantum algorithm is polynomially better in terms of $\log n$ and matches its complexity in terms of $1/\varepsilon$. Technically, our main contribution is the idea of replacing the classical perturbations in gradient descent methods by simulating quantum wave equations, which constitutes the polynomial speedup in $\log n$ for escaping from saddle points. In the end, we will also briefly introduce our recent development in quantum algorithms for nonconvex optimization in general.

Quantum Annealing via Path-Integral Monte Carlo with Data Augmentation
Speaker: **Jianchang Hu, University of Wisconsin-Madison**
This talk considers quantum annealing in the Ising framework for solving combinatorial optimization problems. The path-integral Monte Carlo simulation approach is often used to approximate quantum annealing and implement the approximation by classical computers, which refers to simulated quantum annealing. In this talk we introduce a data augmentation scheme into simulated quantum annealing and develop a new algorithm for its implementation. The proposed algorithm reveals new insights on the sampling behaviors in simulated quantum annealing. Theoretical analyses are established to justify the algorithm, and numerical studies are conducted to check its performance and to confirm the theoretical findings.

IS20: Room: G26	**Inference Methods for Adaptively Collected Data**	**Chair: Susan Murphy**

Treatment Allocation with Strategic Agents
Speaker: **[Evan Munro]MunroE., Stanford University**

There is increasing interest in allocating treatments based on observed individual characteristics: examples include targeted marketing, individualized credit offers, and heterogenous pricing. Treatment personalization introduces incentives for individuals to modify their behavior to obtain a better treatment. This shifts the distribution of covariates, which means the Conditional Average Treatment Effect (CATE) now depends on how treatments are targeted. The optimal rule without strategic behavior allocates treatments only to those with a positive CATE. With strategic behavior, we show that the optimal rule can involve randomization, allocating treatments with less than 100% probability even to those with a positive CATE induced by that rule. We propose a sequential experiment based on Bayesian Optimization that learns the optimal treatment rule without parametric assumptions on individual strategic behavior.

Post-Contextual-Bandit Inference
Speaker: **Nathan Kallus, Cornell University**
Experiments using contextual bandits are increasingly replacing non-adaptive A/B tests in e-commerce, healthcare, and policymaking. In these very same applications, nonetheless, we still want to do statistical inference to draw credible conclusions from the experiment, not just to control the regret. However, the interdependence and vanishing exploration of bandit-collected data breaks usual inferential methods. While this has been addressed in non-contextual settings, the contextual setting poses unique challenges that we tackle for the first time in this paper. We propose the Contextual Adaptive Doubly Robust (CADR) estimator, which is guaranteed to be asymptotically normal, enabling simple Wald confidence intervals. The main technical challenge in constructing CADR is designing adaptive and consistent estimators of stabilization weights. Extensive numerical experiments using 57 OpenML datasets demonstrate that confidence intervals based on CADR uniquely provide correct coverage.

Near-optimal Inference in Adaptive Linear Regression
Speaker: **Koulik Khamaru, University of California, Berkeley**
When data is collected in an adaptive manner, even simple methods like ordinary least squares can exhibit non-normal asymptotic behavior. As an undesirable consequence, hypothesis tests and confidence intervals based on asymptotic normality can lead to erroneous results. We propose an online debiasing estimator to correct these distributional anomalies in least squares estimation. Our proposed method takes advantage of the covariance structure present in the dataset and provides sharper estimates in directions for which more information has accrued. We establish an asymptotic normality property for our proposed online debiasing estimator under mild conditions on the data collection process, and provide asymptotically exact confidence intervals. We additionally prove a minimax lower bound for the adaptive linear regression problem, thereby providing a baseline by which to compare estimators. There are various conditions under which our proposed estimator achieves the minimax lower bound up to logarithmic factors. We demonstrate the usefulness of our theory via applications to multi-armed bandit, autoregressive time series estimation, and active learning with exploration.

Statistical Inference After Adaptive Sampling in Non-Markovian Environments
Speaker: **Kelly Zhang, Harvard University**
There is a great desire to use adaptive sampling methods, such as reinforcement learning (RL) and bandit algorithms, for the real-time personalization of interventions in digital applications like mobile health and education. A major obstacle preventing more widespread use of such algorithms in practice is the lack of assurance that the resulting adaptively collected data can be used to reliably answer inferential questions, including questions about time-varying causal effects. Current methods for statistical inference on such data are insufficient because they (a) make strong assumptions regarding the environment dynamics, e.g., assume a contextual bandit or Markovian environment, or (b) require data to be collected with one adaptive sampling algorithm per user, which excludes data collected by algorithms that learn to select actions by pooling the data of multiple users. In this work, we make initial progress by introducing the adaptive sandwich estimator to quantify uncertainty; this estimator (a) is valid even when user rewards and contexts are non-stationary and highly dependent over time, and (b) accommodates settings in which an online adaptive sampling algorithm learns using the data of all users. Furthermore, our inference method is robust to misspecification of the reward models used by the adaptive sampling algorithm. This work is motivated by our work designing experiments in which RL algorithms are used to select actions, yet reliable statistical inference is essential for conducting primary analyses after the trial is over.

IS30: Room: G11	**Prediction and Sampling with Deep Neural Networks**	**Chair: Jian Huang**

Probabilistic Forecasting with Conditional Generative Networks via Scoring Rule Minimisation
Speaker: **Ritabrata Dutta, University of Warwick**
Probabilistic forecasting consists of stating a probability distribution for a future outcome based on past observations. In meteorology, ensembles of physics-based numerical models are run to get such distribution. Usually, performance is evaluated with scoring rules, functions of the forecast distribution and the observed outcome. With some scoring rules, calibration and sharpness of the forecast can be assessed at the same time. In deep learning, generative neural networks parametrize distributions on high-dimensional spaces and easily allow sampling by transforming draws from a latent variable. Conditional generative networks additionally constrain the distribution on an input variable. In this manuscript, we perform probabilistic forecasting with conditional generative networks trained to minimize scoring rule values. In contrast to Generative Adversarial Networks (GANs), no discriminator is required and training is stable. We perform experiments on two chaotic models and a global dataset of weather observations; results are satisfactory and better calibrated than what achieved by GANs.

Inference and Learning in Infinite Dimensions: Insights from Optimal Transport
Speaker: **Tengyuan Liang, University of Chicago**
I plan to discuss two papers on the topic of Inference and Learning in Infinite Dimensions using insights from optimal transport theory. The first paper introduces a new simulation-based inference procedure to model and sample from multi-dimensional probability distributions given access to i.i.d. samples, circumventing the usual approaches of explicitly modeling the density function or designing Markov chain Monte Carlo. Motivated by the seminal work on distance and isomorphism between metric measure spaces by Memoli (2011) and Sturm (2012), we propose a new notion called the Reversible Gromov-Monge (RGM) distance and study how RGM can be used to design new transform samplers to perform simulation-based inference. Our RGM sampler can also estimate optimal alignments between two heterogeneous metric measure spaces $(\mathscr{X}, \mu, c_{\mathscr{X}})$ and $(\mathscr{Y}, \nu, c_{\mathscr{Y}})$ from empirical data sets, with estimated maps that approximately push forward one measure μ to the other ν, and vice versa. Analytic properties of the RGM distance are derived; statistical rate of convergence, representation, and optimization questions regarding the induced sampler are studied. Synthetic and real-world examples showcasing the effectiveness of the RGM sampler are also demonstrated. Motivated by robust dynamic resource allocation in operations research, the second paper studies the Online Learning to Transport (OLT) problem where the decision variable is a probability measure, an infinite-dimensional object. We draw connections between online learning, optimal transport, and partial differential equations through an insight called the minimal selection principle, originally studied in the Wasserstein gradient flow setting by Ambrosio-et-al. This allows us to extend the standard online learning framework to the infinite-dimensional setting seamlessly. Based on our framework, we derive a novel method called the minimal selection or exploration (MSoE) algorithm to solve OLT problems using mean-field approximation and discretization techniques. In the displacement convex setting, the main theoretical message underpinning our approach is that minimizing transport cost over time (via the minimal selection principle) ensures optimal cumulative regret upper bounds. On the algorithmic side, our MSoE algorithm applies beyond the displacement convex setting, making the mathematical theory of optimal transport practically relevant to non-convex settings common in dynamic resource allocation.

Bayesian Learning via Neural Schrödinger–Föllmer Flows
Speaker: **Francisco Vargas, Cambridge University**
In this work we explore a new framework for approximate Bayesian inference in large datasets based on stochastic control (i.e. Schrödinger bridges). We advocate stochastic control as a finite time and low variance alternative to popular steady-state methods such as stochastic gradient Langevin dynamics (SGLD). Furthermore, we discuss and adapt the existing theoretical guarantees of this framework and establish connections to already existing VI routines in SDE-based models. Part of the practical exploration will focus on limitations of these approaches and potential avenues for overcoming said pitfalls.

Causal Probabilistic Spatio-temporal Fusion Transformers in Two-sided Ride-Hailing Markets
Speaker: **Hongtu Zhu, University of North Carolina Chapel Hill**
Achieving accurate spatio-temporal predictions in large-scale systems is extremely valuable in many real-world applications, such as weather forecasts, retail forecasting, and urban traffic forecasting. So far, most existing methods for multi-horizon, multi-task and multi-target predictions select important predicting variables via their correlations with responses of interest, and thus it is highly possible that many forecasting models generated from those methods are not causal, leading to weak transfer and generalization. The aim of this paper is to develop a collaborative causal spatio-temporal fusion transformer, named CausalTrans, to establish the collaborative causal effects of predictors on multiple forecasting targets, such as supply and demand in ride-sharing platforms. Specifically, we integrate the *causal attention* with the Conditional Average Treatment Effect (CATE) estimation method in causal inference. Moreover, we propose a novel and fast multi-head attention evolved from Taylor expansion instead of *softmax*, reducing time complexity from $O(|\mathscr{V}|^2)$ to $O(|\mathscr{V}|)$, where $|\mathscr{V}|$ is the number of hexagonal regions in a city. We further design a spatial graph fusion mechanism to significantly reduce the parameters' scale. We conduct a wide range of experiments to demonstrate the interpretability of *causal attention*, the effectiveness of various model components, and the time efficiency of our `CausalTrans`. As shown in these experiments, our `CausalTrans` framework can achieve up to 15% error reduction compared with various baseline methods.

Thursday 30-06-2022 10:30-12:30 Invited Sessions

IP22: Room: G03 Reflecting diffusions, stochastic networks and applications Chair: Cristina Costantini

Domains of Attraction of Invariant Distributions of the Infinite Atlas Model
Speaker: **Amarjit Budhiraja, University of North Carolina at Chapel Hill**
The infinite Atlas model describes a countable system of competing Brownian particles where the lowest particle gets a unit upward drift and the rest evolve as standard Brownian motions. The stochastic process of gaps between the particles in the infinite Atlas model does not have a unique stationary distribution and in fact there is a one parameter family $\{p(a), a \geq 0\}$ of mutually singular stationary distributions.

We say that an initial distribution of gaps is in the weak domain of attraction of the stationary measure $p(a)$ if the time averaged laws of the stochastic process of the gaps, when initialized using that distribution, converge to $p(a)$ weakly in the large time limit. We provide general sufficient conditions on the initial gap distribution of the Atlas particles for it to lie in the weak domain of attraction of $p(a)$ for each a. Results on extremality and ergodicity of $p(a)$ will also be presented. This is joint work with Sayan Banerjee.

Diffusion Limits for Multi-class Processor Sharing Queues
Speaker: **Amber Puha, California State University San Marcos**
Consider a single server queue that serves a finite number heterogeneous job types according to the processor sharing service discipline. Measure valued processes that keeps track of the residual service times of all jobs in the system at any given time are a natural descriptor of the system state. Under appropriate asymptotic assumptions, including standard heavy traffic assumptions, we show that (suitably rescaled) measure valued processes corresponding to a sequence of such queues converge in distribution to certain measure valued diffusion processes. An important contribution of this work is to devise a new methodology for establishing state-space collapse via the use of a certain relative entropy functional.

Error Bounds for the Constrained Langevin Approximation for Density Dependent Markov Chains
Speaker: **Felipe Campos, University of California, San Diego**
The stochastic dynamics of chemical reaction networks are often modeled using continuous-time Markov chains. However, except in very special cases, these processes cannot be analysed exactly and their simulation can be computationally intensive. An approach to this problem is to consider a diffusion approximation. The Constrained Langevin Approximation (CLA) is a reflected diffusion approximation for stochastic chemical reaction networks proposed by Leite & Williams. In this work, we extend this approximation to (nearly) density dependent Markov chains, when the diffusion state space is one-dimensional. Then, we provide a bound for the error of the CLA in a strong approximation. Finally, we discuss some applications for chemical reaction networks and epidemic models, illustrating these with examples. This is joint work with Ruth Williams.

Reflecting Diffusions in Non-smooth Domians: Some New Uniqueness Results
Speaker: **Cristina Costantini, University of Chieti-Pescara**
Exhaustive existence and uniqueness results for Reflecting Brownian Motion are available only in a polyhedron with constant direction of reflection on each face and in a smooth cone with radially constant direction of reflection. Only partial results are available for reflecting diffusions in piecewise smooth domains with curved boundaries, although these situations come up in applications, first of all in stochastic networks. This talk will present some recent, published and unpublished, existence and uniqueness results for semimartingale reflecting diffusions in piecewise smooth domains, with a varying direction of reflection on each "face". The first part of the talk will be devoted to 2-dimensional domains. We obtain existence and uniqueness under general, easily verifiable, geometrically meaningful conditions. When the domain is a polygon with a constant direction of reflection on each side, our conditions coincide with Dai and Williams (1995) conditions, which are necessary for existence of Reflecting Brownian Motion. Moreover we allow for cusps and for situations where two "sides" meet smoothly but the direction of reflection is discontinuous. In the second part of the talk some results in higher dimension will be discussed.

IP23: Room: G04 Centennial of the Lindeberg Central Limit Theorem Chair: Adam Jakubowski

The Central Limit Theorem: Lindeberg and After
Speaker: **Rene Schilling, TU Dresden**
We review the genesis of the Lindeberg theorem and the contributions of Levy and Feller.

Asymptotic Normality in Banach Spaces via Lindeberg Method
Speaker: **Alfredas Račkauskas, Vilnius University**
The relation between weak convergence of probabilities on a smooth Banach space and uniform convergence over certain class of smooth functions is established. This leads to an extension of Lindeberg's proof of the central limit theorem in Banach space framework. As an application asymptotic normality for sums of Banach space random variables including triangular arrays and weighted linear processes is proved.

The Lindeberg Central Limit Theorem and the Principle of Conditioning
Speaker: **Adam Jakubowski, Nicolaus Copernicus University**
We recall the Principle of Conditioning - a method of derivation of limit theorems for dependent random variables. In particular, we show how this method works when applied to the Lindeberg Central Theorem. We also discuss several recent applications.

IP26: Room: G05	**Random Walks in Static and Dynamic Random Environments**	**Chair: Jonathon Peterson**

Random Walk on the Simple Symmetric Exclusion Process
Speaker: **Daniel Kious, University of Bath**
In a joint work with Marcelo R. Hilário and Augusto Teixeira, we investigate the long-term behavior of a random walker evolving on top of the simple symmetric exclusion process (SSEP) at equilibrium. At each jump, the random walker is subject to a drift that depends on whether it is sitting on top of a particle or a hole. The asymptotic behavior is expected to depend on the density in [0, 1] of the underlying SSEP. Our first result is a law of large numbers (LLN) for the random walker for all densities except for at most two values and + in [0, 1], where the speed (as a function fo the density) possibly jumps from, or to, 0. Second, we prove that, for any density corresponding to a non-zero speed regime, the fluctuations are diffusive and a Central Limit Theorem holds. Our main results extend to environments given by a family of independent simple symmetric random walks in equilibrium. We will mention a work in progress with Guillaume Conchon–Kerjan and Pierre-Francois Rodriguez in which we study the strict monotonicity of the speed in these models.

Biased Random Walk on Dynamical Percolation
Speaker: **Nina Gantert, Technical University Munich**
We study the biased random walk for dynamical percolation on the d-dimensional lattice.

We establish a law of large numbers and an invariance principle for the random walk using regeneration times. Moreover, we verify that the Einstein relation holds, and we investigate the speed of the walk as a function of the bias. While for $d = 1$ the speed is monotone increasing, we show that this fails in general dimension d.

Based on joint work (in progress) with Sebastian Andres, Dominik Schmid and Perla Sousi.

Random Conductance Models on Simple Point Processes
Speaker: **Alessandra Faggionato, Sapienza University of Rome**
We consider a generic random conductance model on a simple point process (i.e. a random walk with random jump rates admitting a reversible measure). The jump length can be unbounded. Under minimal assumptions, we show that the effective homogenized matrix equals both the diffusion matrix of the random walk and the infinite volume conductivity of the associated random resistor network. We then discuss some low-temperature behavior of the effective homogenized matrix for some random walks relevant in Physics.

Anomalous Fluctuations for Random Walks in Dynamic Random Environments with Resetting
Speaker: **Luca Avena, Leiden University**
We perturb the classical RWRE (Random Walk in Random Environment) model on the one dimensional lattice by resetting the underlying random medium along a given sequence of times, called cooling sequence. Depending on the regularity structure of this cooling sequence, this model may present strong homogenization features as for a standard Random Walk, or can lead to strong spatial trapping effects as those known for RWRE. A surprisingly rich palette of possible limit scenarios have been explored in a series of recent papers. In this talk we focus on recent fluctuations results and show how this model can give rise to non-diffusive scaling with various different limiting laws ranging from Gaussian to tempered stables and mixtures of them. This is joint work with Conrado da Costa and Jonathon Peterson.

IP29: Room: G07	**Geodesics in First-passage Percolation**	**Chair: Daniel Ahlberg**

Empirical Measures, Geodesic Lengths, and a Variational Formula in First-passage Percolation
Speaker: **Erik Bates, University of Wisconsin–Madison**
We consider the standard first-passage percolation model on Z^d, in which each edge is assigned an i.i.d. non-negative weight, and the passage time between any two points is the smallest total weight of a nearest-neighbor path between them. This induces a random “disordered” geometry on the lattice. Our primary interest is in the empirical measures of edge-weights observed along geodesics in this geometry, say from 0 to $[n\xi]$, where ξ is a fixed unit vector. For various dense families of edge-weight distributions, we prove that these measures converge weakly to a deterministic limit as n tends to infinity. The key tool is a new variational formula for the time constant. In this talk, I will explain this formula and discuss its implications for the convergence of both empirical measures and lengths of geodesics.

Transitions for Exceptional Times in Dynamical First-Passage Percolation
Speaker: **Jack Hanson, City University of New York**
Abstract: In first-passage percolation (FPP), we let (t_v) be i.i.d. nonnegative weights on the vertices of a graph and consider the induced weighted graph metric T. When $P(t_v = 0)$ is small, T grows linearly in the graph distance; when it is large, $T(x,y)$ is of

order one even when x and y have large graph distance. In between these is the critical regime in which $T(x,y)$ diverges with graph distance, but does so sublinearly. We study a dynamical version of critical FPP on the triangular lattice where vertices resample their weights according to independent rate-one Poisson processes. We show that for one class of critical distributions, there are exceptional times at which T grows atypically, and for another class, a.s. there are no such times (these two classes cover all but a small set of critical weight distributions). In the former case, we compute the Hausdorff and Minkowski dimensions of the exceptional set and show that they can be but need not be equal.

Stationary Coalescing Walks on the Lattice
Speaker: **Arjun Krishnan, University of Rochester**
We consider translation invariant measures on families of semi-infinite nearest-neighbor paths on the integer lattice that coalesce if they meet. We classify the collective behavior of these walks under mild assumptions: they either coalesce almost surely or form bi-infinite trajectories. Bi-infinite trajectories form measure-preserving dynamical systems, have a common asymptotic direction in 2d, and possess other nice entropic properties. Our theory applies, for example, to geodesics in first- and last-passage percolation, and particle trajectories of exclusion processes.

Joint work with J. Chaika

Superconcentration, Chaos and Multiple Valleys in First-Passage Percolation
Speaker: **Maria Deijfen, University of Stockholm**
We consider a dynamical version of first-passage percolation on the d-dimensional integer lattice with i.i.d. edge weights, where edge weights are resampled independently in time. Let T_n denote the passage time from the origin to the site n steps along the first coordinate axis at time $t=0$, and let $\mu(t)$ denote the expected overlap between the time minimising paths at time 0 and $t>0$. We show that a subdiffusive behaviour of T_n is equivalent to a chaotic behaviour of the time minimising paths, manifested in that $\mu(t)=o(n)$. Known bounds for $Var(T_n)$ thus imply that indeed $\mu(t)=o(n)$ for $t>0$. As a consequence we show that there are many almost disjoint paths with almost optimal passage time. This gives evidence that earlier work by Sourav Chatterjee for certain Gaussian disordered systems reflects a more general principle.

IS8: Room: G22	**Selective Inference**	**Chair: Lucy Gao**

Multiple Testing of Partial Conjunction Null Hypotheses, with application to Replicability Analysis of High-dimensional Studies
Speaker: **Thorsten Dickhaus, University of Bremen**
The partial conjunction null hypothesis is tested in order to discover a signal that is present in multiple studies. The standard approach of carrying out a multiple test procedure on the partial conjunction p-values can be extremely conservative. We suggest alleviating this conservativeness, by eliminating many of the conservative partial conjunction p-values prior to the application of a multiple test procedure. This leads to the following two-step procedure: First, select the set with partial conjunction p-values below a selection threshold; second, within the selected set only, apply a family-wise error rate or false discovery rate controlling procedure on the conditional partial conjunction p-values, where we condition on the selection event. We analyze the proposed methodology theoretically, and we compare it with other recent approaches by means of computer simulations as well as by analyzing real data. This is joint work with Ruth Heller and Anh-Tuan Hoang.

Is the Best Good Enough? Goodness-of-fit After Selection
Speaker: **Joshua Loftus, London School of Economics and Political Science**
Goodness-of-fit tests after model selection are typically biased away from significance: when a data-driven selection process chooses an incorrect model the test will not have power to detect the lack of fit. In this talk we demonstrate a solution strategy using conditional post-selection inference on examples including the F-test as a goodness-of-fit test for high-dimensional linear models and tests for independent effects after thresholding selection rules. We study the conditional power of the tests before and after correction and show improved power to detect lack of fit. The literature on goodness-of-fit in high-dimensional regression is fairly sparse, but we conclude by comparing our method to one competitor and consider their complementary strengths.

Selective Inference for Hierarchical Clustering
Speaker: **Lucy Gao, University of Waterloo**
Testing for a difference in means between two groups is fundamental to answering research questions across virtually every scientific area. Classical tests control the type I error rate when the groups are defined a priori. However, if the groups are instead defined using a clustering algorithm, then applying a classical test yields an extremely inflated type I error rate. Surprisingly, this problem persists even if two separate and independent data sets are used for clustering and for hypothesis testing. In this talk, I will propose a test for a difference in means between two estimated clusters that accounts for the fact that the null hypothesis is a function of the data, using a selective inference framework. Then, I will describe how to efficiently compute exact p-values for clusters obtained using hierarchical clustering. I will also show an application in the context of single-cell RNA-sequencing data, where it is common for researchers to cluster the cells, then test for a difference in mean gene expression between the clusters. This talk is based on joint work with Jacob Bien (University of Southern California) and Daniela Witten (University of Washington).

A (Tight) Upper Bound for the Length of Confidence Intervals with Conditional Coverage
Speaker: **Hannes Leeb, University of Vienna**
We show that two popular selective inference procedures, namely data carving and selection with a randomized response, when combined with the polyhedral method, result in confidence intervals whose length is bounded. This contrasts results for confidence intervals based on the polyhedral method alone, whose expected length is typically infinite. Moreover, we show that these two procedures always dominate corresponding sample-splitting methods in terms of interval length.

IS10: Room: G16	**Spatial and Spatial-Temporal Modelling**	**Chair: Hsin-Cheng Huang**

Accelerated Spatio-Temporal Changes in Arctic Sea Ice from Remote Sensing Data
Speaker: **Noel Cressie, University of Wollongong** and **Bohai Zhang, Nankai University**
Catastrophic climate change comes in different forms. The cryosphere is Earth's air conditioner, but an ever-smaller Arctic sea-ice extent (SIE) sets up a natural feedback loop where more dark ocean in the region absorbs more energy and turns up the thermostat. A spatio-temporal analysis of forty years of remotely sensed Arctic ice/water (i.e., binary) data at a 25 km resolution shows increasing zonal variability and then collapse at lower latitudes. Moreover, there is an indication that the same is underway at higher latitudes. Our talk gives visualisations of these changes through maps, time series of box plots broken down by selected latitude bands, and a plot showing SPOT (spatial proportion over threshold) functions at five-year intervals. These exploratory analyses are followed by fitting a dynamic spatio-temporal logistic autoregressive model that allows forecasting of the following year's Arctic ice/water distribution. Based on a bootstrap from the fitted model, it is inferred through the SPOT function and other summaries that SIE across the decades is in decline and that the decline accelerating.

Asynchronous Change-point Estimation for Spatially Correlated Functional Time Series
Speaker: **Bo Li, Univeristy of Illinois Urbana Champaign**
We propose a new solution under the Bayesian framework to simultaneously estimate mean based asynchronous change-points in spatially correlated functional time series. Unlike previous methods that assume a shared change-point at all spatial locations or ignore spatial correlation, our method treats change-points as a spatial process. This allows our model to respect spatial heterogeneity and exploit spatial correlations to improve estimation. Our method is derived from the ubiquitous cumulative sum (CUSUM) statistic that dominates change-point detection in functional time series. However, instead of directly searching for the maximum of the CUSUM based processes, we build spatially correlated two-piece linear models with appropriate variance structure to locate all change-points at once. The proposed linear model approach increases the robustness of our method to variability in the CUSUM process, which, combined with our spatial correlation model, improves change-point estimation near the edges. We demonstrate through extensive simulation studies that our method outperforms existing functional change-point estimators in terms of both estimation accuracy and uncertainty quantification, under either weak and strong spatial correlation, and weak and strong change signals. Finally, we demonstrate our method using a temperature data set and a coronavirus disease 2019 (COVID-19) study."

Composite Likelihood Inference for Ordinal Periodontal Data with Replicated Spatial Patterns
Speaker: **Jun Zhu, University of Wisconsin–Madison**
Spatial ordinal data observed separately for multiple subjects are common in biomedical research, yet statistical methodology for such ordinal data analysis is limited. The existing methodology often assumes a single realization of spatial ordinal data without replications and thus are not directly applicable. Motivated by a dataset evaluating periodontal disease (PD) status, we propose a multisubject spatial ordinal model that assumes a geostatistical spatial structure within a regression framework through a latent variable representation. For achieving computational scalability within a classical inferential framework, we develop a maximum composite likelihood method for parameter estimation, and establish the asymptotic properties of the parameter estimates. We also develop model diagnostic measures for the dependent data scenario using generalized surrogate residuals. A simulation study suggests sound finite sample properties of the proposed methods. We illustrate our proposed methodology by the motivating PD dataset. A companion R package clordr is available for implementation.

IS13: Room: G26	**Adaptive Learning**	**Chair: Alexandra Carpentier**

Minimax Non-Parametric Testing: Local Adaptivity, Conditional Independence, and Calibration via Permutation
Speaker: **Sivaraman Balakrishnan, Carnegie Mellon University**
I will discuss a sequence of recent works which aim to better understand minimax rates for non-parametric density testing problems (testing goodness-of-fit, two-sample testing and conditional independence testing) in each case focusing on a different aspect of the non-parametric testing problem. For goodness-of-fit testing, I will discuss locally-adaptive tests which adapt to the intrinsic problem difficulty. For two-sample testing, I will discuss the permutation method, focusing in particular on its minimax power. Finally, for conditional independence testing, I will focus on smoothness assumptions that enable tractable CI testing, and also briefly discuss the relevance of these smoothness for calibration via permutation. This talk is based on joint works with: Ilmun Kim, Matey Neykov and Larry Wasserman.

Fundamental Limits for Learning Hidden Markov Model
Speaker: **Elisabeth Gassiat, Université Paris-Saclay**

We study the frontier between learnable and unlearnable hidden Markov models (HMMs). HMMs are flexible tools for clustering dependent data coming from unknown populations. The model parameters are known to be fully identifiable (up to label-switching) without any modeling assumption on the distributions of the populations as soon as the clusters are distinct and the hidden chain is ergodic with a full rank transition matrix. In the limit as any one of these conditions fails, it becomes impossible in general to identify parameters. For a chain with two hidden states we prove non-asymptotic minimax upper and lower bounds, matching up to constants, which exhibit thresholds at which the parameters become learnable. We also provide an upper bound on the relative entropy rate for parameters in a neighbourhood of the unlearnable region which may have interest in itself.

Tracking Most Significant Arm Switches in Bandits
Speaker: **Samory Kpotufe, Columbia University**
In bandit with distribution shifts, one aims to automatically adapt to unknown changes in reward distribution, and restart exploration when necessary. While this problem has been studied for many years, a recent breakthrough of Auer et al. (2018, 2019) provides the first adaptive procedure to guarantee an optimal (dynamic) regret $(LT)^{1/2}$, for T rounds and L changes.

However, while this rate is tight in the worst case, it leaves open whether faster rates are possible, without prior knowledge, if few changes in distribution are actually “severe". To resolve this question, we propose a new notion of significant shift, which only counts very severe changes that clearly necessitate a restart: roughly, these are changes involving only the most severe best arm switches, with large aggregate differences in reward overtime. Thus, our resulting procedure is first to adaptively achieve rates always faster (sometimes significantly) than $(ST)^{1/2}$, where S only counts best arm switches, and at the same time achieves optimal regret when expressed in terms of total variation (which aggregates differences overtime). Our results are expressed in enough generality to also capture non-stochastic adversarial settings.

Optimal Ranking in Crowd-Sourcing Problems
Speaker: **Nicolas Verzelen, INRAE**
Consider a crowd-sourcing problem where we have n experts and d tasks. The average ability of each expert for each task is stored in an unknown matrix M, from which we have incomplete and noise observations. We make no (semi-) parametric assumptions, but assume that both experts and tasks can be perfectly ordered: so that if an expert A is better than an expert B, the ability of A is higher than that of B for all tasks - and that the same holds for the tasks. This implies that if the matrix M, up to permutations of its rows and columns, is bi-isotonic. We focus on the problem of recovering the optimal ranking of the experts in l_2 norm, when the ordering of the tasks is known to the statistician. In other words, we aim at estimating the suitable permutation of the rows of M while the permutation of the columns is known. We provide a minimax-optimal and computationally feasible method for this problem, based on hierarchical clustering, PCA, change-point detection, and exchange of information among the clusters. We prove in particular - in the case where $d > n$ - that the problem of estimating the expert ranking is significantly easier than the problem of estimating the matrix M. This talk is based on a joint ongoing work with Alexandra Carpentier and Emmanuel Pilliat.

IS25: Chancellors Hall	**Conformal Prediction, Semiparametric Statistics, and Causal Inference**	**Chair: Arun Kumar Kuchibhotla**

Conformal Inference of Counterfactuals and Individual Treatment Effects
Speaker: **Lihua Lei, Stanford University**
Evaluating treatment effect heterogeneity widely informs treatment decision-making. At the moment, much emphasis is placed on the estimation of the conditional average treatment effect via flexible machine learning algorithms. While these methods enjoy some theoretical appeal in terms of consistency and convergence rates, they generally perform poorly in terms of uncertainty quantification. This is troubling since assessing risk is crucial for reliable decision-making in sensitive and uncertain environments. In this work, we propose a conformal inference-based approach that can produce reliable interval estimates for counterfactuals and individual treatment effects under the potential outcome framework. For completely randomized or stratified randomized experiments with perfect compliance, the intervals have guaranteed average coverage in finite samples regardless of the unknown data generating mechanism. For randomized experiments with ignorable compliance and general observational studies obeying the strong ignorability assumption, the intervals satisfy a doubly robust property which states the following: the average coverage is approximately controlled if either the propensity score or the conditional quantiles of potential outcomes can be estimated accurately. Numerical studies on both synthetic and real datasets empirically demonstrate that existing methods suffer from a significant coverage deficit even in simple models. In contrast, our methods achieve the desired coverage with reasonably short intervals.

Finite-Sample Efficient Conformal Prediction
Speaker: **Yachong Yang, University of Pennsylvania**
Conformal prediction is a generic methodology for finite-sample valid distribution-free prediction. This technique has garnered a lot of attention in the literature partly because it can be applied with any machine learning algorithm that provides point predictions to yield valid prediction regions. Of course, the efficiency (width/volume) of the resulting prediction region depends on the performance of the machine learning algorithm. In this paper, we consider the problem of obtaining the smallest conformal prediction region given a family of machine learning algorithms. We provide two general-purpose selection algorithms and consider coverage as well as width properties of the final prediction region. The first selection method yields the smallest width prediction region among the family of conformal prediction regions for all sample sizes, but only has an approximate coverage guarantee. The second selection method has a finite sample coverage guarantee but only attains close to the smallest width. The approximate optimal width

property of the second method is quantified via an oracle inequality. Asymptotic oracle inequalities are also considered when the family of algorithms is given by ridge regression with different penalty parameters.

Overparameterized Ridgeless Regression: Can LOOCV and GCV still Estimate Out-of-Sample Test Error when the Training Error is Zero?

Speaker: **Ryan Tibshirani, Carnegie Mellon University**

We examine generalized and leave-one-out cross-validation for ridge regression in a proportional asymptotic framework where the dimension of the feature space grows proportionally with the number of observations. Given i.i.d. samples from a linear model with an arbitrary feature covariance and a signal vector that is bounded in L2 norm, we show that generalized cross-validation for ridge regression converges almost surely to the expected out-of-sample prediction error, uniformly over a range of ridge regularization parameters that includes zero (and even negative values). We prove the analogous result for leave-one-out cross-validation. As a consequence, we show that ridge tuning via minimization of generalized or leave-one-out cross-validation asymptotically almost surely delivers the optimal level of regularization for predictive accuracy, whether it be positive, negative, or zero.

Thursday 30-06-2022	13:30-15:30	Invited Sessions

IP24: Room: G03 Quantum Field Theory and Stochastic Analysis Chair: Hao Shen

Large N Limit of 2D Pure Gauge Theories

Speaker: **Antoine Dahlqvist, University of Sussex**

Quantum Gauge Field Theories are motivated by the standard model introduced in physics to describe elementary particles and their interactions. Space-time is modelled by a manifold, each type of interaction has a symmetry given by a compact Lie group. Since the 70's, it was realised in physics that considering symmetry groups of large size leads to simplified but non-trivial theories, called master fields. The 2D Euclidean Yang-Mills measure is one of the only well defined continuous probability model for a pure gauge theory. We consider a unitary group of size N as symmetry group and fix a 2D manifold for space-time. The interaction field is described by a random process valued in unitary matrices and indexed by loops of the manifold. We shall review some recent progress about the limit of this process as N goes to infinity and discuss how it depends on the topology of the manifold.

Grassmannian Stochastic Analysis and the Stochastic Quantization of Euclidean Fermions

Speaker: **Francesco de Vecchi, University of Bonn**

In this talk we propose a stochastic analysis of Grassmann random variables suitable for the stochastic quantization of Euclidean fermionic quantum field theories. Analysis on Grassmann algebras is developed from the point of view of quantum probability, i.e. using a suitable C-algebra endowed with a positive state. We exploit this setting for building fermionic quantum field theories as invariant measures of suitable Langevin equations driven by a Grassmannian space-time white noise. We apply this method to the stochastic quantization and to the removal of the space cut-off of the (regularized) Euclidean Yukawa model. The talk is based on the joint work with Sergio Albeverio, Luigi Borasi and Massimiliano Gubinelli.

Log-Sobolev Inequalities and Renormalisation

Speaker: **Benoit Dagallier, University of Cambridge**

Recently, Roland Bauerschmidt and Thierry Bodineau combined the Bakry-Emery theory with a renormalisation group approach in the spirit of Wilson and Polchinski. This combination allows one to derive log-Sobolev inequalities for various field theories, as well as for critical Ising models in dimension $d \geq 5$. I will introduce the main ideas in the setting of the φ^4 Euclidean field theory ($d = 2, 3$), for which Roland Bauerschmidt and I recently proved the log-Sobolev inequality.

Stochastic Quantisation of Gauge Theories

Speaker: **Ajay Chandra, Imperial College London**

In this talk I will present some recent progress on the construction of quantum gauge theories using ideas from stochastic PDEs. I will introduce and motivate the stochastic Yang–Mills heat equation and give an overview of how one can prove gauge covariance of the dynamics and also build a state space of "connections modulo gauge equivalence" that is compatible with the dynamic. This is joint work with Ilya Chevyrev, Martin Hairer, and Hao Shen.

IP25: Room: G05 High-dimensional Extremes and their Applications Chair: Gennady Samorodnitsky

Tail Inverse Regression for Dimension Reduction with Extreme Response

Speaker: **Anne Sabourin, Telecom Paris, Institut Polytechnique de Paris**

We consider the problem of dimensionality reduction for prediction of a real valued target Y to be explained by a covariate vector X of dimension p, with a particular focus on extreme values of Y which are of particular concern for risk management.

The general purpose is to reduce the dimensionality of the statistical problem through an orthogonal projection on a lower dimensional subspace of the covariate space. Inspired by the sliced inverse regression (SIR) methods, we develop a novel framework (TIREX, Tail Inverse Regression for EXtreme response) relying on an appropriate notion of tail conditional independence in order to estimate an extreme sufficient dimension reduction (SDR) space of potentially smaller dimension than that of a classical SDR space.

We prove the weak convergence of tail empirical processes involved in the estimation procedure and we illustrate the relevance of the proposed approach on simulated and real world data.

Authors: Anass Aghbalou, François Portier, Anne Sabourin, Chen Zhou. Working paper arXiv:2108.01432.

Distributed inference in extreme value statistics

Speaker: **Chen Zhou, Erasmus Universiteit Rotterdam**

The availability of massive datasets allows for conducting extreme value statistics using more observations drawn from the tail of an underlying distribution. When large datasests are distributedly stored and cannot be combined into one oracle sample, a divide-and-conquer algorithm is often invoked to construct a distributed estimator. If the distributed estimator possesses the same asymptotic behavior as the hypothetical oracle estimator based on the oracle sample, then it is regarded as satisfying the oracle property. In a series of works, we introduce a set of tools for establishing the oracle property of most estimators in extreme value statistics. The tools are based on the (multivariate) tail empirical process and the tail quantile process.

Learning Extremal Graphical Structures in High Dimensions

Speaker: **Michaël Lalancette, University of Toronto**

Multiple characterizations and models exist for extremal dependence, the dependence structure of multivariate data in unobserved tail regions. However, statistical inference for extremal dependence uses merely a fraction of the available data, drastically reducing the effective sample size and creating challenges even in moderate dimension. Recently introduced graphical models for multivariate extremes allow for enforced sparsity in moderate- to high-dimensional settings, reducing the effective dimension. In this work, we propose a novel, scalable method for selection of extremal graphical models that makes no assumption on the underlying graph structure, as opposed to existing approaches. It exploits existing tools for Gaussian graphical model selection such as the graphical lasso and neighborhood selection. Model selection consistency is established in sparse regimes where the dimension is allowed to be exponentially larger than the effective sample size. Based on joint work with Sebastian Engelke and Stanislav Volgushev (https://arxiv.org/abs/2111.00840).

Spectral Learning of Multivariate Extremes

Speaker: **Richard Davis, Columbia University**

A spectral clustering algorithm for analyzing the dependence structure of multivariate extremes is proposed. More specifically, we focus on the asymptotic dependence of multivariate extremes characterized by the angular or spectral measure in extreme value theory. This work studies the theoretical performance of spectral clustering based on a random k-nearest neighbor graph constructed from an extremal sample, i.e., the angular part of random vectors for which the radius exceeds a large threshold. In particular, the asymptotic distribution of extremes arising from a linear factor model is derived and it is shown that, under certain conditions, spectral clustering can consistently identify the clusters of extremes arising in this model. Leveraging this result we propose a simple consistent estimation strategy for learning the angular measure. Our theoretical findings are complemented with numerical experiments illustrating the finite sample performance of the proposed methods. (This is joint work with Marco Avella Medina and Gennady Samorodnitsky.)

IP27: Room: G07	Stable Random Walks, Capacity and Fractional Diffusions Session in honour of Mark M. Meerschaert	Chair: Stilian Stoev

Attraction to and Repulsion from Patches on the Hypersphere and Hyperplane for Isotropic d-dimensional α-stable Processes with Index in $\alpha \in (0,1]$ and $d \leq 2$.

Speaker: **Andreas Kyprianou, University of Bath**

Consider a d-dimensional ?-stable processes with index in $\alpha \in (0,1)$ and $d \leq 2$. Suppose that D is a region of the unit sphere $S^{d-1} = \{x \in R^d : |x| = 1\}$. We construct the aforesaid stable Lévy process conditioned to approach Δ continuously, either from inside S^{d-1}, from outside S^{d-1} or in an oscillatory way; all of which have zero probability. Our approach also extends to the setting of hitting bounded domains of $(d-1)$-dimensional hyperplanes. We appeal to a mixture of methods, appealing to the modern theory of self-similar Markov process as well as the classical potential analytic view. Joint work with Tsogzolmaa Saizmaa (National University of Mongolia), Sandra Palau (UNAM, Mexico) and Mateusz Kwasniki (Technical University of Wroclaw).

Calculating Alpha-capacity via Stable Random Walks

Speaker: **John Nolan, American University**

An approach is given for computing alpha-capacity of complicated objects in d-dimensons through the use of simulated continuous time random walks with isometric stable increments. A method we call Walk-In-Out-Balls is described to simulate points from the equilibrium measure of a set K. These points are then used to estimate the alpha-capacity.

A Study on the Stability of the Mean Value Property for the Fractional Laplacian

Speaker: **Claudia Bucur, University of Insubria**

As well known, harmonic functions satisfy the mean value property, namely the average of a harmonic function over any ball equals the functions value at the center of the ball. Is this property stable for sets, meaning that if the mean value property holds on a set for all harmonic functions, is this set necessarily a ball? This question was investigated by several authors, and was elegantly settled by Ülkü Kuran. More recently, Giovanni Cupini, Nicola Fusco, Ermanno Lanconelli and Xiao Zhong proved a quantitative stability result for the mean value formula, showing that a suitable “mean value gap“ is bounded from below by the “gap“between the set and a ball. In this talk, we present some results on the non local counterparts of these results. In particular we establish that if the value of any fractional harmonic function at a given point equals a suitable fractional mean value on a domain, that domain is the ball centered at the given point. We then focus on some quantitative versions of this question in terms of different possible “gap“ functions. Differently from the classical case, some of our arguments rely on purely non local properties that do no have classical counterparts, such as the fact that “all functions are locally fractional harmonic up to a small error“

On Coupled Continuous Time Random Walks

Speaker: **Hans-Peter Scheffler, University of Siegen**

We analyze the limiting behavior of coupled continuous time random walks and related processes. Moreover, we present formulas for the distribution of the those processes and their governing equations.

IP28: Room: G04 Vector- and Function-valued Random Fields: models, Structure and Regularity Chair: Tailen Hsing

Vector-valued Random Fields for Color Images
Speaker: **Hermine Bierme, University of Tours**
The RGB color model is an additive color model in which the red, green and blue colors are added to reproduce colors. We consider the color monogenic framework to define random fields according to natural colors. This approach is based on the monogenic representation of a greyscale image, based on Riesz transform and particularly well-adapted to detect directionality of self-similar of operator scaling Gaussian fields. This talk is dedicated to the memory of Mark Marvin Meerschaert (1955-2020).

Wavelet Random Matrices and High-dimensional Eigenanalysis
Speaker: **Gustavo Didier, Tulane University**
In this talk, we combine two mathematical frameworks that are rarely considered jointly: (i) high dimensional probability theory; (ii) fractal analysis. This is done by bringing together the study of large random matrices and wavelet-domain scaling analysis. We propose to characterize low-frequency behavior in high-dimensional systems by means of a random matrix framework involving the three-way limit as the sample size, the dimension and the scale go to infinity. In this framework, we show that the scaling laws governing the behavior of scale-invariant, non-Markovian systems eventually emerge in the eigenspectrum of large wavelet random matrices.

Function-valued Random Fields: Tangents, Intrinsic Stationarity, and Self-similarity
Speaker: **Stilian Stoev, University of Michigan**
We study random fields taking values in a separable Hilbert space H. First, we focus on their local structure and establish a counterpart to Falconer's characterization of tangent fields. That is, we show (under general conditions) that the tangent fields to a H-valued process are self-similar and almost all of them have stationary increments. We go a bit further and study higher-order tangent fields. This leads naturally to the study of self-similar intrinsic random functions (IRF) taking values in a Hilbert space. To this end, we begin by extending Matheron's theory of scalar-valued IRFs and provide the spectral representation of H-valued IRFs. We then use this theory to characterize large classes of operator self-similar H-valued IRF processes, which in the Gaussian case can be viewed as the H-valued counterparts to fractional Brownian fields. These general results may find applications to the study of long-range dependence for random fields taking values in a Hilbert space as well as to modeling function-valued spatial data.

Sample Path and Extreme Value Properties of Multivariate Gaussian Random Fields
Speaker: **Yimin Xiao, Michigan State University**
In this talk, we present some recent results on sample path and extreme value properties of a large class of multivariate Gaussian random fields including multivariate Matérn Gaussian fields, operator fractional Brownian motion, vector-valued operator-scaling random fields, and matrix-valued Gaussian random fields. These results illustrate explicitly the effects of the dependence structures among the coordinate processes on the sample path and extreme value properties of multivariate Gaussian random fields.

IS7: Room: G22 Causal Inference Chair: Rajen Shah

The Role of Principal Stratification in Pharmaceutical Experimentation: the Estimand Strategy
Speaker: **Fabrizia Mealli, University of Florence**
The ICH E9(R1) Addendum has provided a framework for discussing interesting estimands that may be of relevance in RCTs affected by the intercurrent events of various types. The Addendum is of fundamental importance because it provides principles regarding the analysis of RCTs and observational studies with intercurrent events. One of the strategy that is proposed is the Principal Stratum strategy. Here, we will discuss issues regarding the application of such strategy, offering guidelines regarding which data should be collected at baseline and post-treatment to support assumptions and analysis. Various approaches to Principal Stratification analysis will be reviewed with associated carefully planned sensitivity analysis to assess robustness of conclusions to deviations from such assumptions, as well as to choice of an analytic approach more broadly (e.g., parametric vs semiparametric or non parametric). Examples will be used to illustrate concepts.

Towards deriving graphical rules for efficient in causal graphical models
Speaker: **Andrea Rotnitzky, Universidad Torcuato Di Tella and Harvard University**
Causal graphs are responsible in great part for the explosive growth and adoption of causal inference in modern epidemiology and medical research. This is so because graphical models facilitate encoding and communicating causal assumptions and reasoning in an intuitive fashion, requiring minimal, if any at all, mathematical dexterity. Applying simple graphical rules, it is possible to easily communicate biases due to confounding and selection, which data are needed to correct for these biases, and to derive which statistical target parameter quantifies a causal effect of interest. Yet, causal graphical models encode a well defined statistical semiparametric model and little, if any, work has been done to investigate and derive simple graphical rules to encode efficiency in estimation. In this talk, I will present work towards bridging this gap. Given a causal graphical model, I will derive a set of graphical rules for determining the optimal adjustment set in a point exposure problem. This is the subset of variables in the graph that both suffices to control for confounding under the model and yields a non-parametric estimator of the population average treatment effect (ATE) of a dynamic, i.e. personalized, or static point exposure on an outcome with smallest asymptotic variance under any law in the model. I will then discuss the conditions for existence of a universally optimal adjustment set when the search is restricted to

the realistic scenario in which only a subset of the graph variables is observable. For the problem of estimating the effect of a time dependent treatment I will discuss an impossibility result. Finally, I will describe graphical rules for constructing a reduced graph whose nodes represent only those variables that are informative for ATE and such that efficient estimation of ATE under the reduced graph and under the original graph agree.

Stochastic Causal Programming for Bounding Treatment Effects
Speaker: **Ricardo Silva, University College London**
Causal effect estimation is important for numerous tasks in the natural and social sciences. However, identifying effects is impossible from observational data without making strong, often untestable assumptions. We consider algorithms for the partial identification problem, bounding treatment effects from multivariate, continuous treatments over multiple possible causal models when unmeasured confounding makes identification impossible. We consider a framework where observable evidence is matched to the implications of constraints encoded in a causal model by norm-based criteria. This generalizes classical approaches based purely on generative models. Casting causal effects as objective functions in a constrained optimization problem, we combine flexible learning algorithms with Monte Carlo methods to implement a family of solutions under the name of stochastic causal programming. In particular, we present ways by which such constrained optimization problems can be parameterized without likelihood functions for the causal or the observed data model, reducing the computational and statistical complexity of the task. (Joint work with Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner and Niki Kilbertus.)

Debiased Inverse Propensity Score Weighting for Estimation of Average Treatment Effects with High-Dimensional Confounders
Speaker: **Rajen Shah, University of Cambridge**
We consider estimation of average treatment effects given observational data with high-dimensional pretreatment variables. Existing methods for this problem typically assume some form of sparsity for the regression functions. In this talk, we will introduce a debiased inverse propensity score weighting (DIPW) scheme for average treatment effect estimation that delivers root n consistent estimates of the average treatment effect when the propensity score follows a sparse logistic regression model; the regression functions are permitted to be arbitrarily complex. Our theoretical results quantify the price to pay for permitting the regression functions to be unestimable, which rows up as an inflation of the variance of the estimator compared to the semiparametric efficient variance by at most $O(1)$ under mild conditions. Given the lack of assumptions on the regression functions, averages of transformed responses under each treatment may also be estimated at the root n rate, and so for example, the variances of the potential outcomes may be estimated. We show how confidence intervals centred on our estimates may be constructed, and also discuss an extension of the method to estimating projections of the heterogeneous treatment effect function.

IS15: Room: G26	**Statistical Machine Learning**	**Chair: Raaz Dwivedi**

Online Learning in the Real World: Delay, Optimism, and Subseasonal Forecasting
Speaker: **Genevieve Flaspohler, MIT**
Real-time climate and weather forecasting is a challenging prediction problem that online learning algorithms, which can provide regret guarantees against adversarial environments, are particularly well suited to address. However, real-world forecasting has many challenges: feedback is often delayed, forecasters must do well over short time intervals, and the underlying loss sequence is highly non-stationary. Inspired by these demands, this talk presents three new algorithms for real-world online learning—DORM, DORM+, and AdaHedgeD—that require no parameter tuning and have optimal regret guarantees under delayed feedback. These algorithms arise from a novel reduction of delayed online learning to optimistic online learning that reveals how optimistic hints can mitigate the regret penalty caused by delay. This delay-as-optimism perspective is paired with a new analysis of optimistic learning that exposes its robustness to hinting errors and a new meta-algorithm for learning effective hinting strategies in the presence of delay. Finally and importantly, the online learners have been applied to achieve state-of-the-art results in real-world forecasting. DORM+ is currently being adapted for use at NOAA's Climate Prediction Center. This talk will cover the theory and application of developing online learning with delay and optimism.

Gradient-Free Kernel Stein Discrepancy
Speaker: **Matthew Fisher, University of Newcastle**
Stein discrepancies have emerged as a powerful statistical tool, being applied to fundamental statistical problems including parameter inference, goodness-of-fit testing, and sampling. The canonical Stein discrepancies require the derivatives of a statistical model to be computed, and in return provide theoretical guarantees of convergence detection and control. However, for complex statistical models, the stable numerical computation of derivatives can require bespoke algorithmic development and render Stein discrepancies impractical. In this talk, we investigate a collection of non-canonical Stein discrepancies that are gradient free, meaning that derivatives of the statistical model are not required. Sufficient conditions for convergence detection and control will be discussed, and applications to sampling and variational inference will be presented.

Controlled Discovery and Localization of Signals via Bayesian Linear Programming (BLiP)
Speaker: **Lucas Janson, Harvard University**
In many statistical problems, it is necessary to simultaneously discover signals and localize them as precisely as possible. For in-

stance, genetic fine-mapping studies aim to discover causal genetic variants, but the strong local dependence structure of the genome makes it hard to identify the exact locations of those variants. So the statistical task is to output as many regions as possible and have those regions be as small as possible, while controlling how many outputted regions contain no signal. The same problem arises in any application where signals cannot be perfectly localized, such as locating stars in astronomical sky surveys and change point detection in time series data. However, there are two competing objectives: maximizing the number of discoveries and minimizing the size of those discoveries (all while controlling false discoveries), so our first contribution is to propose a single unified measure we call the resolution-adjusted power that formally trades off these two objectives and hence, in principle, can be maximized subject to a constraint on false discoveries. We take a Bayesian approach, but the resulting posterior optimization problem is intractable due to its non-convexity and high-dimensionality. Thus our second contribution is Bayesian Linear Programming (BLiP), a method which overcomes this intractability to jointly detect and localize signals in a way that verifiably nearly maximizes the expected resolution-adjusted power while provably controlling false discoveries. BLiP is very computationally efficient and can wrap around any Bayesian model and algorithm for approximating the posterior distribution over signal locations. Applying BLiP on top of existing state-of-the-art Bayesian analyses of UK Biobank data (for genetic fine-mapping) and the Sloan Digital Sky Survey (for astronomical point source detection) increased the resolution-adjusted power by 30-120% with just a few minutes of computation. BLiP is implemented in the new packages pyblip (Python) and blipr (R). This is joint work with Asher Spector.

Counterfactual Inference in Sequential Experimental Design
Speaker: **Raaz Dwivedi, UC Berkeley**
We consider the problem of counterfactual inference in sequentially designed experiments wherein a collection of N units each undergo a sequence of interventions for T time periods, based on policies that sequentially adapt over time. Our goal is counterfactual inference, i.e., estimate what would have happened if alternate policies were used, a problem that is inherently challenging due to the heterogeneity in the outcomes across units and time. To tackle this task, we introduce a suitable latent factor model where the potential outcomes are determined by exogenous unit and time level latent factors. Under suitable conditions, we show that it is possible to estimate the missing (potential) outcomes using a simple variant of nearest neighbors. First, assuming a bilinear latent factor model and allowing for an arbitrary adaptive sampling policy, we establish a distribution-free non-asymptotic guarantee for estimating the missing outcome of any unit at any time; under suitable regularity conditions, this guarantee implies that our estimator is consistent. Second, for a generic non-parametric latent factor model, we establish that the estimate for the missing outcome of any unit at time T satisfies a central limit theorem as T goes to infinity, under suitable regularity conditions. Finally, en route to establishing this central limit theorem, we prove a non-asymptotic mean-squared-error bound for the estimate of the missing outcome of any unit at time T. Our work extends the recently growing literature on inference with adaptively collected data by allowing for policies that pool across units and also compliment the matrix completion literature when the entries are revealed sequentially in an arbitrarily dependent manner based on prior observed data.

IS31: Room: G16	**Safe, Anytime-valid Inference**	**Chair: Rina Barber**

Game-Theoretic Probability and Statistics
Speaker: **Volodya Vovk, Royal Holloway University of London**
Game-theoretic probability is an approach to the foundations of probability motivated by Philip Dawid's prequential principle and going outside Kolmogorov's measure-theoretic axioms of probability. It replaces an overarching stochastic picture by a more modest picture of isolated islands of stochasticity. I will argue that, while game-theoretic probability is not indispensable for game-theoretic statistics, it can serve as a convenient, intuitive, and philosophically clean foundation. The main mathematical tools of game-theoretic statistics, e-values (a recent alternative to p-values) and test martingales, become even more natural when interpreted as concepts of game-theoretic probability.

Sequential Tests with Optional Stopping and Anytime-Valid Effect Size Estimation for Two-Sample Data Streams using E-Processes
Speaker: **Rosanne Turner, CWI Amsterdam**
We develop sequential tests and corresponding anytime-valid confidence sequences for sequential effect size estimation based on E-variables. We introduce a general method for constructing E-processes that can be used for testing in 2-sample streams, for general null hypotheses. In contrast to earlier methods developed in the sequential testing literature, our approach is valid for both balanced and unbalanced experiments and allows for arbitrary, user-specified notions of effect size. With two Bernoulli streams as a running example, we illustrate the power of our test and show that expected stopping times often are lower than or comparable to stopping times of classical methods that do not allow for optional stopping, such as Fisher's exact test. We also illustrate how anytime-valid confidence sequences are constructed with two different notions of effect size: log odds ratio and difference in mean.

Testing Exchangeability: Fork-Convexity, Supermartingales, and E-processes
Speaker: **Johannes Ruf, London School of Economics and Political Science**
Suppose we observe an infinite series of coin flips $X1, X2, \ldots$, and wish to sequentially test the null that these binary random variables are exchangeable. Nonnegative supermartingales (NSMs) are a workhorse of sequential inference, but we prove that they are powerless for this problem. First, utilizing a geometric concept called fork-convexity (a sequential analog of convexity), we show that any process that is an NSM under a set of distributions, is also necessarily an NSM under their "fork-convex hull". Sec-

ond, we demonstrate that the fork-convex hull of the exchangeable null consists of all possible laws over binary sequences; this implies that any NSM under exchangeability is necessarily nonincreasing. Since testing arbitrary deviations from exchangeability is information-theoretically impossible, we focus on Markovian alternatives. We combine ideas from universal inference and the method of mixtures to derive a "safe e-process", which is a nonnegative process with expectation at most one under the null at any stopping time, and is upper bounded by a martingale, but is not itself an NSM.

E-detectors: A Nonparametric Framework for Sequential Change-point Detection
Speaker: **Alessandro Rinaldo, Carnegie Mellon University**
Quickest sequential change-point detection is a classic problem with a variety of applications. We develop a new and general framework for univariate quickest change point detection when the pre- and post-change distributions are nonparametrically specified (and thus composite). Our procedures come with nonasymptotic bounds on the average run length (frequency of false alarms). For certain classes of nonparametric distributions (including those of sub-Gaussian or sub-exponential type), we also provide near-optimal bounds on the detection delay following a change-point. Our main technical contribution is the formalization of the notion of an e-detector, a sums of e-processes that are started at consecutive times. We first introduce simple Shiryaev–Roberts and CUSUM-style e-detectors, and then show how to design their mixtures in order to achieve both statistical and computational efficiency. We demonstrate their performance in detecting changes in the mean of a bounded random variable without relying on the i.i.d. assumption, with an application to tracking the performance of a sports team over multiple seasons.

Monday 27-06-2022	13:30-15:30	Topic Contributed Sessions (TC6)

TC6: Room: G04 Recent advances in rank-based inference Chair: Marc Hallin

On Universally Consistent and Fully Distribution-Free Rank Tests of Vector Independence
Speaker: **Hongjian Shi, Technical University of Munich**
Rank correlations have found many innovative applications in the last decade. In particular, suitable rank correlations have been used for consistent and distribution-free tests of independence between pairs of random variables. However, the traditional concept of ranks relies on ordering data and is, thus, tied to univariate observations. As a result, it has long remained unclear how one may construct distribution-free yet consistent tests of independence between random vectors. In this talk, I will discuss how this problem can be addressed via a general framework for designing multivariate dependence measures and associated test statistics based on the recently introduced concept of center-outward ranks and signs, a multivariate generalization of traditional ranks. In this framework, we obtain new multivariate Hájek asymptotic representation results and use them for local power analyses that demonstrate the statistical efficiency of our tests.

Regression and Autoregression Rank Score Processes and their Functionals
Speaker: **Jana Jurečková, Charles University and Academy of Sciences, Prague**
Regression and autoregression rank score processes, accompanied with their dual quantile processes, provide surprisingly efficient tools of the statistical inference. Their invariance and equivariance properties enable to follow indicators of interest, risk measures and trends, even in the presence of covariates which are not under our control. Unlike the classical statistical methods, these processes and their functionals provide reliable results even in the presence of measurement errors, under sequential dependence in the dataset, etc. We shall discuss their properties and new posible applications, not yet shown in the literature.

Ancillarity and Semiparametric Efficiency
Speaker: **Bo Zhou, Durham University**
The classical notions of (maximal) ancillarity and (minimal) sufficiency in statistical experiments are related through Basu's Theorem. We embed these notions in the Hajek-Le Cam theory of convergence of experiments leading to a unified theory of asymptotically optimal semiparametric inference for a wide class of models, including LAN, LAMN and LABF experiments. For each of these, we provide examples where our theory leads to new results on semiparametrically optimal inference. This is joint work with Bas Werker.

Center-outward Rank Tests for Multiple-output Regression and MANOVA
Speaker: **Šárka Hudecová, Charles University**
A concept of center-outward multivariate ranks and signs based on measure transportation ideas has been introduced recently. In this contribution we describe the usage of these concepts in a multiple-output regression model, which contains, as particular cases, the two-sample location and MANOVA models. Asymptotic properties of the resulting rank and sign tests are provided and their finite sample performance is illustrated by a simulation study.

Monday 27-06-2022 16:00-17:30 Topic Contributed Sessions (TC1 – 5, 7–8)

TC1: Room: G03 Semiparametric Approaches and Randomization Test in Causal Inference Chair: Yu Cheng

Instrumental Variable Estimation of Complier Causal Treatment Effect with Interval-Censored Data
Speaker: **Limin Peng, Emory University**
Assessing causal treatment effect on a time-to-event outcome is of key interest in many scientific investigations. Instrumental variable (IV) is a useful tool to mitigate the impact of endogenous treatment selection to attain unbiased estimation of causal treatment effect. Existing development of IV methodology, however, hasn't attended to outcomes subject to interval censoring, which are ubiquitously present in studies with intermittent follow-up but are challenging to handle in terms of both theory and computation. In this work, we fill in this important gap by studying a general class of causal semiparametric transformation models with interval-censored data. We propose a nonparametric maximum likelihood estimator of the complier causal treatment effect. Moreover, we design a reliable and computationally stable EM algorithm which has a tractable objective function in the maximization step via the use of Poisson latent variables. The asymptotic properties of the proposed estimators, including the consistency, asymptotic normality, and semiparametric efficiency, are established with empirical process techniques. We conduct extensive simulation studies and an application to a colorectal cancer screening dataset, showing satisfactory finite-sample performance of the proposed method as well as its prominent advantages over naive methods.

Imputation-Based Q-Learning for Optimizing Dynamic Treatment Regimes with Right-Censored Survival Outcome
Speaker: **Abdus Wahed, University of Pittsburgh**
Q-learning is one of the most commonly used methods for optimizing dynamic treatment regimes (DTRs) in multi-stage decision-making. Right-censored survival outcome poses a significant challenge to Q-Learning due to its reliance on parametric models that are subject to misspecification and sensitive to missing covariates. In this talk, we discuss an imputation-based Q-learning (IQ-learning) where the semiparametric Cox proportional hazard (CPH) model is employed to estimate optimal treatment rules at each stage, and then weighted hot-deck multiple imputations (MI) and direct-draw MI are used to predict optimal potential survival times. Missing data are handled using inverse probability weighting and MI, and the non-random treatment assignment among the observed is accounted for using a propensity-score approach. We investigate the performance of IQ-learning via extensive simulations. We demonstrate IQ-Learning by developing an optimal DTR for leukemia treatment based on a randomized trial with observational follow-up.

Multi-Threshold Structural Equation Model
Speaker: **Jialiang Li, National University of Singapore**
In this article, we consider the instrumental variable estimation for causal regression parameters with multiple unknown structural changes across sub-populations. We propose a multiple change point detection method to determine the number of thresholds and estimate the threshold locations in the two-stage least square procedure. After identifying the estimated threshold locations, we use the Wald method to estimate the parameters of interest, that is, the regression coefficients of the endogenous variable. Based on some technical assumptions, we carefully establish the consistency of estimated parameters and the asymptotic normality of causal coefficients. Simulation studies are included to examine the performance of the proposed method. Finally, our method is illustrated via an application of the Philippine farm households data for which some new findings are discovered.

Multiple conditional randomization tests
Speaker: **Qingyuan Zhao, University of Cambridge**
We establish a general sufficient condition on constructing multiple "nearly independent" conditional randomization tests, in the sense that the joint distribution of their p-values is almost uniform under the global null. This property implies that the tests are jointly valid and can be combined using standard methods. Our theory generalizes existing techniques in the literature that use independent treatments, sequential treatments, or post-randomization, to construct multiple randomization tests. In particular, it places no condition on the experimental design, allowing for arbitrary treatment variables, assignment mechanisms and unit interference. The flexibility of this framework is illustrated through developing conditional randomization tests for lagged treatment effects in stepped-wedge randomized controlled trials. A weighted Z-score test is further proposed to maximize the power when the tests are combined. We compare the efficiency and robustness of the commonly used mixed-effect models and the proposed conditional randomization tests using simulated experiments and real trial data. This talk is based on joint work with Yao Zhang.

TC2: Room: G04 Algebraic Structures in Statistics Chair: Jane Ivy Coons

Statistics for Phylogenetic Trees with Tropical Geometry
Speaker: **Anthea Monod, Imperial Colleage London**
We study the problem of statistical analysis on sets of phylogenetic trees. The setting is challenging due to the discrete geometric structure of the trees themselves, which results in a nonlinear, non-smooth tree space, as well as size and dimensionality aspects (datasets can be very large, trees themselves can also be large). We focus on a tree space based on a variant of algebraic geometry and study its structural and geometric characteristics in comparison to the current standard; in particular, we discuss their implications on statistics and data analysis, offering a novel perspective to statistics on phylogenetic trees as well as a novel statistical

perspective in tropical geometry. We will present results of a comparative study on influenza data.

Interventional Model Equivalence via Algebraic Geometry
Speaker: **Liam Solus, KTH Royal Institute of Technology**
By viewing a statistical model as the intersection of an algebraic variety with the space of canonical parameters we can arrive at new proofs of classic theorems about model equivalence, as well as generalizations thereof. One case where this approach is particularly useful is the problem of learning a causal model from a mixture of observational and interventional data. Here, the goal is to learn a causal graph on the system of variables from the available data. In this talk we will see that, with the help of algebraic geometry, we can provide structural characterizations of interventional model equivalence that generalize classic results for DAG models to a more granular family of models capable of encoding context-specific causal relations. We will look at some applications of these results to real data and discuss the implications for generalizations of classic causal discovery algorithms to the context-specific setting.

On the Strongly Robustness Property: When Markov and Graver Basis Coincidence
Speaker: **Dimitra Kosta, University of Edinburgh**
To every toric ideal one can associate an oriented matroid structure, consisting of a graph and another toric ideal, called bouquet ideal. The connected components of this graph are called bouquets. Bouquets are of three types; free, mixed and non mixed. We prove that the cardinality of the following sets - the set of indispensable elements, minimal Markov bases, the Universal Markov basis and the Universal Gröbner basis of a toric ideal - depends only on the type of the bouquets and the bouquet ideal. These results enable us to introduce the strongly robustness simplicial complex and show that it determines the strongly robustness property. For codimension 2 toric ideals, we study the strongly robustness simplicial complex and prove that robustness implies strongly robustness.

Gaussian Conditional Independence Beyond Graphical Models
Speaker: **Tobias Boege, Max Planck Institute for Mathematics in the Sciences**
Graphical models are widely useful and benign statistical models. They allow the concise specification of dependencies among a set of random variables and their algorithmic manipulation.For regular Gaussian distributions, parametrizations and efficient inference algorithms for these models are available. This talk gives an overview of the algebraic, algorithmic and topological complexity of general, non-graphical, Gaussian conditional independence models. It is shown that the desirable properties of graphical models do not continue to hold and, in fact, Gaussian CI models are as complicated in these regards as any semi-algebraic set can possibly be.

TC3: Room: G05	Statistical Methods for Innovative Experimental Designs	Chair: Sofia Villar

Bayesian Algorithms for Adaptive Trials that Trade Off Statistical Analysis with Benefits to Participants
Speaker: **Joseph J Williams, University of Toronto**
Multi-armed bandit algorithms like TS (Thompson Sampling) can be used to conduct adaptive experiments, in which maximising reward means that data is used to progressively assign more participants to more effective arms. Such assignment strategies increase the risk of statistical hypothesis tests identifying a difference between arms when there is not one, and failing to conclude there is a difference in arms when there truly is one (Rafferty et al. 2019). We tackle this by introducing a novel heuristic algorithm, called TS-PostDiff (Posterior Probability of Difference). TSPostDiff takes a Bayesian approach to mixing TS and Uniform Random: the probability a participant is assigned using Uniform Random allocation is the posterior probability that the difference between two arms is "small" (below a certain threshold), allowing for more Uniform Random exploration when there is little or no reward to be gained. We evaluate TSPostDiff against state-of-the-art strategies. The empirical and simulation results help characterise the trade-offs of these approaches between Reward, FPR (False Positive Rate), and Statistical Power, and under which circumstances each is effective. We quantify the advantage of TS-PostDiff in performing well across multiple differences in arm means (effect sizes), showing the benefits of adaptively changing randomisation/exploration in TS in a "Statistically Considerate" manner: to reduce FPR and increase Statistical Power when differences are small or zero and there is less reward to be gained, while exploiting more when differences may be large. This highlights the considerations important for future algorithm development and analysis to better balance reward and statistical analysis.

A Bayesian Decision-theoretic Randomisation Procedure for Multi-armed Clinical Trials with Normally Distributed Outcomes
Speaker: **Faye Williamson, Newcastle University**
We present a response-adaptive randomisation (RAR) procedure for multi-armed clinical trials with normally distributed outcomes and unknown variances. This procedure is set in the Bayesian decision-theoretic framework and aims to maximise the Bayes-expected total reward in a trial, where the reward is taken to be a measure of patient benefit. In this talk, we utilise the near-optimal Gittins index solution of the multi-armed bandit problem to propose a forward-looking randomisation procedure, referred to as the Forward-Looking Gittins Index (FLGI), which adapts to known future information as well as past observations. We illustrate the proposed procedure by simulations in the context of phase II cancer trials. Results show that, in a multi-armed setting, there are both efficiency and patient benefit gains of using RAR procedures, such as the FLGI, with a continuous outcome instead of artificially dichotomising to form a binary one. These gains persist even if an anticipated low rate of missing data is imputed online using an approach introduced in this work.

Optimal Designs for Testing the Efficacy of Heterogeneous Experimental Groups
Speaker: **Rosamarie Frieri, University of Bologna**
The problem of comparing the means of several experimental groups is well-known in the statistical literature and is relevant in many applied fields. The majority of the papers deal with optimal designs for the estimation of treatment effects, while the problem of hypothesis testing has received limited attention. We presented a unified framework for deriving optimal designs for testing the homogeneity of several treatments, which includes the general ANOVA set-up with heteroscedastic errors. The optimal designs are generalized Neyman allocations involving only two treatment groups. In addition, with a particular interest in the clinical trial context, we derived optimal designs maximizing the power of the classical Wald test subject to the constraint that the allocation proportions are ordered according to the magnitude of the treatment effects. However, such ordering is generally a-priori unknown, so the proposed allocations are locally optimal designs that can be implemented via response-adaptive randomization procedures. We illustrated the properties of the suggested allocations both theoretically and numerically - also in comparisons with other existing designs - showing very good performance in terms of both statistical power and ethical demand.

Multinomial Thompson Sampling for adaptive experiments with rating scales data
Speaker: **Nina Deliu, Sapienza University of Rome**
Bandit algorithms such as Thompson Sampling (TS) have been argued for decades as useful for conducting adaptively-randomised experiments. By skewing the allocation ratio towards superior arms, they can substantially improve participants' welfare with respect to particular outcomes of interest. For example, as we illustrate in this work, they may use participants ratings for understanding and assigning promising text messages for managing mental health issues more often. In such setting, model-based algorithms as TS are generally implemented based on a normal model assumption for the reward variable, which is typically unrealistic. Alternatively, the outcome is dichotomised–sometimes based on arbitrary rules–and a binary model is considered. When dealing with categorical rating scale outcomes, these common practices may lead to suboptimal performances both during the decision-making process (e.g., increased regret) and secondary statistical analyses (e.g., increased bias or frequentist errors). Guided by the evidence of our two-armed field experiment, in this work we introduce Multinomial TS for rating scale data, and show its improved empirical performance compared to common modelling approaches. We then provide an insightful exploration of Multinomial TS's arm-allocation rates, along with their asymptotic behaviour, and motivate an Augmented Multinomial TS alternative, corrective of a potentially aggressive (i.e., with arm-allocation rates close to 0 or 1) behaviour of TS under identical arms scenarios.

TC4: Room: G07 Advances in Time Series Analysis Chair: Greta Goracci

Classification with High-dimensional Time Series
Speaker: **Kung-Sik Chan, University of Iowa**
Multivariate time-series (MTS) data are prevalent in diverse domains and often high dimensional. In this talk, I will introduce two approaches for developing classifiers with high-dimensional MTS. The first approach assumes that the true underlying dynamics comprise a core low-dimensional dynamics corrupted by some background white noise, resulting in a spiked spectrum. Within this framework, we have developed non-parametric methods for estimating a spiked population spectrum and its rank, and develop several spike-spectrum based classifiers for high-dimensional time series. In the general case that the underlying spectra need not be spiked, we propose the random projection ensemble classifiers for time series (RPECTS). The RPECTS method first applies dimension reduction in the time domain via randomly projecting the time-series variables into some low dimensional space, followed by measuring the disparity via some novel base classifier between the data and the candidate generating processes in the projected space. We derive optimal weighted majority voting schemes for pooling information from the base classifiers for multiclass classification, and introduce new base frequency-domain classifiers based on Whittle likelihood (WL), Kullback-Leibler divergence (KL), Eigen-Distance (ED) and Chernoff divergence (CH). We illustrate the methods with simulation and real applications. The talk is based on joint work with Fuli Zhang.

High-dimensional Time Series Segmentation via Factor-adjusted Vector Autoregressive Modelling
Speaker: **Haeran Cho, University of Bristol**
Piecewise stationarity is a widely adopted assumption for modelling non-stationary time series. However fitting piecewise stationary vector autoregressive (VAR) models to high-dimensional data is challenging as the number of parameters increases as a quadratic of the dimension. Recent approaches to address this have imposed sparsity assumptions on the parameters of the VAR model, but such assumptions have been shown to be inadequate when datasets exhibit strong (auto)correlations. We propose a piecewise stationary time series model that accounts for pervasive serial and cross-sectional correlations through a factor structure, and only assumes that any remaining idiosyncratic dependence between variables can be modelled by a sparse VAR model. We propose an accompanying two-stage change point detection methodology which fully addresses the challenges arising from not observing either the factors or the idiosyncratic VAR process directly. Its consistency in estimating both the total number and the locations of the change points in the latent components, is established under conditions considerably more general than those in the existing literature. We demonstrate the competitive performance of the proposed methodology on simulated datasets and an application to US blue chip stocks data. This is joint work with Hyeyoung Maeng, Idris A. Eckley and Paul Fearnhead (Lancaster).

The validity of bootstrap testing in the threshold framework
Speaker: **Simone Giannerini, University of Bologna**
We consider bootstrap-based testing for threshold effects in non-linear threshold autoregressive (TAR) models. It is well-known that classic tests based on asymptotic theory tend to be oversized in the case of small, or even moderate sample sizes, or when the estimated parameters indicate non-stationarity, as often witnessed in the analysis of financial or climate data. To address the issue we propose a supremum Lagrange Multiplier test statistic (sLMb), where the null hypothesis specifies a linear autoregressive (AR) model against the alternative of a TAR model. We consider a recursive bootstrap applied to the sLMb statistic and establish its validity. This result is new, and requires the proof of non-standard results for bootstrap analysis in time series models; this includes a uniform bootstrap law of large numbers and a bootstrap functional central limit theorem. These new results can also be used as a general theoretical framework that can be adapted to other situations, such as regime-switching processes with exogenous threshold variables, or testing for structural breaks. The Monte Carlo evidence shows that the bootstrap test has correct empirical size even for small samples, and also no loss of empirical power when compared to the asymptotic test. Moreover, its performance is not affected if the order of the autoregression is estimated based on information criteria. Finally, we analyse a panel of short time series to assess the effect of warming on population dynamics. (Joint work with Greta Goracci and Anders Rahbek.)

High dimensional threshold regression with common stochastic trends
Speaker: **Lorenzo Trapani, University of Nottingham**
We study inference for threshold regression in the context of a large panel factor model with common stochastic trends. We develop a Least Squares estimator for the threshold level, deriving almost sure rates of convergence and proposing a novel, testing based, way of constructing confidence intervals. We also investigate the properties of the PC estimator for the loadings and common factors in both regimes, and develop a procedure to estimate the number of common trends in each regime. Our theoretical findings are corroborated through a comprehensive set of Monte Carlo experiments and an application to US mortality data. (Joint work with Daniele Massacci.)

TC5: Room: G16	**Graphical Models**	**Chair: Alexandros Grosdos**

Towards Standard Imsets for Maximal Ancestral Graphs
Speaker: **Robin Evans, University of Oxford**
Imsets, introduced by Studený (see Studený, 2005 for details), are an algebraic method for representing conditional independence models. They have many attractive properties when applied to such models, and they are particularly nice when applied to directed acyclic graph (DAG) models. In particular, the standard imset for a DAG is in one-to-one correspondence with the independence model it induces, and hence is a label for its Markov equivalence class.We present a proposed extension to standard imsets for maximal ancestral graph (MAG) models, using the parameterizing set representation of Hu and Evans (2020). By construction, our imset also represents the Markov equivalence class of the MAG. We show that for many such graphs our proposed imset defines the model, though there is a subclass of graphs for which the representation does not. We prove that it does work for MAGs where there are no adjacent bidirected edges without an ancestral relation, as well as for a large class of purely bidirected models. If there is time, we will also discuss applications of imsets to structure learning in MAGs.

Staged Trees and Asymmetry-Labeled DAGs
Speaker: **Manuele Leonelli, IE University**
Bayesian networks are a widely-used class of probabilistic graphical models capable of representing symmetric conditional independence between variables of interest using the topology of the underlying graph. They can be seen as a special case of the much more general class of models called staged trees, which can represent any type of non-symmetric conditional independence. In this talk I will formalize the relationship between these two models and introduce a minimal Bayesian network representation of the staged tree, which can be used to read conditional independences in an intutitive way. Furthermore, I will define a new labeled graph, termed asymmetry-labeled directed acyclic graph, whose edges are labeled to denote the type of dependence existing between any two random variables. Various datasets are used to illustrate the methodology, highlighting the need to construct models which more flexibly encode and represent non-symmetric structures. The talk is based on joint work with Federico Carli and Gherardo Varando.

Total Positivity in Graphical Models for Extremes
Speaker: **Frank Röttger, Université de Genève**
Engelke and Hitz (2020) recently introduced a general theory for conditional independence and graphical models for extremes. For Hüsler–Reiss distributions, the extremal analogue of Gaussians, it follows that their precision matrices similarly to Gaussians encode the extremal graphical structure. Multivariate total positivity of order 2 (MTP_2) is a strong form of positive dependence that induces many interesting properties in graphical modeling. A multivariate Gaussian is MTP_2 when its precision matrix is an M-matrix, i.e. when all the non-diagonal entries in the precision matrix are non-positive. We introduce the notion of extremal MTP_2 ($EMTP_2$) and show that many classical models are always $EMTP_2$. A Hüsler–Reiss random vector is $EMTP_2$ if and only if its precision matrix is the Laplacian matrix of a connected graph with positive edge weights. We propose an estimator for the parameters of the Hüsler–Reiss distribution under $EMTP_2$ as the solution of a convex optimization problem with Laplacian constraint. We prove that this estimator is consistent and typically yields a sparse model with possibly non-decomposable extremal graphical structure.We construct a block descent algorithm and demonstrate on real data that our $EMTP_2$ estimator outperforms other available graphical

estimators.

Causal Structure Discovery between Clusters of Nodes Induced by Latent Factors
Speaker: **Chandler Squires, Massachusetts Institute of Technology**
We consider the problem of learning the structure of a causal directed acyclic graph (DAG) model in the presence of latent variables. We define latent factor causal models (LFCMs) as a restriction on causal DAG models with latent variables, which are composed of clusters of observed variables that share the same latent parent and connections between these clusters given by edges pointing from the observed variables to latent variables. LFCMs are motivated by gene regulatory networks, where regulatory edges, corresponding to transcription factors, connect spatially clustered genes. We show identifiability results on this model and design a consistent three-stage algorithm that discovers clusters of observed nodes, a partial ordering over clusters, and finally, the entire structure over both observed and latent nodes. We evaluate our method in a synthetic setting, demonstrating its ability to almost perfectly recover the ground truth clustering even at relatively low sample sizes, as well as the ability to recover a significant number of the edges from observed variables to latent factors. Finally, we apply our method in a semi-synthetic setting to protein mass spectrometry data with a known ground truth network, and achieve almost perfect recovery of the ground truth variable clusters.

TC7: Room: G22 New progress and directions in nonparametric statistics Chair: Hongjian Shi

Nonparametric Regression on Lie groups with Measurement Errors
Speaker: **Jeong Min Jeon, KU Leuven**
We study the problem of estimating the regression and density functions for predictors taking values on compact Lie groups and contaminated by measurement errors. Our methodology and theory are based on harmonic analysis on Lie groups. We derive their rates of convergence and asymptotic distributions. Our numerical studies show that our estimators outperform competitors ignoring measurement errors or the geometric structures of Lie groups.

Honest Confidence Bands for Quantile Curves
Speaker: **Lutz Dümbgen, University of Bern**
An essential building block of distributional regression are conditional quantiles. This talk reviews some procedures for regression quantiles and presents new honest confidence bands for such curves under isotonicity and other shape constraints.

Limiting Laws for Optimal Transport Plans on Finite Spaces
Speaker: **Yoav Zemel, University of Cambridge**
Optimal transport is now a popular tool in statistics, machine learning, and data science. The majority of studies regarding the asymptotic properties of optimal transport have focussed on the Wasserstein distance itself, i.e., the optimal objective value. In many situations, however, it is the transport plan (or map) that is more informative, as it allows the practitioner to understand *where*, and not only how much, transport is taking place. We thus study the asymptotics of optimal transport plans in the (practically relevant) case of finite ground space. Possible limiting distributions are derived, and it is shown that the limiting distributions are non-standard with complexity that depends on the degeneracy of the optimal transport linear program and its dual. In particular, if the dual is degenerate (as it is in regular ground spaces), the asymptotic distribution depends on the way the empirical optimal solution is chosen. The results are not specific to optimal transport and hold for general linear programs. This talk is Based on joint work with Marcel Klatt and Axel Munk.

TC8: Room: G26 New Change-point Algorithms and Theoretical Results Chair: Toby Hocking

Optimistic Search Strategy: Change-point Detection for Large-scale Data via Adaptive Logarithmic Queries
Speaker: **Solt Kovacs, ETH Zurich**
As a classical and ever reviving topic, change point detection is often formulated as a search for the maximum of a gain function describing improved fits when segmenting the data. Searching through all candidate split points on the grid for finding the best one requires $O(T)$ evaluations of the gain function for an interval with T observations. If each evaluation is computationally demanding (e.g. in high-dimensional models), this can become infeasible. Instead, we propose optimistic search strategies with $O(\log T)$ evaluations exploiting the specific structure of the gain function. Towards solid understanding of our strategies, we investigate in detail the Gaussian change in mean setup. For some of our proposals, we prove asymptotic minimax optimality for single and multiple change point scenarios. Our search strategies generalize far beyond the theoretically analyzed setup. We illustrate, as an example, the massive computational speedup in change point detection for high-dimensional Gaussian graphical models. More generally, we demonstrate empirically that optimistic search methods lead to competitive estimation performance while heavily reducing run-time, both in classical univariate, as well as in high-dimensional problems.

FOCuS: Online Changepoint Detection via Functional Pruning CUSUM Statistics
Speaker: **Gaetano Romano, University of Lancaster**
Many modern applications of online change-point detection require the ability to process high-frequency observations in real-time, sometimes with limited available computational resources. Online algorithms for detecting a change-in-mean often involve using a moving window or specifying the expected size of change. Such choices affect which changes the algorithms have most power

to detect. We introduce an algorithm, Functional Online CuSUM (FOCuS), which is equivalent to running these earlier methods simultaneously for all sizes of window, or all possible values for the size of change, with a computational cost logarithmic in the number of observations. We demonstrate FOCuS practical utility through its state-of-the-art performance at detecting anomalous behaviour in computer server data.

Detecting changepoints in periodic data
Speaker: **Owen Li, University of Lancaster**
Traditional changepoint approaches consider changepoints to occur linearly in time; one changepoint happens after another, and they are not linked. However, data processes may exhibit periodic behaviour and so changepoints will occur regularly, e.g. sleeping patterns and daily routine behaviour. Using linear changepoint approaches in these settings will miss global changepoint features which affect changepoints on the more local (periodic) level, for example the introduction of lockdowns affecting sleeping patterns. Being able to tease these global changepoint features from the more local (periodic) ones is beneficial for inference. We propose a deterministic periodic changepoint method using a periodic (circular) time perspective. This is done by adapting the Segment Neighbourhood changepoint method to the periodic time perspective. We then integrate this local changepoint model into the pruned exact linear time (PELT) search algorithm to identify the optimal global changepoint positions. We demonstrate that the method detects both local and global changepoints with high accuracy on simulations and motivating digital health applications.

Random Forests for Change Point Detection
Speaker: **Malte Londschien, ETH Zurich**
We propose a novel multivariate nonparametric multiple change point detection method using classifiers. We construct a classifier log-likelihood ratio that uses class probability predictions to compare different change point configurations. We propose a computationally feasible search method that is particularly well suited for random forests, denoted by changeforest. However, the method can be paired with any classifier that yields class probability predictions, which we illustrate by also using a k-nearest neighbor classifier. We provide theoretical results motivating our choices. In a large simulation study, our proposed changeforest method achieves improved empirical performance compared to existing multivariate nonparametric change point detection methods. An efficient implementation of our method is made available for R, Python, and Rust users in the changeforest software package. This is joint work with Peter Bühlmann and Solt Kovács. arXiv: https://arxiv.org/abs/2205.04997

Tuesday 28-06-2022 16:00-17:30 Topic Contributed Sessions (TC9, 10, 12 – 16)

TC9: Room: G03 New models and evaluation methods for changepoint detection Chair: Rebecca Killick

Multiple Change-point Detection with a priori Constraints on Consecutive Changes
Speaker: **Vincent Runge, University of Evry**
The accurate detection of multiple change-points in univariate time series is a daily task for many engineers and scientists. Often, practitioners can have prior knowledge about the type of changes they are looking for. For example in genomic data, biologists expect peaks: up changes followed by down changes. Integrating such priors is important and requires dedicated algorithms. We propose a generic algorithm (available in 'gfpop' R package) able to deal with many priors constraining the successive segment parameters. This algorithm can be seen as a Hidden Markov Chain model with continuous state space. gfpop works for a user-defined graph and several loss functions: Gauss, Poisson, Binomial, Biweight and Huber. The underlying penalized optimization problem is solved by the functional pruning optimal partitioning (fpop) algorithm. Integrating constraints in this algorithm leads to complex algorithmic update rules for piecewise functions. We will focus on some new ideas for decreasing time complexity and will compare functional pruning performances for various implementations.

Automatic Calibration of Change-point Detection Methods
Speaker: **Charles Truong, Centre Borelli**
In numerous applications involving time series, change-point detection is a common step of the data processing pipeline. This step detects the instants when certain user-defined characteristics of the signal abruptly change, for instance, the mean, the variance, the spectral content, etc. A significant difficulty of change detection methods is the calibration. Finding an appropriate set of calibration parameters often requires a laborious manual trial-and-error procedure and expert knowledge in both the application domain (e.g. biostatistics or econometrics) and change-point methods. To alleviate this issue, we propose in this work an automatic calibration procedure that only relies on labels provided by an expert. To that end, the combinatorial optimisation problem at the heart of the detection methods is modified to yield a differentiable (in the parameters) formulation that can then be used as an objective function to learn calibration parameters in a supervised manner. In addition to detection accuracy, this formulation is able to integrate a number of constraints if needed, such as dimension reduction, scaling, sparse representation, etc. We show on several real-world examples that our strategy can learn a correct detection algorithm, with very little input from the domain expert.

Optimizing ROC Curves with a Sort-Based Surrogate Loss Function for Binary Classification and Changepoint Detection
Speaker: **Toby Dylan Hocking, Northern Arizona University**
Receiver Operating Characteristic (ROC) curves are plots of true positive rate versus false positive rate which are useful for evaluating binary classification models, but difficult to use for learning since the Area Under the Curve (AUC) is non-convex. ROC curves can also be used in other problems that have false positive and true positive rates such as changepoint detection. We show that in this more general context, the ROC curve can have loops, points with highly sub-optimal error rates, and AUC greater than one. This observation motivates a new optimization objective: rather than maximizing the AUC, we would like a monotonic ROC curve with AUC=1 that avoids points with large values for Min(FP,FN). We propose a relaxation of this objective that results in a new surrogate loss function called the AUM, short for Area Under Min(FP, FN). Whereas previous loss functions are based on summing over all labeled examples or pairs, the AUM requires a sort and a sum over the sequence of points on the ROC curve. We show that AUM directional derivatives can be efficiently computed and used in a gradient descent learning algorithm. In our empirical study of supervised binary classification and changepoint detection problems, we show that our new AUM minimization learning algorithm results in improved AUC and comparable speed relative to previous baselines.

An Evaluation of Change Point Detection Algorithms
Speaker: **Gerrit Van der Burg, Alan Turing Institute**
Change point detection (CPD) is an important part of time series analysis, as the presence of a change point indicates an abrupt and significant change in the data generating process. While many algorithms for CPD have been proposed, comparatively little attention has been paid to evaluating their performance on real-world time series. Algorithms are typically evaluated on simulated data and a small number of commonly-used series with unreliable ground truth. Clearly this does not provide sufficient insight into the performance of these algorithms. Therefore, instead of developing yet another change point detection method, we consider it vastly more important to properly evaluate existing algorithms on real-world data. In this work we present a data set specifically designed for the evaluation of CPD algorithms that consists of 37 time series from various application domains. Each series was annotated by five human annotators to provide ground truth on the presence and location of change points. We analyze the consistency of the human annotators, and describe evaluation metrics that can be used to measure algorithm performance in the presence of multiple ground truth annotations. Next, we present a benchmark study where 14 algorithms are evaluated on each of the time series in the data set. Our aim is that this data set will serve as a proving ground in the development of novel change point detection algorithms. (Joint work with Chris Williams.)

TC10: Room: G04 Statistical and Causal Inference in Algorithmic Fairness and Recourse Chair: Nicolai Meinshausen

Statistical vs. Causal Fair Predictions
Speaker: **Drago Plecko, ETH Zurich**
Many definitions of fairness are based on statistical criteria, such as independence relations. We discuss some of these definitions and show that if such criteria are used as constraints when optimizing regressors/classifiers, the resulting predictor might have unintended properties, from a causal perspective. We discuss an approach, based on a causal diagram G, that alleviates the problem, and also allows for natural, causally interpretable relaxations of the possibly prohibitively strong independence relations.

Algorithmic Recourse: from theory to practice
Speaker: **Amir-Hossein Karimi, Max Planck Institute for Intelligent Sustem Tubingen**
Algorithmic recourse is concerned with aiding individuals who are unfavorably treated by automated decision-making systems to overcome their hardship, by offering recommendations that would result in a more favorable prediction when acted upon. Such recourse actions are typically obtained through solving an optimization problem that minimizes changes to the individual's feature vector, subject to various plausibility, diversity, and sparsity constraints. Whereas previous works offer solutions to the optimization problem in a variety of settings, they critically overlook real-world considerations pertaining to the environment in which recourse actions are performed. The present work emphasizes that changes to a subset of the individual's attributes may have consequential down-stream effects on other attributes, thus making recourse a fundamentally causal problem. Here, we model such considerations using the framework of structural causal models, and highlight pitfalls of not considering causal relations through examples and theory. Such insights allow us to reformulate the optimization problem to directly optimize for minimally-costly recourse over a space of feasible actions (in the form of causal interventions) rather than optimizing for minimally-distant "counterfactual explanations". We offer both the optimization formulations and solutions to deterministic and probabilistic recourse, on an individualized and sub-population level, overcoming the steep assumptive requirements of offering recourse in general settings. Finally, using synthetic and semi-synthetic experiments based on the German Credit dataset, we demonstrate how such methods can be applied in practice under minimal causal assumptions. Finally, we show how to offer fair and adversarially robust recourse under the causal assumptions considered above.

TC12: Room: G07 Graphical Modeling of Multivariate Functional Data Chair: Alexander Petersen

Functional Differential Graph Estimation with Fully and Discretely Observed Curves
Speaker: **Mladen Kolar, University of Chicago**
We consider the problem of estimating the difference between two functional undirected graphical models with shared structures. In many applications, data are naturally regarded as a vector of random functions rather than a vector of scalars. For example, electroencephalography (EEG) data are more appropriately treated as functions of time. In such a problem, not only can the number of functions measured per sample be large, but each function is itself an infinite dimensional object, making estimation of model parameters challenging. This is further complicated by the fact that the curves are usually only observed at discrete time points. We first define a functional differential graph that captures differences between two functional graphical models and formally characterize when the functional differential graph is well defined. We then propose a method, FuDGE, that directly estimates the functional differential graph without first estimating each individual graph. This is particularly beneficial in settings where the individual graphs are dense, but the differential graph is sparse. We show that FuDGE consistently estimates the functional differential graph even in a high-dimensional setting for both discretely observed and fully observed function paths. We illustrate finite sample properties of our method through simulation studies. We also propose a competing method, the Joint Functional Graphical Lasso, which generalizes the Joint Graphical Lasso to the functional setting. Finally, we apply our method to EEG data to uncover differences in functional brain connectivity between a group of individuals with alcohol use disorderand a control group. Th is joint work Boxin Zhao and Y. Samuel Wang.

High-dimensional Nonparametric Functional Graphical Models via the Additive Partial Correlation Operator
Speaker: **Eftychia Solea, National School of Statistics and Information Analysis, France**
This article develops a novel approach for estimating a high-dimensional and nonparametric graphical model for functional data. Our approach is built on a new linear operator, the functional additive partial correlation operator, which extends the partial correlation matrix to both the nonparametric and functional setting. We show that its nonzero elements can be used to characterize the graph, and we employ sparse regression techniques for graph estimation. Moreover, the method does not rely on any distributional assumptions and does not require the computation of multi-dimensional kernels, thus avoiding the curse of dimensionality. We establish both estimation consistency and graph selection consistency of the pro- posed estimator, while allowing the number of nodes to grow with the increasing sample size. Through simulation studies, we demonstrate that our method performs better than existing methods in cases where the Gaussian or Gaussian copula assumption does not hold. We also demonstrate the performance of the proposed method by a study of an electroencephalography dataset to construct a brain network.

Adaptive Functional Thresholding for Sparse Covariance Function Estimation in High Dimensions
Speaker: **Qin Fang, London School of Economics and Political Science**
Covariance function estimation is a fundamental task in multivariate functional data analysis and arises in many applications. In this

paper, we consider estimating sparse covariance functions for high-dimensional functional data, where the number of random functions p is comparable to, or even larger than the sample size n. Aided by the Hilbert–Schmidt norm of functions, we introduce a new class of functional thresholding operators that combine functional versions of thresholding and shrinkage, and propose the adaptive functional thresholding of the sample covariance function capturing the variability of individual functional entries. We investigate the convergence and support recovery properties of our proposed estimator when p can grow exponentially with n. To handle the practical scenario where random functions are partially observed, we also develop a nonparametric smoothing approach to estimate entries of the covariance function, and propose fast binned approximations of estimators under the high-dimensional setting. The theoretical properties of the resulting estimators under different measurement schedules are established. Finally, we demonstrate that our proposed adaptive functional thresholding estimators significantly outperform the competing estimators through an extensive set of simulations and the functional connectivity analysis of two neuroimaging datasets.

Partial Separability and Functional Graphical Models for Multivariate Gaussian Processes
Speaker: **Alexander Petersen, Brigham Young University**
The covariance structure of multivariate functional data can be highly complex, especially if the multivariate dimension is large, making extensions of statistical methods for standard multivariate data to the functional data setting challenging. For example, Gaussian graphical models have recently been extended to the setting of multivariate functional data by applying multivariate methods to the coefficients of truncated basis expansions. This talk will address the general problem of covariance modeling for multivariate functional data and, more particularly, functional Gaussian graphical models. As a first step, a new notion of separability, termed partial separability, for the covariance operator of multivariate functional data is proposed, and several characterizations are provided. Next, the partial separability structure is shown to be particularly useful in order to provide a well-defined functional Gaussian graphical model that can be identified with a sequence of finite-dimensional graphical models. This motivates a simple and efficient estimation procedure through application of the joint graphical lasso, for which asymptotic consistency of edge selection is established. Empirical performance is assessed through simulation and analysis of functional brain connectivity during a motor task.

TC13: Room: G11	New Directions in Theory and Practice of Formally Private Synthetic Data	Chair: Aleksandra Slavkovic

Data Augmentation MCMC for Bayesian Inference from Privatized Data
Speaker: **Ruobin Gong, Rutgers University**
Differentially private mechanisms protect privacy by introducing additional randomness into the data. Restricting access to only the privatized data makes it challenging to perform valid statistical inference on parameters underlying the confidential data. Specifically, the likelihood function of the privatized data requires integrating over the large space of confidential databases and is typically intractable. For Bayesian analysis, this results in a posterior distribution that is doubly intractable, rendering traditional MCMC techniques inapplicable. We propose an MCMC framework to perform Bayesian inference from the privatized data, which is applicable to a wide range of statistical models and privacy mechanisms. Our MCMC algorithm augments the model parameters with the unobserved confidential data, and alternately updates each one conditional on the other. For the potentially challenging step of updating the confidential data, we propose a generic approach that exploits the privacy guarantee of the mechanism to ensure efficiency. In particular, we give results on the computational complexity, acceptance rate, and mixing properties of our MCMC. We illustrate the efficacy and applicability of our methods on a naïve-Bayes log-linear model as well as on a linear regression model. This is joint work with Nianqiao Ju, Jordan Awan, and Vinayak Rao.

Formal Privacy for Partially Private Data
Speaker: **Jeremy Seeman, Pennsylvania State University**
Differential privacy (DP) requires that any statistic based on confidential data be released with additional noise for privacy. Such a restriction can be logistically impossible to achieve, for example due to policy-mandated disclosure in the present or unsanitized data releases in the past. Still, we want to preserve DP-style privacy guarantees for future data releases in excess of this pre-existing public information. In this paper, we present a privacy formalism, -DP relative to Z, extending Pufferfish privacy, that accommodates DP-style semantics in the presence of public information. We introduce two mechanisms for releasing partially private data (PPD) and prove their desirable properties such as asymptotic negligibility of errors due to privacy and congeniality with as-is public information. We demonstrate theoretically and empirically how statistical inference from PPD degrades with post-processing, and propose alternative inference algorithms for estimating statistics from PPD. This collection of the framework, mechanisms, and inferential tools aims to help practitioners overcome the real logistical barriers introduced when public information is an unavoidable component of the data release process.

Mechanisms for Global Differential Privacy under Bayesian Data Synthesis
Speaker: **Jingchen (Monika) Hu, Vassar College**
We review, propose, and compare several Bayesian data synthesizers with different differential privacy guarantees that can be used by data stewards for microdata dissemination with privacy protection. The pseudo posterior mechanism achieves an asymptotic differential privacy guarantee and a variant of it can provide faster convergence. The newly proposed censoring mechanism embedded in the pseudo posterior mechanism censors the pseudo likelihood of every record within [exp(-epsilon/2), exp(epsilon/2)], which provides a stronger, non-asymptotic differential privacy guarantee. Through a series of simulation studies with bounded, univariate

data and an application to sample of the Survey of Doctoral Recipients where a beta regression synthesizer is utilized, we demonstrate that the pseudo posterior mechanism creates synthetic data with the highest utility at the price of a weaker, asymptotic privacy guarantee, while the censoring mechanism embedded in the pseudo posterior mechanism produces synthetic data with a stronger, non-asymptotic privacy guarantee at the cost of slightly reduced utility. The perturbed histogram is included for comparison.

TC14: Room: G16	Sequential Methods for Big Data Analysis	Chair: Alexander Tartakovsky

The Effect of Prior Information on Optimal Sequential Control of Generalized Multiple Testing Metrics
Speaker: **Jay Bartroff, University of Southern California**
The gamma-FDP and k-FWER multiple testing error metrics have become popular alternatives to the FDR and FWER. I will discuss asymptotically optimal procedures for controlling either of these metrics for sequential data in the context of prior information on the number of false null hypotheses.

Detection of Transient Changes and Application in Statistical Genetics
Speaker: **Michael Baron, American University**
We discuss methods of detecting multiple transient changes in long sequences of data, where the initial distribution can change at unknown times and later return to the original state. The number of changes, the moments of change, and durations of each period are unknown, although prior distributions of change-points and distribution parameters within each segment may be available. We propose change-point detection methods for different scenarios and explore their application to a problem of detecting instability regions in genome coverage processes. Accurate detection and estimation of changes in patterns of coverage depth and comparison of distributions within instability segments between healthy tissues and tumors is potentially becoming a powerful cancer diagnostics tool.

Point Processes with Infinite Intensity: a New Model for Bursty Spatial Sources
Speaker: **Sergei Zuev, Chalmers University**
Point processes gained recognition in modelling different aspects of modern systems: network infrastructure, population distribution, disease prevalence, mobile customers, etc. In many cases the underlying system shows highly irregular spatial characteristics. For instance, the number of requests to web-servers varies hugely, the number of phone calls during popular events increases in orders of magnitude compared to normal periods. Such phenomena call for development of point process models exhibiting such bursty behaviour, in time or in space. We introduce a wide class of point processes which generally have infinite expected number of points in a bounded domain, yet they appear unavoidably as a limit in various superposition schemes. We give insight into their cluster structure and further distributional properties and show how these can be estimated from data.

Optimal Joint Sequential Detection and Identification of Changes in Multiple Data Streams
Speaker: **Alexander Tartakovsky, AGT StatConsult**
Modern complex engineering and information systems generate large volumes of data. The objective is to develop novel optimal or nearly optimal decision-making strategies for rapid detection and identification of changes in multiple data streams or multiple channels for efficient Big Data analysis. We will discuss several challenging applications such as quick detection and isolation of anomalies associated with COVID, the appearance of near-Earth space objects, and attacks on computer networks. This is joint work with Prof. Serge Pergamenchtchikov from University of Rouen.

TC15: Room: G22	Advances in Directional Statistics	Chair: Thomas Verdebout

Finite Sample Smeariness for Directional Data
Speaker: **Stephan Huckeman, University of Goettingen**
Intrinsic means on circles, tori and spheres usually exhibit asymptotic normality scaled with root of sample size, just as their Euclidean kin. On the circle and on tori certain distributions can be constructed featuring arbitrary smaller rates, to some extent we can also do this on spheres. It turns out that this exotic phenomenon also affects large classes of rather common distributions, for example all distributions fully supported on a circle: for finite sample sizes, variances of intrinsic means scale with less than sample size. This phenomenon is called finite sample smeariness (FSS). FFS of Type 1 renders classical quantile based tests completely inapplicable, and for FSS of Type 2 they may apply only for very high sample sizes. In contrast, we show that suitably designed bootstrap tests keep the level. We explore this phenomenon of FSS and find it, among others, in directional wind data of European cities over the last 20 years. While quantile based tests, not correcting for FSS, find a multitude of significant wind changes, under our boostrap tests, this multitude condenses to a few years featuring significant wind changes. (This is joint work with Shayan Hundrieser and Benjamin Eltzner.)

Geological Features, Parallel Lines and Circular Statistics
Speaker: **Peter Jupp, University of St Andrews**
Some geological features appear to lie near equally-spaced parallel lines. If this impression is correct, it can be informative about

mechanisms that formed the features. Correctness of the impression can be assessed using quantal models in circular statistics analogous to those of Kendall (1974). The direction of the lines and the spacing between them can be estimated. This is joint work with Ian Goudie, Robert Goudie and Richard Batchelor.

Spatial Quantiles on the Hypersphere
Speaker: **Davy Paindaveine, University of Brusells**
A concept of quantiles for distributions on the unit hypersphere R^d is proposed. The innermost quantiles are Frechet medians, i.e. the L^1-analog of Frechet means. Since these medians may be non-unique, we define a quantile field around each such median m. The corresponding quantiles are directional in nature: they are indexed by a scalar order between 0 and 1 and a unit vector in the tangent space to the hypersphere at the median. To ensure computability in any dimension, our quantiles are essentially obtained by considering the Euclidean Chaudhuri spatial quantiles in a suitable stereographic projection of the hypersphere onto its tangent space at the median. Despite this link with their Euclidean antecedent, studying our quantiles requires understanding the nature of the Chaudhuri quantile in a version of the projective space where all points at infinity are identified. We thoroughly investigate the properties of the proposed quantiles, and study in particular the asymptotic behaviour of their sample versions, which requires controlling the impact of estimating the median. Our spherical quantile concept also allows for companion concepts of ranks and depth on the hypersphere. This is joint with Dimitri Konen.

Asymptotic Power of Sobolev Tests for Uniformity on Hyperspheres
Speaker: **Thomas Verdebout, University of Brusells**
One of the most classical problems in multivariate statistics is considered, namely, the problem of testing isotropy, or equivalently, the problem of testing uniformity on the unit hypersphere. Rather than restricting to tests that can detect specific types of alternatives only, we consider the broad class of Sobolev tests. While these tests are known to allow for omnibus testing of uniformity, their non-null behavior and consistency rates, unexpectedly, remain largely unexplored. To improve on this, we thoroughly study the local asymptotic powers of Sobolev tests under the most classical alternatives to uniformity, namely, under rotationally symmetric alternatives. We show in particular that the consistency rate of Sobolev tests does not only depend on the coefficients defining these tests but also on the derivatives of the underlying angular function at zero.

TC16: Room: G26	Statistical Inference for Continuous-time Stochastic Processes	Chair: Vladimir Minin

Pseudo-marginal Inference for CTMCs on Infinite Spaces via Monotonic Likelihood Approximations
Speaker: **Miguel Biron-Lattes, , University of British Columbia**
Bayesian inference for Continuous-Time Markov Chains (CTMCs) on countably infinite spaces is notoriously difficult because evaluating the likelihood exactly is intractable. One way to address this challenge is to first build a non-negative and unbiased estimate of the likelihood—involving the matrix exponential of finite truncations of the true rate matrix—and then to use the estimates in a pseudo-marginal inference method. In this work, we show that we can dramatically increase the efficiency of this approach by avoiding the computation of exact matrix exponentials. In particular, we develop a general methodology for constructing an unbiased, non-negative estimate of the likelihood using doubly-monotone matrix exponential approximations. We further develop a novel approximation in this family—the skeletoid—as well as theory regarding its approximation error and how that relates to the variance of the estimates used in pseudo-marginal inference. Experimental results show that our approach yields more efficient posterior inference for a wide variety of CTMCs. This is joint work with Alexandre Bouchard.

Variational Inference of Evolutionary Trees
Speaker: **Julia Palacios, Stanford University**
Rooted binary trees are mathematical objects of great importance used to model hierarchical data and evolutionary relationships with applications in many fields including evolutionary biology and genetic epidemiology. Bayesian phylogenetic inference usually explores the posterior distribution of trees via Markov Chain Monte Carlo methods, however assessing uncertainty and summarizing distributions remains challenging for these types of structures. Here, we exploit recently proposed distance metrics of unlabeled binary trees and genealogies (equipped with branch lengths) to define the Fréchet mean and variance as summaries of these tree distributions and provide an efficient combinatorial optimization algorithm for their computation. We propose a new class of models based on such distances that efficiently allows variational inference of model parameters in phylodynamics and phylogenetics. We show the applicability of our proposed methods for studying SARS-CoV-2 from molecular sequences of infected individuals.

Functional Models for Time Varying Random Objects
Speaker: **Paromita Dubey, University of Southern California**
In recent years, samples of time-varying object data such as time-varying networks that are not in a vector space have been increasingly collected. These data can be viewed as elements of a general metric space that lacks local or global linear structure and therefore common approaches that have been used with great success for the analysis of functional data, such as functional principal component analysis, cannot be applied directly.

In this talk, I will propose some recent advances along this direction. First, I will discuss ways to obtain dominant modes of variations in time varying object data. I will describe metric covariance, a new association measure for paired object data lying in a metric space (Ω, d) that we use to define a metric auto-covariance function for a sample of random Ω-valued curves, where Ω will not have a vector space or manifold structure. The proposed metric auto-covariance function is non-negative definite when the squared metric d^2 is of negative type. The eigenfunctions of the linear operator with the metric auto-covariance function as the kernel can be used as building blocks for an object functional principal component analysis for Ω-valued functional data, including time-varying probability distributions, covariance matrices and time-dynamic networks. Then I will describe how to obtain analogues of functional principal components for time-varying objects by applying weighted Fréchet means which serve as projections of the random object trajectories in the directions of the eigenfunctions, leading to Ω-valued Fréchet integrals. This talk is based on joint work with Hans-Georg Müller.

Fitting Stochastic Epidemic Models to Gene Genealogies Using Linear Noise Approximation
Speaker: **Vladimir Minin, University of California, Irvine**
Phylodynamics is a set of population genetics tools that aim at reconstructing demographic history of a population based on molecular sequences of individuals sampled from the population of interest. One important task in phylodynamics is to estimate changes in (effective) population size. When applied to infectious disease sequences such estimation of population size trajectories can provide information about changes in the number of infections. To model changes in the number of infected individuals, current phylodynamic methods use non-parametric approaches (e.g., Bayesian curve-fitting based on change-point models or Gaussian process priors), parametric approaches (e.g., based on differential equations), and stochastic modeling in conjunction with likelihood-free Bayesian methods. The first class of methods yields results that are hard to interpret epidemiologically. The second class of methods provides estimates of important epidemiological parameters, such as infection and removal/recovery rates, but ignores variation in the dynamics of infectious disease spread. The third class of methods is the most advantageous statistically, but relies on computationally intensive particle filtering techniques that limits its applications.We propose a Bayesian model that combines phylodynamic inference and stochastic epidemic models, and achieves computational tractability by using a linear noise approximation (LNA) — a technique that allows us to approximate probability densities of stochastic epidemic model trajectories. LNA opens the door for using modern Markov chain Monte Carlo tools to approximate the joint posterior distribution of the disease transmission parameters and of high dimensional vectors describing unobserved changes in the stochastic epidemic model compartment sizes (e.g., numbers of infectious and susceptible individuals). In a simulation study, we show that our method can successfully recover parameters of stochastic epidemic models.We apply our estimation technique to Ebola genealogies estimated using viral genetic data from the 2014 epidemic in Sierra Leone and Liberia.

Thursday 30-06-2022	13:30-15:30	Topic Contributed Sessions (TC17)

TC17: Room: G21A	Machine Learning Methods in Biomedical Studies	Chair: Rui Feng

Bayesian Multiple Instance Learning with Application to Cancer Detection Using TCR Repertoire Sequencing Data
Speaker: **Sherry Wang, Southern Methodist University University of Texas Southwestern Medical Center**
As a branch of machine learning, multiple instance learning (MIL) learns from a collection of labeled bags, each containing a set of instances. The learning process is weakly supervised due to ambiguous instance labels. Since its emergence, MIL has been applied to solve various problems including content-based image retrieval, object tracking/detection, and computer-aided diagnosis. In biomedical research, the use of MIL has been focused on medical image analysis and molecule activity prediction. We explore a novel and important biomedical application of MIL, cancer detection using T-cell receptor (TCR) sequences, where we apply state-of-the-art MIL methods and examine their performance. We further devise a Bayesian MIL method based on hierarchical probit regression (MICProB), which contributes a significant portion to the suite of statistical methodologies for MIL. We demonstrate the competitiveness of MICProB in simulation and real data examples. More importantly, MICProB is capable of identifying primary instances and enjoys great advantages in providing a transparent model structure, straightforward statistical inference, and favorable explainability, while none of the existing optimization-based approaches developed in the computer science domain can.

Network Functional Varying Coefficient Model
Speaker: **Yanyuan Ma, Pennsylvania State Univerity**
We consider functional responses with network dependence observed for each individual at irregular time points. To model both the inter-individual dependence as well as within-individual dynamic correlation, we propose a network functional varying coefficient (NFVC) model. The response of each individual is characterized by a linear combination of responses from its connected nodes and its own exogenous covariates. All the model coefficients are allowed to be time dependent. The NFVC model adds to the richness of both the classical network autoregression model and the functional regression models. To overcome the complexity caused by the network inter-dependence, we devise a special non-parametric least squares type estimator, which is feasible when the responses are observed at irregular time points for different individuals. The estimator takes advantage of the sparsity of the network structure to reduce the computational burden. To further conduct the functional principal component analysis, a novel within-individual covariance function estimation method is proposed and studied. Theoretical properties of our estimators are analyzed, which involve techniques related to empirical processes, nonparametrics, functional data analysis and various concentration inequalities. We analyze a social network data to illustrate the powerfulness of the proposed procedure.

RNA Sequencing Differential Analysis Using Machine Learning Algorithms
Speaker: **Dongmei Li, University of Rochester**
RNA sequencing has been widely used in biomedical research to identify novel genes and cell types associated with different treatment conditions or diseases. With the decreasing of sequencing costs in recent years, large amount of sequencing data has been generated from NIH funded consortium grants across multiple institutions. Given the availability of large sample size in sequencing data, application of machine learning and deep learning algorithms into RNA sequencing data analyses getting more popular in recent years. Our study focuses on examining the performance of machine learning and deep learning algorithms in RNA sequencing differential analysis comparing with popular statistical methods, through both simulation studies and real data examples. The performance of different algorithms will be indicated by the false discovery rate control, power, and stability.

Layer-wise Propagation Relevance for Deep Learning with Graph Structure
Speaker: **Rui Feng, University of Pennsylvania**
Neural networks have demonstrated strong capabilities of discovering potential complex patterns for predicting outcomes. They typically need to be trained on a large amount of data. However, biomedical studies often have limited sample sizes but large numbers of feature variables that may correlate, interact, and jointly affect outcomes. Previously, we proposed Peel Learning (PL), a novel neural network that incorporates the prior relationship among features. In each layer of learning, overall structure is peeled into multiple local substructures. Within the substructure, dependency among features is reduced by projecting children features onto parents' space and removing the redundant components. The overall structure is gradually reduced in size over layers and the parameters are optimized through backpropagation. PL has showed improved prediction accuracy, especially in small data, compared with several existing methods. To understand the contribution of each predictor to individual prediction or overall model fitting, we developed a corresponding relevance index in the framework of Layer-wise Relevance Propagation (LRP) in this study. Its performance was evaluated in simulated and real data.

Monday 27-06-2022	10:30-12:30	Contributed Talks

CTS3: Room: G21A Clinical Data, Missing and Censored Data Chair: Yichuan Zhao

Optimal Design Strategies in Clinical Trials with Variance Heterogeneity
Speaker: **Lida Mavrogonatou, University of Cambridge**
Optimal patient allocation strategies in clinical trials with variance heterogeneity. Abstract: In a multi-objective experiment such as a clinical trial, optimisation with respect to different objectives, such as efficiency, statistical power and patient outcome, can result in conflicting patient allocation strategies. Typically, such conflict is resolved with the adoption of a balanced allocation as it has been shown to maximise both estimation accuracy and power under certain assumptions while maintaining a controlled approach with respect to the allocation of patients to each treatment. Homogeneity in the variances of patient responses is one of these assumptions but can be easily violated if treatments differ in outcome mean thus compounding the conflict between the alternative allocation strategies further. We explore the interrelation of the objectives of interest under different allocation strategies and make suggestions for dealing with potential conflicts which is a problem frequently encountered in clinical trials. We consider blocked designs, allowing for heterogeneous variances not only among treatments but also blocks. Lastly, we present a Bayesian response-adaptive framework for the implementation of the resulting patient allocations.

Assessing the Use of Variational Inference for Large Real-world Clinical Data
Speaker: **Brian Buckley, University College Dublin**
With growing acceptance by clinical regulators of the value of real-world evidence to supplement clinical trials, there is increasing interest in the use of Bayesian analysis for real-world clinical studies. This interest has been limited by the computational challenges applying the Markov-chain Monte-carlo (MCMC) approach to large real-world clinical data that incorporate missing and adulterated data along with many discrete variables. MCMC is considered the gold standard for Bayesian inference because, in the limit, MCMC is guaranteed to eventually converge to the true posterior distribution. Approximate Bayesian inference via optimization of the Variational evidence lower bound, usually called Variational Inference (VI), has been successfully demonstrated for other applications. We investigate the performance and characteristics of currently available R and Python VI software implementations for clinical data. Four R implementations and four Python implementations are compared. The implementations include several algorithms for VI: coordinate ascent mean-field VI, stochastic VI, automatic differentiation VI and two application-specific R packages. Since clinical data analyses commonly employ logistic regression of the response variable we choose this approach for comparative purposes. The standard R Generalized Linear Model function glm() was used to validate a MCMC logistic model as the comparison baseline. The R package rjags was used to implement the MCMC baseline model. Several datasets were used to fully explore the performance and characteristics of the methods and implementations over a wide variety of realistic clinical data. We used Pima Indian Type 2 Diabetes data as a standard benchmark data set for the initial comparison. The Pima Indian data contains 768 observations with 9 continuous variables and no missing data. The methods are further applied to a large real-world clinical dataset, Optum EHR, containing 1,133,214 paediatric patients with a rare Type 1 form of diabetes, extending the analysis to complex data containing discrete and continuous variables incorporating missing and adulterated data. We conclude the study with a simulation analysis to explore in more detail where particular differences occur. We find that automatic VI approaches require more effort and technical knowledge to set up for accurate posterior estimation and are very sensitive to stopping time. The stopping time characteristic require very high iteration rates that negate some of the advantages of automatic VI, with algorithm run times longer than MCMC in one case. We find that several data characteristics common in clinical data, for example a very high proportion of zeroes, significantly affect the posterior accuracy of automatic VI methods compared with conditionally conjugate mean-field methods. We propose further investigation of automatic VI methods with a view to improving posterior accuracy and computational runtime from default settings.

Seeing the Unseen - a Generalized Scheme for Missing Mass Estimation
Speaker: **Amichai Painsky, Tel Aviv University**
Consider a finite sample from an unknown distribution over a countable alphabet. The missing mass refers to the probability of symbols that do not appear in the sample. Estimating the missing mass is a basic problem in machine learning and related fields, which dates back to the early work of Laplace, and the more recent seminal contribution of Good and Turing. The missing mass problem has many applications in a variety of domains. For example, given the observed population of Covid mutations, what is the probability that we encounter a new variant that has not been seen before? In this work we introduce a generalized framework for missing mass estimation. Our framework provides novel risk bounds and improves upon currently known estimation schemes. Importantly, it is easy to apply and does not require additional modeling assumptions. This makes it a favorable choice for practical applications. Furthermore, we show that by utilizing our scheme, we improve the estimation accuracy of large alphabet probability distributions.

Estimating Heterogeneous Treatment Effects with Right-censored Data via Causal Survival Forests
Speaker: **Yifan Cui, National University of Singapore**
Forest-based methods have recently gained popularity for non-parametric treatment effect estimation. Building on this line of work, we introduce causal survival forests, which can be used to estimate heterogeneous treatment effects in survival and observational settings where outcomes may be right-censored. The approach relies on orthogonal estimating equations to robustly adjust for both

censoring and selection effects under unconfoundedness. In the experiments, we find our approach to perform well relative to a number of baselines.

Instrumental Variable Quantile Regression under Random Right Censoring
Speaker: **Lorenzo Tedesco, KU Leuven**
The paper studies a quantile regression model with endogenous variables and random right censoring. The endogeneity issue is solved using instrumental variables. The proposed semi- parametric model assumes that the structural quantile of the outcome variable is linear in the covariates and the censoring variable is independent of the other variables. No strict conditions on the distributions of the covariates are required. Identification results are presented. We propose an estimation procedure for the parameter of interest and derive its asymptotic properties. The finite sample performance of the estimator is evaluated through numerical simulations. The method is illustrated by an application to a bone marrow transplant data set.

Monday 27-06-2022	13:30-15:30	Contributed Talks

CTS2: Room: G16 Inference on Change-Points Chair: Yudong Chen

High-Dimensional Data Segmentation Under a Sparse Regression Model
Speaker: **Dominic Owens, University of Bristol**
We develop a data segmentation methodology in the high-dimensional regression setting, where the parameters are permitted to undergo multiple changes. The proposed methodology, MOSUMSR, proceeds in two steps: First, it adopts a moving window-based procedure that compares the local parameter estimates to detect and locate multiple change points, which is followed by a location refinement step. We show that the combined methodology consistently estimates both the total number and the locations of the change points and further, it achieves minimax optimality (up to logarithmic factors) both in separation and localisation rates. Also, it is competitive computationally thanks to a subsampling strategy that reduces the number of Lasso-based estimators required for the first-step screening. We further propose a multiscale extension of MOSUMSR, and show it performs well in comparison with existing methodologies on simulated datasets, and in applications to house price and equity premium datasets. An implementation of the proposed methodology is available at https://github.com/Dom-Owens-UoB/sparse_mosum.

Multiple Hypothesis Testing from the Change-point Detection Viewpoint – Surprises and New Results
Speaker: **Anica Kostic, London School of Economics and Political Science**
Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses is an important problem in multiple testing literature. In the sequence of sorted p-values (p-value plot), false null p-values tend to be smaller and concentrated at the beginning. This suggests approximate piecewise linear shape of the p-value plot, with a change-point separating small from large p-values. Our proposed method for estimating the false null proportion (Difference Of Slopes) utilizes the idea of estimating that change-point in slope in the p-value plot. The resulting estimator is a conservative estimator of the proportion, which is a desirable property when using adaptive FDR procedures. It performs best in sparse cases, when the false null proportion is small. A possible extension involves estimating multiple change-points and thus grouping p-values based on their local FDR value.

Cross-validation for Change-point Regression: Pitfalls and Solutions
Speaker: **Florian Pein, Lancaster University**
Cross-validation is the standard approach for tuning parameter selection in many non-parametric regression problems. However its use is less common in change-point regression, perhaps as its prediction error-based criterion may appear to permit small spurious changes and hence be less well-suited to estimation of the number and location of change-points. We show that in fact the problems of cross-validation with squared error loss are more severe and can lead to systematic under- or over-estimation of the number of change-points, and highly suboptimal estimation of the mean function in simple settings where changes are easily detectable. We propose two simple approaches to remedy these issues, the first involving the use of absolute error rather than squared error loss, and the second involving modifying the holdout sets used. For the latter, we provide conditions that permit consistent estimation of the number of change-points for a general change-point estimation procedure. We show these conditions are satisfied for optimal partitioning using new results on its performance when supplied with the incorrect number of change-points. Numerical experiments show that the absolute error approach in particular is competitive with common change-point methods using classical tuning parameter choices when error distributions are well-specified, but can substantially outperform these in misspecified models. This is joint work with Rajen Shah.

Robust Inference for Change-points using Confidence Sets
Speaker: **Shakeel Gavioli-Akilagun, London School of Economics**
Multiple change-point detection has become popular with the routine collection of complex non stationary time series. An equally important but comparatively neglected question concerns quantifying the level of uncertainty around each putative change point. Though a handful of procedures exist in the literature, most all make assumptions on the density of the contaminating noise which are impossible to verify in practice. Moreover, most procedures are only applicable in the canonical piecewise-constant mean (median, or quantile) setting. We present a procedure which, under minimal assumptions, returns localised regions of a data sequence which must contain a change point at some global significance level chosen by the user. Our procedure is based on properties of confidence sets for the underlying regression function obtained by inverting certain multi-resolution tests, and is immediately applicable to change points in higher order polynomials. We will discuss some appealing theoretical properties of our procedure, and show its good practical performance on real and simulated data.

Variance Change-point Detection with Credible Sets
Speaker: **Lorenzo Cappello, Universitat Pompeu Fabra**
We introduce a novel approach to detect changes in the variance of a Gaussian sequence model, focusing on quantifying the uncertainty in the change-point locations. We do that by framing the problem as a product of multiple single changes in the scale parameter. We fit the model through an iterative procedure similar to what is done for additive models. The novelty is that each iteration returns a probability distribution on time instances, which captures the uncertainty in the change point location. Leveraging a recent result in the literature, we can show that our proposal is a variational approximation of the exact model posterior distribution.

We study the convergence of the algorithm and the change-point localization rate. We illustrate our approach in simulations studies.

Inference in High-dimensional Online Change-point Detection
Speaker: **Yudong Chen, London School of Economics**
We introduce and study two new inferential challenges associated with the sequential detection of change in a high-dimensional mean vector. First, we seek a confidence interval for the change-point, and second, we estimate the set of indices of coordinates in which the mean changes. We propose an online algorithm that produces an interval with guaranteed nominal coverage, and whose length is, with high probability, of the same order as the average detection delay, up to a logarithmic factor. The corresponding support estimate enjoys control of both false negatives and false positives. Simulations confirm the effectiveness of our methodology, and we also illustrate its applicability on both US excess deaths data from 2017–2020 and SP 500 data from the 2007–2008 financial crisis.

CTS23: Room:G21A	**Complex Probability Models**	**Chair: Alexandre Pannier**

Level Densities for General β-ensembles: An Operator-valued Free Probability Perspective
Speaker: **Andrej Srakar, IER and University of Ljubljana**
Random point processes corresponding to β-ensembles for arbitrary $\beta > 0$, or, equivalently, log gases at inverse temperature β, are being subject to intense study. The orthogonal, unitary, and symplectic ensembles (β = 1, 2, or 4, respectively) are now well understood, but other values of β are believed to also be relevant in theory (e.g. relevant for the study of self-adjoint and Schrödinger operators) and applications (e.g. in logistics). For certain rational values of β, β-ensembles are related to Jack polynomials, but for general β much less is known. In a seminal article, Dumitriu and Edelman (2002) constructed tridiagonal random matrix models for general β-Hermite andβ-Laguerre ensembles and defined open problems for research on general β-ensembles, including finding a unified formula for the level density in general β-case. In general, level density is defined as distribution of a random eigenvalue of an ensemble (by the Wigner semicircular law, the limiting distribution of the eigenvalue is semicircular). We exploit the product nature of Dumitriu and Edelman's construction of tridiagonal random matrix models and derive the formula for the level density in the general β-case depending on the multivariate Fuss-Narayana polynomials and concepts from operator-valued free probability theory. We study perturbation invariability of the level densities (Wang and Yan, 2005; Kozhan, 2017) and discuss extensions to problems of sampling general β-ensembles (referring to Li and Menon, 2013; Olver et al., 2015; Srakar and Verbic, 2020) and limiting entropy in β-ensembles related point processes (Mészaros, 2020).

Permutations Avoiding a Pattern of Length Three under Mallows Distributions
Speaker: **Ross Pinsky, Technion-Israel Institute of Technology**
The Mallows distributions with parameter $q > 0$ on the permutations in S_n is the distribution for which the probability of $\sigma \in S_n$ is proportional to $q^{\mathrm{Inv}(\sigma)}$, where $\mathrm{Inv}(\sigma)$ counts the number of inversions in σ. We consider permutations avoiding a pattern of length three under the family of Mallows distributions. In particular, for any pattern $\tau \in S_3 \setminus \{321\}$, we obtain rather precise results on the asymptotic probability as $n \to \infty$ that a permutation under the Mallows distribution with parameter $q \in (0,1)$ avoids the pattern. By a duality between the parameters q and $\frac{1}{q}$, we also obtain rather precise results on the above probability for $q > 1$ and any pattern $\tau \in S_3 \setminus \{123\}$.

Coexistence in Discrete Time Multi type Competing Frog Models
Speaker: **Rishideep Roy, IIM Bangalore**
We study coexistence in discrete time multi-type frog models. We first show that for two types of particles on $\mathbb{Z}^d$, for $d \geq 2$, for any jumping parameters $p_1, p_2 \in (0,1]$, coexistence occurs with positive probability for sufficiently rich deterministic initial configuration. We extend this to the case of random distribution of initial particles. We study the question of coexistence for multiple types and show positive probability coexistence of 2^d types on $\mathbb{Z}^d$ for rich enough initial configuration. We also show an instance of infinite coexistence on $\mathbb{Z}^d$ for $d \geq 3$ provided we have sufficiently rich initial configuration.

Nonlinear Semigroups and their Generators with respect to Γ-convergence
Speaker: **Jonas Blessing, University of Konstanz**
Motivated by model uncertainty and stochastic optimal control problems, we develop a general theory of non-linear semigroups. In contrast to the linear theory, the domain of the generator is, in general, not invariant under semigroup. In order to overcome this issue, we consider so-called invariant Lipschitz sets which turn out to be a suitable domain for a weaker notion of the generator. This is defined using Γ-convergence in an appropriate function space. Furthermore, we show that the Γ-generator uniquely characterizes the non-linear semigroup and is determined by the evaluation at smooth functions. In particular, we obtain that different approximation schemes lead to the same semigroup. As an application, we show that stochastic optimal control problems can be approximated by simple deterministic controls. Moreover, non-parametric uncertainty over Wasserstein balls can be as well approximated by a suitable parametrization. This talk is based on a joint work with Robert Denk, Michael Kupper (both University of Konstanz) and Max Nendel (Bielefeld University).

Pathwise Large Deviations for White Noise Chaos Expansions
Speaker: **Alexandre Pannier, Imperial College London**
We consider a family of continuous processes, measurable with respect to a white noise measure, and characterised by their Wiener

chaos expansion. We provide sufficient conditions for their large deviations principle to hold in path space, thereby refreshing a problem left open by Pérez-Abreu in 1993 in the Brownian motion case. The proof is based on the weak convergence approach to large deviations: it involves demonstrating the convergence in distribution of certain perturbations of the original process, and thus the main difficulties lie in analysing and controlling the perturbed multiple stochastic integrals. Moreover, adopting this representation offers a new perspective on pathwise large deviations and induces a variety of applications thereof.

Tuesday 28-06-2022	10:30-12:30	Contributed Talks

CTS4: Room:G16	Modern Regression	Chair: Alastair Young

Splitting Strategies for Selective Inference
Speaker: **Alastair Young, Imperial College London** We consider the problem of providing valid inference for a selected parameter in a sparse regression setting. Many approaches have been proposed in recent years to overcome the bias generated by the selection step and ensure inferential validity. Here, we consider a simple alternative to data splitting based on randomising the response vector, which allows for higher selection and inferential power than the former and is applicable with an arbitrary selection rule. This is joint work with Daniel Garcia Rasines (ICMAT, Madrid).

Regression Diagnostics meets Forecast Evaluation: Conditional Calibration, Reliability Diagrams, and Coefficient of Determination
Speaker: **Johannes Resin, Heidelberg Institute for Theoretical Studies**
A common principle in model diagnostics and forecast evaluation is that fitted or predicted distributions ought to be reliable, ideally in the sense of auto-calibration, where the outcome is a random draw from the posited distribution. Extant practice relies on weaker, unconditional notions, such as probabilistic calibration. Conditional concepts give rise to hierarchies of calibration. A predictive distribution is conditionally T-calibrated if it can be taken at face value in terms of the functional T. For identifiable functionals T, such as moments, quantiles, and expectiles, auto-calibration implies T-calibration. The notion of T-calibration also applies to stand-alone point forecasts or regression output. We introduce population versions of T-reliability diagrams and revisit a score decomposition into miscalibration (MCB), discrimination (DSC), and uncertainty (UNC). Stable and efficient estimators of T-reliability diagrams and score components arise via nonparametric isotonic regression and the pool-adjacent-violators algorithm. For in-sample model diagnostics, we propose a universal coefficient of determination that nests and reinterprets the classical R^2 in least squares (mean) regression and its natural analogue R^1 in quantile regression, yet applies to T-regression in general. Joint work with Tilmann Gneiting.

Least Squares Estimation of a Quasiconvex Regression Function
Speaker: **Rohit Patra, University of Florida**
We develop a new approach for the estimation of a multivariate function based on the economic axioms of quasiconvexity (and monotonicity). On the computational side, we prove the existence of the quasiconvex constrained least squares estimator (LSE) and provide a characterization of the function space to compute the LSE via a mixed integer quadratic programme. On the theoretical side, we provide finite sample risk bounds for the LSE via a sharp oracle inequality. Our results allow for errors to depend on the covariates and to have only two finite moments. We illustrate the superior performance of the LSE against some competing estimators via simulation. Finally, we use the LSE to estimate the production function for the Japanese plywood industry and the cost function for hospitals across the US.

Stochastic Tree Ensembles for Regularized Nonlinear Regression
Speaker: **Jingyu He, City University of Hong Kong**
This paper develops a novel stochastic tree ensemble method for nonlinear regression, named XBART. By combining regularization and stochastic search strategies from Bayesian modeling with computationally efficient techniques from recursive partitioning algorithms, XBART attains state-of-the-art performance at prediction and function estimation. Simulation studies demonstrate that XBART provides accurate pointwise estimates of the mean function and does so faster than popular alternatives, such as BART, XGBoost, and neural networks on a variety of test functions. Additionally, it is demonstrated that using XBART to initialize the standard BART MCMC algorithm considerably improves credible interval coverage and reduces total run-time. Finally, we establish basic theoretical results: the single tree version of the model is asymptotically consistent and the Markov chain produced by the ensemble version of the algorithm has a unique stationary distribution.

Predictive Model Degrees of Freedom in Linear Regression
Speaker: **Yoonkyung Lee, Ohio State University**
Overparametrized interpolating models have drawn increasing attention from machine learning. Some recent studies suggest that regularized interpolating models can generalize well. This phenomenon seemingly contradicts the conventional wisdom that interpolation tends to overfit the data and performs poorly on test data. Further, it appears to defy the bias-variance trade-off. As one of the shortcomings of the existing theory, the classical notion of model degrees of freedom fails to explain the intrinsic difference among the interpolating models since it focuses on estimation of in-sample prediction error. This motivates an alternative measure of model complexity which can differentiate those interpolating models and take different test points into account. In particular, we propose a measure with a proper adjustment based on the squared covariance between the predictions and observations. Our analysis with least squares method reveals some interesting properties of the measure, which can reconcile the "double descent" phenomenon with the classical theory. This opens doors to an extended definition of model degrees of freedom in modern predictive settings.

Elastic Gradient Descent and Elastic Gradient Flow: LARS Like Algorithms Approximating Solution Paths of Elastic Net
Speaker: **Oskar Allerbo, Chalmers University and University of Gothenburg**

It is well known that early stopping can work as a regularizer when training statistical models. For linear regression, the connection between the optimization algorithm forward stagewise regression and the lasso penalty is well established, while more recently the connection between gradient descent with early stopping and ridge regression was investigated. The elastic net is a convex combination of lasso and ridge regression which poses the question “Can the elastic net be formulated as an optimization algorithm?” Here we introduce elastic gradient descent, an iterative optimization algorithm with solution paths approximating those of the elastic net. We furthermore let optimization step size go to zero, obtaining a piecewise analytical solution, which we refer to as elastic gradient flow. We compare elastic gradient descent and the elastic net on real an simulated data, and show that elastic gradient descent provides significantly better sensitivity and mean squared prediction error, than the elastic net, without sacrificing specificity. Furthermore, when comparing the solution paths of elastic gradient descent and the elastic net, elastic gradient descent proposes fewer alternative models along the path, thus providing more stable model selection results.

Tuesday 28-06-2022	13:30-15:30	Contributed Talks

CTS6: Room: G16	**Stochastic Differential Equations**	**Chair: Chengcheng Ling**

Taming Singular Stochastic Differential Equations: A Numerical Method
Speaker: **Chengcheng Ling, TU Berlin**
We consider a generic and explicit tamed Euler–Maruyama scheme for multidimensional time-inhomogeneous stochastic differential equations with multiplicative Brownian noise. The diffusive coeffcient is uniformly elliptic, Hölder continuous and weakly differentiable in the spatial variables while the drift satisfies the Ladyzhenskaya–Prodi–Serrin condition, as considered by Krylov and Röckner (2005). In the discrete scheme, the drift is tamed by replacing it by an approximation. A strong rate of convergence of the scheme is provided in terms of the approximation error of the drift in a suitable and possibly very weak topology. A few examples of approximating drifts are discussed in detail. The parameters of the approximating drifts can vary and—under suitable conditions—be fine-tuned to achieve the standard 1/2-strong convergence rate with a logarithmic factor.

A Law of Large Numbers for interacting diffusions via a mild formulation
Speaker: **Florian Bechtold, Bielefeld University**
Consider a system of n weakly interacting particles driven by independent Brownian motions. In many instances, it is well known that the empirical measure converges to the solution of a partial differential equation, usually called McKean-Vlasov or Fokker-Planck equation, as n tends to infinity. We propose a relatively new approach to show this convergence by directly studying the stochastic partial differential equation that the empirical measure satisfies for each fixed n. Under a suitable control on the noise term, which appears due to the finiteness of the system, we are able to prove that the stochastic perturbation goes to zero, showing that the limiting measure is a solution to the classical McKean-Vlasov equation. In contrast with known results, we do not require any independence or finite moment assumption on the initial condition, but the only weak convergence. The evolution of the empirical measure is studied in a suitable class of Hilbert spaces where the noise term is controlled using two distinct but complementary techniques: rough paths theory and maximal inequalities for self-normalized processes. Based on joint work with Fabio Coppini.

On Regularization by Noise of an Averaged Version of the Navier–Stokes Equations
Speaker: **Theresa Lange, Bielefeld University**
In [T16], the author constructs an averaged version of the deterministic three-dimensional Navier-Stokes equations (3D NSE) which experiences blow-up in finite time. In the last decades, various works have studied suitable perturbations of singular deterministic PDEs in order to prevent or delay such behavior. A promising example is given by a particular choice of stochastic transport noise closely studied in [FL21] and in [FGL21]. In this talk, we analyze the model in [T16] in view of these results and discuss the regularization skills of this noise in the context of the averaged 3D NSE. This is joint work with Martina Hofmanová.

- [T16] T. Tao, Finite time blowup of an averaged three-dimensional Navier-Stokes equation. Journal of the American Mathematical Society, Vol. 29, No. 3, pp. 601-674 (2016)
- [FL21] F. Flandoli, D. Luo, High mode transport noise improves vorticity blow-up control in 3D Navier-Stokes equations. Probability Theory and Related Fields, Vol. 180, No. 1, pp. 309-363 (2021)
- [FGL21] F. Flandoli, L. Galeati, D. Luo, Delayed blow-up by transport noise. Communications in Partial Differential Equations, Vol. 46, No. 9, pp. 1757-1788 (2021)

The Galerkin Analysis for the Random Periodic Solution of Semilinear Stochastic Evolution Equations
Speaker: **Yue Wu, University of Strathclyde**
Periodicity is widely exhibited in a large number of natural phenomena like oscillations, waves, or even lying behind many complicated ensembles such as biological and economic systems. However, periodic behaviors are often found to be subject to a random perturbation or under the influence of noise. It was only until recently that the random periodic solution was endowed with a proper definition, which is compatible with definitions of both the stationary solution and the deterministic periodic solution. In general, random periodic solutions cannot be solved explicitly. One may treat the numerical approximation that stays sufficiently close to the true solution as a good substitute to study stochastic dynamics. It is worth mentioning here that this is a numerical approximation of an infinite time horizon problem. In this talk, we will address the numerical analysis of the random periodic solution to the semilinear stochastic evolution equations. We propose a Galerkin-type exponential integrator scheme and establish its convergence rate of the strong error to the mild random periodic solution. We finally conclude with the best order of convergence that is arbitrarily close to 0.5.

Nonuniqueness in law for stochastic hypodissipative Navier-Stokes equations
Speaker: **Andre Schenke, Bielefeld University**
We study the incompressible hypodissipative Navier–Stokes equations with dissipation exponent $0 < \alpha < 1/2$ on the three-dimensional torus perturbed by an additive Wiener noise term and prove the existence of an initial condition for which distinct probabilistic weak solutions exist. To this end, we employ convex integration methods to construct a pathwise probabilistically strong solution, which violates a pathwise energy inequality up to a suitable stopping time. This paper seems to be the first in which such solutions are constructed via Beltrami waves instead of intermittent jets or flows in a stochastic setting. Joint work with Marco Rehmeier.

Wednesday 29-06-2022	10:30-12:30	Contributed Talks

CTS5: Room: G03	Bayesian Inference	Chair: Konstantinos Kalogeropoulos

Stochastic Approximation Techniques for Bayesian Uncertainty Directed Trial Designs
Speaker: **Sandra Fortini, University of Bocconi**
The power of Stochastic Approximation in the study of the asymptotic behavior of trial designs has been illustrated in Laruelle and Pagès (2013, Ann. Appl. Prob.), and further exploited in Baldi and Zagoraiou (2015, Bernoulli). However, Bayesian uncertainty directed trial designs (Ventz et al. 2018, JASA) pose a new challenge since in this case the randomization probabilities depend on the posterior distribution of the unknown parameters, as well as on past allocations. We show how Stochastic Approximation can be designed to control the estimation process and the randomization probabilities at the same time. Convergence and asymptotic normality of the allocation proportions and of the randomization probabilities are proved under mild assumptions. Explicit expressions are given for limits and for asymptotic variances. The practical implications for clinical trials designs are discussed. The talk is based on Bonsaglio et al. (2022, JSPI).

Robust Estimation in the Regression Setting
Speaker: **Yannick Baraud, University of Luxembourg**
In a regression setting, we present some non-asymptotic results on the performance of a new class of estimators, called rho-estimators, which possess the properties of being optimal and robust.

Robust Estimation of a Regression Function in Exponential Families
Speaker: **Juntong Chen, University of Luxembourg**
We consider the problem of estimating a regression function when the distribution of the data is modelled by an exponential family. Several interesting problems are under this setting, for example logit, Poisson and exponential regressions. Our estimation strategy is based on Rho-estimation and we present a non-asymptotic exponential inequality for the deviation of their risk. We deduce from this inequality that the estimator is robust to contamination, the presence of outliers and model misspecification. We also provide a uniform risk bound over the class of Hölderian functions and prove the optimality of the estimator over this class up to a logarithmic factor. Finally, we carry out a simulation study in order to compare the performance of Rho-estimators to the maximum likelihood estimator and median-based ones.

Sequential Bayesian Learning for Hidden Semi-Markov Models
Speaker: **Patrick Aschermayr, London School of Economics and Political Science**
In this paper, we explore the sequential estimation setting of state space models, in particular the Hidden semi-Markov Model (HSMM), a flexible extension of the well-known Hidden Markov Model (HMM) that allows the underlying stochastic process to be a semi-Markov chain. HSMMs are typically less often used than their basic HMM counterpart due to the increased computational challenges to evaluate the likelihood function. Moreover, while both models are of sequential in nature, parameter estimation is mainly conducted via batch estimation methods. A major motivation of this paper is thus to provide methods to estimate HSMMs (1) in a computationally feasible time and (2) in a sequential setting. We provide and verify an efficient computational scheme for Bayesian parameter estimation on HSMMs. Moreover, we demonstrate predictive model performance by estimating the parameters for a financial times series for HSMMs against other popular discrete State Space Models (SSM), and show how this algorithm can be used for classification and model selection.

Sequential Learning and Economic Benefits from Dynamic Term Structure Models
Speaker: **Konstantinos Kalogeropoulos, London School of Economics and Political Science**
This work explores the statistical and economic importance of restrictions on the dynamics of risk compensation, from the perspective of a real-time Bayesian learner who predicts bond excess returns using a dynamic term structure model (DTSM). We propose a novel methodological framework that successfully handles sequential model search and parameter estimation over the restriction space landscape in real time, allowing investors to revise their beliefs when new information arrives, thus informing their asset allocation and maximizing their expected utility. Our setup provides the entire predictive density of returns, allowing us to revisit the evident puzzling behaviour between statistical predictability and meaningful out-of-sample economic benefits for bond investors. Empirical results reveal the importance of different sets of restrictions across market conditions and monetary policy actions. Furthermore, our results reinforce the argument of sparsity in the market price of risk specification since we find strong evidence of out-of-sample predictability only for those models that allow for level risk to be priced. Most importantly, such statistical evidence is turned into economically significant utility gains, across prediction horizons. The sequential version of the stochastic search variable selection (SSVS) scheme developed offers an important diagnostic as it monitors potential changes in the importance of different risk prices over time and provides further improvement during periods of macroeconomic uncertainty, where results are more pronounced. This is joint work with Tomasz Dubiel-Teleszynski

CTS7: Room: G16 Functional Data – Theory and Methods Chair: Jongmin Lee

Nonparametric Estimation of Covariance and Autocovariance Operators on the Sphere
Speaker: **Alessia Caponera, EPFL**
We propose nonparametric estimators for the second-order central moments of spherical random fields within a functional data context. We consider a measurement framework where each field among an identically distributed collection of spherical random fields is sampled at a few random directions, possibly subject to measurement error. The collection of fields could be i.i.d. or serially dependent. Though similar setups have already been explored for random functions defined on the unit interval, the nonparametric estimators proposed in the literature often rely on local polynomials, which do not readily extend to the (product) spherical setting. We therefore formulate our estimation procedure as a variational problem involving a generalized Tikhonov regularization term. The latter favours smooth covariance/autocovariance functions, where the smoothness is specified by means of suitable Sobolev-like pseudo-differential operators. Using the machinery of reproducing kernel Hilbert spaces, we establish representer theorems that fully characterise the form of our estimators. We determine their uniform rates of convergence as the number of fields diverges, both for the dense (increasing number of spatial samples) and sparse (bounded number of spatial samples) regimes. We moreover validate and demonstrate the practical feasibility of our estimation procedure in a simulation setting, assuming a fixed number of samples per field. Our numerical estimation procedure leverages the sparsity and second-order Kronecker structure of our setup to reduce the computational and memory requirements by approximately three orders of magnitude compared to a naive implementation would require.

Equivariant Estimation of Frechet Means
Speaker: **Andrew McCormack, Duke University**
The Frechet mean generalizes the concept of a mean to a metric space setting. In this work we consider equivariant estimation of Frechet means for parametric models on metric spaces that are Riemannian manifolds. The geometry and symmetry of such a space is encoded by its isometry group. Estimators that are equivariant under the isometry group take into account the symmetry of the metric space. For some models there exists an optimal equivariant estimator, which necessarily will perform as well or better than other common equivariant estimators, such as the maximum likelihood estimator or the sample Frechet mean. We derive the general form of this minimum risk equivariant estimator and in a few cases provide explicit expressions for it. New results for finding the Frechet mean for distributions with radially decreasing densities are presented and used to find expressions for the minimum risk equivariant estimator. Simulation results show that the adaptive equivariant estimator performs favorably relative to alternative estimators.

Learning the Regularity of Curves in Functional Data Analysis and Applications
Speaker: **Valentin Patilea, CREST Ensai**
Combining information both within and across trajectories, we propose simple estimators for the local regularity of the trajectories of a stochastic process. Independent trajectories are measured with errors at randomly sampled time points. Non-asymptotic bounds for the concentration of the estimator are derived. Given the estimate of the local regularity, we build a nearly optimal local polynomial smoother from the curves from a new, possibly very large sample of noisy trajectories. We derive non-asymptotic pointwise risk bounds uniformly over the new set of curves. As another application, we build minimax optimal mean and covariance functions estimators. Our estimators perform well in simulations. Real data sets illustrate the effectiveness of the new approaches.

Robust Spherical Principal Curves
Speaker: **Jongmin Lee, Seoul National University**
Principal curve is a nonlinear generalization of principal components and go through the mean of data lying on Euclidean space In this paper, we propose L1-type and Huber-type principal curves through the median of data to robustify the principal curves for a dataset that may contain outliers. We further investigate the stationarity of the proposed robust principal curves on the 2-sphere. Results from numerical experiments on the 2-sphere and 4-sphrere, including real motion capture data analysis, manifest promising empirical features of the proposed method.

Wednesday 29-06-2022	16:00-17:40	Contributed Talks

CTS8: Room: G03	**Risk, Insurance and Finance**	**Chair: Alejandra Quintos**

Aggregated Markov Chain Models in Life Insurance: Properties and Valuation
Speaker: **Jamaal Ahmad, University of Copenhagen**
Abstract In this talk, we introduce an aggregated Markov chain model in life insurance, where we assign a number of micro states to each biometric macro state, leading to sojourn times in macro states that are dependent and inhomogeneous phase-type distributed. By deriving distributional properties of the associated marked point process and multivariate counting process, we show how this model in general leads to a path-dependency in the states of the insured. While this may complicate valuation of life insurance payments in general, we show that the properties of the inhomogeneous phase-type distribution enables us to compute reserves and premiums explicitly in terms of ordinary differential equations. The talk is ended with a numerical example to illustrate applicability of the results in examples known from actuarial literature.

Multi-asset Optimal Execution and Statistical Arbitrage Strategies under Ornstein–Uhlenbeck Dynamics
Speaker: **Fayçal Drissi, Université Paris 1**
Optimal execution of large orders has mainly been addressed in the case of single-asset portfolios. In practice, optimal execution problems often involve large portfolios comprising numerous assets, and models should consequently account for risks at the portfolio level. In this work, we address multi-asset optimal execution in a framework where prices have multivariate Ornstein-Uhlenbeck dynamics. This model captures the complex multivariate dynamics of portfolios and is well suited when there exist one or several linear combinations of asset prices that are stationary. We use the tools of stochastic optimal control and simplify the initial multi-dimensional Hamilton-Jacobi-Bellman equation into a system of ordinary differential equations (ODEs) involving a Matrix Riccati ODE for which classical existence theorems do not apply. By using a priori estimates obtained thanks to optimal control tools, we nevertheless prove an existence and uniqueness result for the latter ODE, providing a rigorous solution to the problem. Using examples based on data from the foreign exchange and stock markets, we eventually illustrate our results and discuss their implications for both optimal execution and statistical arbitrage. Finally, we briefly show how to use the new theoretical result on matrix Riccati ODEs to address the problem of uncertainty faced by an agent wishing to execute large orders on multiple assets. We propose a model coupling Bayesian learning and dynamic programming techniques where the agent has only à priori knowledge about the distribution of the future drift of the assets composing her portfolio.

Optimal Make-take Fees in a Shared Order Book
Speaker: **Philippe Bergault, Ecole Polytechnique**
In order to attract liquidity and improve trading quality on its platform, an exchange may set a suitable contract with market makers. In this paper, we consider the case of two exchanges sharing a single limit order book. This situation is present in particular on the electricity market in Europe, through the Single Intraday Coupling (SIDC). Using a principal-agent approach, we provide the optimal contract that each exchange should offer to its market maker, as well as the optimal quotes the market makers should display. In particular, we show that there exists a "spillover" effect: when one of the exchanges proposes a suitable make-take fees policy to its market maker, the other exchange also benefits from it.

Dependent Stopping Times and an Application to Credit Risk Theory
Speaker: **Alejandra Quintos, Columbia University**
Stopping times are used in applications to model random arrivals. A standard assumption in many models is that the stopping times are conditionally independent, given an underlying filtration. This is a widely useful assumption, but there are circumstances where it seems to be unnecessarily strong. In the first part of the talk, we use a modified Cox construction, along with the bivariate exponential introduced by Marshall & Olkin (1967), to create a family of stopping times, which are not necessarily conditionally independent, allowing for a positive probability for them to be equal. We also present a series of results exploring the special properties of this construction. In the second part of the talk, we present an application of our model to Credit Risk. We characterize the probability of a market failure which is defined as the default of two or more globally systemically important banks (G-SIBs) in a small interval of time. The default probabilities of the G-SIBs are correlated through the possible existence of a market-wide stress event. We derive various theorems related to market failure probabilities, such as the impact of increasing the number of G-SIBs in an economy and the impact of changing the initial conditions of the economy's state variables. We also show that if there are too many G-SIBs, a market failure is inevitable, i.e., the probability of a market failure tends to one as the number of G-SIBs tends to infinity.

CTS9: Room: G04	**Stochastic Analysis**	**Chair: Andreas Sojmark**

Asymptotic Properties of Measure-valued Pólya Urn Processes
Speaker: **Hristo Sariev, Bulgarian Academy of Sciences**
Measure-valued Pólya urn processes (MVPP) form a general class of random processes with reinforcement that acts as an extension of the classical k-color Pólya urn towards a continuum of possible colors. Formally, an MVPP is a Markov chain on the space of finite measures with an additive structure, representing the evolving composition of an urn with infinitely many colors and random reinforcement. The analysis, which focuses on the asymptotic properties of the normalized urn composition, can be differentiated

based on the nature of its limit. In this talk, we discuss the different reinforcement conditions that lead to random or deterministic and discrete or continuous limit composition, and draw comparisons with k-color urn schemes. We present in detail MVPPs that reinforce only the color of the observed ball, in which case the process is best understood as a randomly reinforced version of Blackwell and MacQueen's Pólya sequence. We discuss the problem of dominance between colors with reference to applications in randomized clinical trials, species sampling and Bayesian inference.

Robuts Super-replication with Transaction Costs for Continuous Processes
Speaker: **Huy Chau, University of Manchester**
We formulate a superhedging theorem in the presence of transaction costs and model uncertainty. Asset prices are assumed continuous and uncertainty is modelled in a parametric setting. Our proof relies on a new topological framework in which no Krein–Smulian type theorem is available.

Covering Systems of Congruences
Speaker: **Robert Hough, Stony Brook University**
I will discuss my solution of the minimum modulus problem for covering systems, a \$ 1000 problem of Paul Erdös, solved using the probabilistic method. Further developments due to myself, Nielsen and Balister, Bollobás, Morris, Sahasrabudhe and Tiba make progress on Erdös' odd problem for coverings.

Continuous Time Random Walks and Convergence of their Stochastic Integrals
Speaker: **Andreas Sojmark, London School of Economics and Political Science**
Deviating from the classical Black–Scholes paradigm of mathematical finance, there is a growing literature thinking instead of asset prices as driven by so-called Continuous Time Random Walks (CTRW). Since the value of investments will then take the form of stochastic integrals involving CTRWs, and since there is a particular interest in utilising certain scaling limits, one is led naturally to important questions about functional convergence of stochastic integrals for CTRWs. In this talk, we will start by discussing some of the main intricacies arising when seeking to answer these questions, and we will then proceed to present both positive and negative results on the functional convergence. While we are particularly motivated by the financial interpretations, we will use these mainly as examples to illustrate a more general framework, which is both easy to apply and encompasses several other interesting applications coming e.g. from physics and time series analysis.

CTS10: Room: G05	Time Series	Chair: Simone Giannerini

Time Series Analysis for Interval-Valued Data
Speaker: **Lynne Billard, University of Georgia**
Many time series data record individual observations as intervals, such as stock market values with daily high-low values, or minimum and maximum monthly temperatures, recorded over time. Maximum likelihood estimators, and their asymptotic properties, of the autocovariance/autocorrelation functions are derived by exploiting the ideas of composite likelihood and pairwise likelihood of Lindsay (1988) and Davis and Yau (2011). A simulation study shows that the new estimators perform considerably better than those obtained previously.

Modelling and Inference for Discrete Time-series using Bayesian Context Trees
Speaker: **Ioannis Papageorgiou, University of Cambridge**
general Bayesian modelling and inference framework is introduced for discrete time-series, based on variable-memory Markov chains. This is a rich and flexible class of higher-order Markov chains, that admit parsimonious representations by allowing the memory length of the process to depend on the values of the most recent observed symbols. Along with the modelling framework, a collection of methodological and algorithmic tools are developed, that allow for efficient Bayesian inference with empirical time-series data. A general prior structure is introduced on the relevant models and parameters, and it is shown that it facilitates exact Bayesian inference. In particular, it is shown that the prior predictive likelihood (averaged over all models and parameters) can be computed exactly by using a version of the context-tree weighting algorithm. Two other algorithms are also provided, which are able to identify the a posteriori most likely models, along with their exact posterior probabilities. Importantly, all three algorithms have only linear complexity in the length of the data. Moreover, a family of variable-dimensional MCMC samplers as well as exact (i.i.d.) samplers are developed for sampling from the joint posterior on models and parameters. The proposed methodology opens the door to a wide range of applications and is found to be successful in numerous important tasks, including model selection, estimation, prediction, segmentation and entropy estimation. The performance of the resulting methods is illustrated on simulated and real-world data from applications in genetics, neuroscience, finance, and animal communication, where they are found to perform at least as well as state-of-the-art approaches. This is joint work with V. Lungu and I. Kontoyiannis.

Semi-supervised Clustering of Time-dependent Categorical Sequences under Positive Constraints
Speaker: **Igor Melnykov, University of Minnesota Duluth**
We consider sequences of categorical values observed over time and assume that the transitions in these sequences occur according to a first-order Markov model. Adopting a finite mixture model and describing the initial state and transition probabilities as functions of time, we estimate all the parameters involved in the model using the EM algorithm under hard positive constraints. In the presence

of such constraints, certain observations are required to be placed in the same class in the solution. Using the proposed methodology, we investigate the association between the education background and major life events of the British Household Panel Survey participants. The analysis reveals which groups of individuals in the study show similar life patterns and points to the relationships that exist between these life outcomes and level of education.

Factor-augmented Tree Ensembles
Speaker: **Filippo Pellegrino, London School of Economics and Political Science**
This manuscript proposes to extend the information set of time-series regression trees with latent stationary factors extracted via state-space methods. In doing so, this approach generalises time-series regression trees on two dimensions. First, it allows to handle predictors that exhibit measurement error, non-stationary trends, seasonality and/or irregularities such as missing observations. Second, it gives a transparent way for using domain-specific theory to inform time-series regression trees. As a byproduct, this technique sets the foundations for structuring powerful ensembles. Their real-world applicability is studied under the lenses of empirical macro-finance.

CTS11: Room: G07	**State Space Models**	**Chair: Thomas Hotz**

Robust Estimation in Finite State Space Hidden Markov Models
Speaker: **Alexandre Lecestre, University of Luxembourg**
We consider stationary hidden Markov models with finite state space and estimate the stationary law of consecutive observations with a focus on robustness properties of our estimator. This means we do not assume the process to be exactly a hidden Markov chain nor to be exactly stationary and we consider the possible presence of outliers. We proved a non-asymptotic sample bound on the resulting squared Hellinger error when the emission densities of our model belong to exponential families. We can derive convergence rates assuming a well specified setting a posteriori, those rates are optimal up to logarithmic factors. It is possible to derive a bound for the estimation of parameters with additional assumptions.

On the Observability of State Space Models with Gaussian Errors and unknown Variance
Speaker: **Ariane Hanebeck, Technische Universität München**
State space models are an important tool for analyzing time series. They assume that observed observations depend on unobserved states over a so-called observation equation. The temporal behavior of the states is explained by the state equation. An important property of state space models is the observability: Given observations, we want to know if it is possible to infer the underlying states and parameters. One way to do this for certain state space models is to use Kalman's observability matrix. However, for the models that are considered in this paper, the variance of the errors is unknown which rules out using this method. We introduce a novel method that investigates the observability of state space models with Gaussian errors. For this, we use discrete approximations (aka deterministic samples) of the Gaussian distribution to get the behavior of the state space models when optimal but noisy data is given. With this technique it is not only possible to answer the question of observability but also to provide a measure of observability.

A Lagged Particle Filter for Stable Filtering of Certain High-dimensional State-space Models
Speaker: **Hamza Ruzayqat, King Abdullah University of Science and Technology**
In this talk I will consider the problem of high-dimensional filtering of state-space models (SSMs) at discrete times. This problem is particularly challenging as analytical solutions are typically not available and many numerical approximation methods can have a cost that scales exponentially with the dimension of the hidden state. Inspired by lag-approximation methods for the smoothing problem, we introduce a lagged approximation of the smoothing distribution that is necessarily biased. For certain classes of SSMs, particularly those that forget the initial condition exponentially fast in time, the bias of our approximation is shown to be uniformly controlled in the dimension and exponentially small in time. We develop a sequential Monte Carlo (SMC) method to recursively estimate expectations with respect to our biased filtering distributions. Moreover, we prove for a class of SSMs that can contain dependencies amongst coordinates that as the dimension d goes to infinity the cost to achieve a stable mean square error in estimation, for classes of expectations, is of $O(Nd^2)$ per-unit time, where N is the number of simulated samples in the SMC algorithm. Our methodology is implemented on several challenging high-dimensional examples including the conservative shallow-water model.

State Space Models as a Flexible Framework for Monitoring Epidemics
Speaker: **Thomas Hotz, TU Ilmenau**
To monitor epidemics, different indicators such as incidences, hospitalisation, and deaths play an important role. The corresponding data form a multivariate time series whose development over time is governed by several effects: the non-linear reproduction equation describing the spread of the disease, delays due to reporting and the patient-dependent progression of the disease, as well as under-reporting. We show that state space models offer a flexible framework to incorporate these effects. They can easily be fitted even to incomplete data either using an extended Kalman filter or following the likelihood-based approach proposed by Durbing & Koopman (2001). Using publicly available data on COVID-19, we demonstrate how the predicted evolution of states can be used both to monitor the epidemic as well as to generate short-term forecasts.

CTS12: Room: G11 Random Graph and Networks Chair: Giorgos Minas

Fluctuations of Subgraph Counts in Graphon-based Random Graphs

Speaker: **Anirban Chatterjee, University of Pennsylvania**

Given a graphon W and a finite simple graph H, with vertex set $V(H)$, denote by $X_n(H,W)$ the number of copies of H in a W-random graph on n vertices. The asymptotic distribution of $X_n(H,W)$ was recently obtained by Hladký, Pelekis, and Šileikis in the case where H is a clique. In this paper, we extend this result to any fixed graph H. Towards this, we introduce a notion of H-regularity of graphons and show that if the graphon W is not H-regular, then $X_n(H,W)$ has Gaussian fluctuations with scaling $n^{|V(H)|-\frac{1}{2}}$. On the other hand, if W is H-regular, then the fluctuations are of order $n^{|V(H)|-1}$ and the limiting distribution of $X_n(H,W)$ can have both Gaussian and non-Gaussian components, where the non-Gaussian component is a (possibly) infinite weighted sum of centered chi-squared random variables with the weights determined by the spectral properties of a graphon derived from W. Our proofs use the asymptotic theory of generalized U-statistics developed by Janson and Nowicki. We also investigate the structure of H-regular graphons for which either the Gaussian or the non-Gaussian component of the limiting distribution (but not both) is degenerate. Interestingly, there are also H-regular graphons W for which both the Gaussian or the non-Gaussian components are degenerate, that is, $X_n(H,W)$ has a degenerate limit even under the scaling $n^{|V(H)|-1}$. We give an example of this degeneracy with $H = K_{1,3}$ (the 3-star) and also establish non-degeneracy in a few examples. This naturally leads to interesting open questions on higher-order degeneracies.

Community Detection with Contextual Multilayer Networks

Speaker: **Sagnik Nandy, University of Pennsylvania**

Let us consider the community detection problem when we observe m networks and a high dimensional covariate matrix, all encoding the same community structure among n subjects. In the asymptotic regime where the number of features p and the number of subjects n grow proportionally, we derive an exact formula of asymptotic minimum mean square error (MMSE) for estimating the common community structure in the balanced two-block case. The formula implies the necessity of integrating information from multiple data sources. Consequently, it induces a sharp threshold of phase transition between the regime where detection (i.e., weak recovery) is possible and the regime where no procedure performs better than a random guess. The asymptotic MMSE depends on the covariate signal-to-noise ratio in a more subtle way than the phase transition threshold does. In the special case of $m = 1$, our asymptotic MMSE formula complements the pioneering work of Deshpande et. al. (2018) which found the sharp threshold when $m = 1$. We map the problem to an appropriate phase synchronization problem with Gaussian matrices and use its properties to derive the asymptotic MMSE. The main technical contribution is the development of a novel AMP algorithm that integrates information from multiple data sources.

Subsampling Based Community Detection for Large Networks

Speaker: **Sayan Chakrabarty, University of Illinois at Urbana-Champaign**

Large networks are increasingly prevalent in many scientific applications. Statistical analysis of such large networks become prohibitive due to exorbitant computation cost and high memory requirements. In this article, we develop a subsampling based divide-and-conquer algorithm, SONNET, for community detection in large networks. The algorithm splits the original network into multiple subnetworks with a common overlap, and applies a suitable community detection algorithm on each subnetwork. The results from individual subnetworks are aggregated using a label matching method to get the final community labels. This method saves both memory and computation costs significantly as one needs to store and process only the smaller subnetworks. This method is also parallelizable which makes it even faster.

Population-level Balance in Signed Networks

Speaker: **Weijing Tang, University of Michigan**

Statistical network models are useful for understanding the underlying formation mechanism and characteristics of complex networks. However, statistical models for signed networks have been largely unexplored. In signed networks, there exist both positive (e.g., like, trust) and negative (e.g., dislike, distrust) edges, which are commonly seen in real-world scenarios. The positive and negative edges in signed networks lead to unique structural patterns, which pose challenges for statistical modeling. In this paper, we introduce a statistically principled latent space approach for modeling signed networks and accommodating the well-known balance theory, i.e., “the enemy of my enemy is my friend” and “the friend of my friend is my friend”. The proposed approach treats both edges and their signs as random variables, and characterizes the balance theory with a novel and natural notion of population-level balance. This approach guides us towards building a class of balanced inner-product models, and towards developing scalable algorithms via projected gradient descent to estimate the latent variables. We also establish non-asymptotic error rates for the estimates, which are further verified through simulation studies. In addition, we apply the proposed approach to an international relation network, which provides an informative and interpretable model-based visualization of countries during World War II.

Stochastic Simulation, Analysis and Inference for Reaction Networks

Speaker: **Giorgos Minas, University of St Andrews**

The continuous biotechnological advances, particularly over the last two decades, continue to provide larger and more informative datasets. They promise a more insightful understanding of biological processes and biomedical advances. Analogous mathematical and statistical advances are required to support these technological advances. A critical challenge to overcome is that biological data are often highly variable due to multiple sources of non-trivial variability affecting them. Stochastic models for reaction networks can describe biological processes incorporating their stochasticity. They can also describe epidemiological, ecological, and sociological processes. In this talk, I will introduce the main approaches for developing stochastic models of reaction networks. I will then describe a new approach for stochastic modelling of oscillators that achieves a suitable balance between model accuracy, computational speed, and scalability to large systems. We will discuss methods for long-time stochastic simulation, analysis of parameter sensitivities using the Fisher Information matrix, and Bayesian statistical inference using MCMC methods based on time-series data.

CTS13: Room: G16	Causal Inference	Chair: Nils Sturma

Causal Survival Analysis from Theory to Practice
Speaker: **Imke Mayer, Charité – Universitätsmedizin Berlin**
Causal survival analysis consists in estimating the effect of a treatment on time-to-event outcome(s). We focus on estimating the restricted mean survival time (RMST), the average survival time from baseline to a pre-specified time, between treated and control groups on right-censored data from an observational study. After stating the identifiability conditions, we review and compare different causal estimation methods, both parametric and non-parametric, which require modeling of the propensity score, the survival outcome and the censoring. We illustrate these methods on observational clinical data to answer a medical question about the effect of transfusion on one-year mortality for patients in intensive care. We discuss the interpretability of the findings of this study from a methodological point of view and explain the methodological challenges for causal survival analysis raised by this study (missing values, selection of the study population, selection of the adjustment variables) in the perspective of guiding future practitioners.

A Semiparametric Method for Evaluating Causal Effects in the Presence of Error-Prone Covariates
Speaker: **Jianxuan Liu, Syracuse University**
The goal of most empirical studies in social sciences and medical research is to determine whether an alteration in an intervention or a treatment will cause a change in the desired outcome response. Unlike randomized designs, establishing the causal relationship based on observational studies is a challenging problem because the ceteris paribus condition is violated. When the covariates of interest are measured with errors, evaluating the causal effects becomes a thorny issue. We propose a semiparametric method to establish the causal relationship, which yields a consistent estimator of the average causal effect. The method we proposed results in locally efficient estimators of the covariate effects. We study their theoretical properties and demonstrate their finite sample performance on simulated data. We further apply the proposed method to the Stroke Recovery in Underserved Populations (SRUP) study by the National Institute on Aging.

Causal Regularization: On the Trade-off between In-sample and Out-of-sample Risk Guarantees
Speaker: **Lucas Kania, Carnegie Mellon University**
In recent decades, a number of ways of dealing with causality in practice, such as propensity score matching, the PC algorithm and invariant causal prediction, have been introduced. Besides its interpretational appeal, the causal model provides the best out-of-sample prediction guarantees. We study the identification of causal-like models from in-sample data that provide out-of-sample risk guarantees when predicting a target variable from a set of covariates. Whereas ordinary least squares provides the best in-sample risk with limited out-of-sample guarantees, causal models have the best out-of-sample guarantees but achieve an inferior in-sample risk. By defining a trade-off of these properties, we introduce causal regularization. As the regularization is increased, it provides estimators whose risk is more stable across sub-samples at the cost of increasing their overall in-sample risk. The increased risk stability is shown to lead to out-of-sample risk guarantees. We provide finite sample risk bounds for all models and prove the adequacy of cross-validation for attaining these bounds. Finally, we apply the methodology to econometric and genetic datasets, and discuss a possible extension to generalized linear models.

Half-Trek Criterion for Identifiability of Latent Variable Models
Speaker: **Nils Sturma, Technical University of Munich**
We consider linear structural equation models with latent variables and develop a criterion to certify whether the direct causal effects between the observable variables are identifiable based on the observed covariance matrix. Each model corresponds to a directed graph whose edges represent the direct effects that appear as coefficients in the equation system defining the model. Prior research has developed a variety of methods to decide identifiability of direct effects in a latent projection framework, in which the confounding effects of the latent variables are represented by correlation among noise terms. This approach is effective when the confounding is sparse and effects only small subsets of the observed variables. In contrast, the new latent-factor half-trek criterion (LF-HTC) we develop operates on the original unprojected latent variable model and is able to certify identifiability in settings, where some latent variables may also have dense effects on many or even all of the observables. Our LF-HTC is an effective sufficient criterion and when restricting the search steps to consider subsets of latent variables of bounded size, the criterion can be verified in time that is polynomial in the size of the graph.

CTS14: Room: G22	Advanced Theoretical Statistics	Chair: Patrick Rebeschini

Calibrating the Scan Statistic: Finite Sample Performance vs. Asymptotics
Speaker: **Guenther Walther, Stanford University**

Recent results for the univariate scan problem have shown that using scale-dependent critical values allows to attain asymptotically optimal detection simultaneously for all signal lengths, but this procedure has been criticized for losing too much power for short signals in a finite sample setting. We explain this discrepancy between asymptotics and finite sample results and we propose a new finite sample criterion. We present three calibrations for scan statistics that perform well across a range of relevant signal lengths. Joint work with Andrew Perry.

Minimax Rates for Conditional Density Estimation via Empirical Entropy
Speaker: **Blair Bilodeau, University of Toronto**
We consider the task of estimating a conditional density using i.i.d. samples from a joint distribution, which is a fundamental problem with applications in both classification and uncertainty quantification for regression. For joint density estimation, minimax rates have been characterized for general density classes in terms of uniform (metric) entropy, a well-studied notion of statistical capacity. When applying these results to conditional density estimation, the use of uniform entropy—which is infinite when the covariate space is unbounded and suffers from the curse of dimensionality—can lead to suboptimal rates. Consequently, minimax rates for conditional density estimation cannot be characterized using these classical results. We resolve this problem for well-specified models, obtaining matching (within logarithmic factors) upper and lower bounds on the minimax Kullback–Leibler risk in terms of the empirical Hellinger entropy for the conditional density class. The use of empirical entropy allows us to appeal to concentration arguments based on local Rademacher complexity, which—in contrast to uniform entropy—leads to matching rates for large, potentially nonparametric classes and captures the correct dependence on the complexity of the covariate space. Our results require only that the conditional densities are bounded above, and do not require that they are bounded below or otherwise satisfy any tail conditions.

Finite Mixture Models: a Bridge with Stochastic Geometry and Choquet Theory
Speaker: **Michele Caprio, Duke University**
In the context of a finite mixture model whose components and weights are unknown, if the number of components is a function of the amount of data collected, we give the growth rate of its expected value. If instead the number of identifiable components is known but the components themselves and the weights are not, we show that a Dirichlet process retrieves the Choquet measure supported on the identifiable components. In turn, this gives us the weights. Finally, we propose a novel algorithm that finds the most parsimonious model, in terms of the number of components, that well fits the data.

Robust Estimation under a Shape Constraint
Speaker: **Hélène Halconruy, University of Luxembourg**
The problem of estimating a density under a shape constraint has been widely studied since a seminal paper by Grenander (1956). To estimate a non-increasing density on the half-line, Grenander devised an eponymous estimator that coincides with the maximum likelihood estimator (MLE) on all non-increasing densities on the half-line. The MLE has nice adaptation properties which make it the favourite and almost exclusive estimator to tackle other shapeconstrained estimation problems such as convexity, monotonicity, log-concavity (in higher dimension)... However, it has some drawbacks; in particular, it requires the knowledge of some information about the support of the target density, perhaps unknown in practice. In a joint work with Y. Baraud and G. Maillard, we design, in the one-dimensional case, an estimator that retains the minimax and adaptation properties of the MLE for shape-constrained density estimation and that is robust, i.e., it remains stable with respect to a slight deviation from the ideal situation of truly i.i.d. data whose density satisfies the required shape. Indeed, our estimator still performs well when the equidistribution condition is slightly violated, or there is a small portion of outliers in the sample, or the true density of the data (in the i.i.d. case) does not satisfy the shape constraint but is close enough (for some loss) to a density that does. In this talk, I will present our $\updownarrow$-estimation based procedure (Baraud, 2021), give its hazard bounds for the total variation distance, and illustrate it with the estimation of a piecewise monotone density.

Concentration without Bernstein
Speaker: **Patrick Rebeschini, University of Oxford**
The local Rademacher complexity framework is one of the most successful general-purpose toolboxes for establishing sharp excess risk bounds for statistical estimators based on the framework of empirical risk minimization. Applying this toolbox typically requires using the Bernstein condition, which often restricts applicability to convex and proper settings. Recent years have witnessed several examples of problems where optimal statistical performance is only achievable via non-convex and improper estimators originating from aggregation theory, including the fundamental problem of model selection. These examples are currently outside of the reach of the classical localization theory. In this work, we build upon the recent approach to localization via offset Rademacher complexities, for which a general high-probability theory has yet to be established. Our main result is an exponential-tail excess risk bound expressed in terms of the offset Rademacher complexity that yields results at least as sharp as those obtainable via the classical theory. However, our bound applies under an estimator-dependent geometric condition (the "offset condition") instead of the estimator-independent (but, in general, distribution-dependent) Bernstein condition on which the classical theory relies. Our results apply to improper prediction regimes not directly covered by the classical theory. (Based on joint work with V. Kanade and T. Vaškevičius)

CTS15: Room: G26 Genetics Chair: William Rosenberger

A New Phylogenetic Association Test based on a Chinese Restaurant Process-esque Model
Speaker: **Julie Zhang, Stanford University**
Present day molecular data, such as DNA, informs about the past evolutionary dynamics of populations. Inference of such dynamics and the sample phylogeny is relevant in many areas of biology, including evolutionary genetics and phylodynamics of infectious diseases. An important question to investigate is whether a certain observable trait is associated with the phylogenetic tree structure or whether the observed patterns of genetic data reflect patterns of population structure. We propose a new model for evolutionary trees motivated by the Chinese Restaurant Process that includes a parameter representing the degree of preferential attachment. This model with no preferential attachment is equivalent to a structured coalescent model with instantaneous migration and coalescence. We then derive a non-parametric test for phylogenetic trait association of binary traits and demonstrate the power of the test. Finally, we apply our test to a cancer data set and to SARS-CoV-2 evolutionary trees.

Modelling Biomarker Variability in Joint Analysis of Longitudinal and Time-to-event Data
Speaker: **Chunyu Wang, University of Manchester**
Modelling biomarker variability in joint analysis of longitudinal and time-to-event data. Abstract: The role of visit-to-visit variability of a biomarker in predicting related disease has been recognized in medical science. Existing measures of biological variability are criticized for being entangled with random variability resulted from measurement error or being unreliable due to limited measurements per individual. We propose a new measure to quantify the biological variability of a biomarker by evaluating the fluctuation of each individual-specific trajectory behind longitudinal measurements. Given a mixed-effects model for longitudinal data with the mean function over time specified by cubic splines, our proposed variability measure can be mathematically expressed as a quadratic form of random effects. A Cox model is assumed for time-to-event data by incorporating the defined variability as well as the current level of the underlying longitudinal trajectory as covariates, which, together with the longitudinal model, constitutes the joint modelling framework. Asymptotic properties of maximum likelihood estimator are established for the present joint model. Estimation is implemented via EM algorithm with fully exponential Laplace approximation used in E-step to reduce the computation burden due to the increase of random effect dimension. Simulation studies are conducted to reveal the advantage of the joint modelling method over the two-stage method. Finally, we apply our model to investigate the effect of systolic blood pressure variability on cardiovascular events in the Medical Research Council elderly trial.

Data-driven Design of Targeted Gene Panels for Estimating Immunotherapy Biomarkers
Speaker: **Jacob Bradley, University of Edinburgh**
Tumour mutation burden and other exome-wide biomarkers are used to determine which patients will benefit from immunotherapy. However, the cost of whole exome sequencing limits the widespread use of such biomarkers. Here, we introduce a data-driven framework for the design of targeted gene panels for estimating a broad class of biomarkers including tumour mutation burden and tumour indel burden. Our approach allows the practitioner to select a targeted gene panel of prespecified size and construct an estimator that only depends on the selected genes. We demonstrate the performance of our proposal using data from various cancer types.

Two-Stage Enrichment Designs With a Continuous Biomarker
Speaker: **William Rosenberger, George Mason University**
Two-stage enrichment designs can be used to target the benefiting population in clinical trials based on patients' biomarkers. In the case of continuous biomarkers, we show that using a bivariate model that treats biomarkers as random variables more accurately identifies a treatment-benefiting enriched population than assuming biomarkers are fixed. Additionally, we show that under the bivariate model, the maximum likelihood estimators (MLEs) follow a randomly scaled mixture of normal distributions. Using random normings, we obtain asymptotically standard normal MLEs and construct hypothesis tests. Finally, in a simulation study, we demonstrate that our proposed design is more powerful than a single stage design when outcomes and biomarkers are correlated; the model-based estimators have smaller bias and mean square error (MSE) than weighted average estimators. This work recently appeared in 2021 *Annals of Statistics.*

Thursday 30-06-2022	09:00-10:00	Contributed Talks

CTS16: Room: G03 Data Privacy Chair: Daniel Alabi

Edge Differentially Private Estimation in the β-model via Jittering and Method of Moments
Speaker: **Fengting Yi, Southwestern University of Finance and Economics**
A standing challenge in data privacy is the trade-off between the level of privacy and the efficiency of statistical inference. Here we conduct an in-depth study of this trade-off for parameter estimation in the β-model (Chatterjee, Diaconis and Sly, 2011) for edge differentially private network data re- leased via jittering (Karwa, Krivitsky and Slavković, 2017). Unlike most previous approaches based on maximum likelihood estimation for this network model, we proceed via method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels – corresponding to stricter privacy – than has been to date. Over this new range we discover our proposed estimator for the parameters exhibits an interesting phase transition, with both its convergence rate and asymptotic variance following one of three different regimes of behavior depending on the level of privacy. Because identification of the operable regime is difficult to impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases. In fact, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of vertices), which would appear to be the first result of its kind. Numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

Hypothesis Testing for Differentially Private Linear Regression
Speaker: **Daniel Alabi, Harvard University**
In this work, we design differentially private hypothesis tests for the following problems in the general linear model: testing a linear relationship and testing for the presence of mixtures. The majority of our hypothesis tests are based on differentially private versions of the F-statistic for the general linear model framework, which are uniformly most powerful unbiased in the non-private setting. We also present another test for testing mixtures, based on the differentially private nonparametric tests of Couch, Kazan, Shi, Bray, and Groce (CCS 2019), which is especially suited for the small dataset regime. We show that the differentially private F-statistic converges to the asymptotic distribution of its non-private counterpart. Through a suite of Monte Carlo based experiments, we show that our tests achieve desired significance levels and have a high power that approaches the power of the non-private tests as we increase sample sizes or the privacy-loss parameter. This is joint work with Salil Vadhan (Harvard University).

CTS17: Room: G04 Monte Carlo Methods Chair: Guo-Jhen Wu

Analysis of Langevin Monte Carlo from Poincaré to Log-Sobolev
Speaker: **Mufan Li, University of Toronto**
Classically, the continuous-time Langevin diffusion converges exponentially fast to its stationary distribution π under the sole assumption that π satisfies a Poincaré inequality. Using this fact to provide guarantees for the discrete-time Langevin Monte Carlo (LMC) algorithm, however, is considerably more challenging due to the need for working with chi-squared or Rényi divergences, and prior works have largely focused on strongly log-concave targets. In this work, we provide the first convergence guarantees for LMC assuming that π satisfies either a Latała–Oleszkiewicz or modified log-Sobolev inequality, which interpolates between the Poincaré and log-Sobolev settings. Unlike prior works, our results allow for weak smoothness and do not require convexity or dissipativity conditions.

Analysis and Optimization of Certain Parallel Monte Carlo Methods in the Low Temperature Limit
Speaker: **Guo-Jhen Wu, KTH Royal Institute of Technology**
Metastability is a formidable challenge to Markov chain Monte Carlo methods. In this talk we present methods for algorithm design to meet this challenge. The design problem we consider is temperature selection for the infinite swapping scheme, which is the limit of the widely used parallel tempering scheme obtained when the swap rate tends to infinity. We use a recently developed tool for the large deviation properties of the empirical measure of a metastable small noise diffusion to transform the variance reduction problem into an explicit graph optimization problem. The nodes in the graph optimization problem correspond to metastable states of the noiseless dynamics. Our first analysis of the optimization problem is in the setting of a double well model, and it shows that the optimal selection of temperature ratios is a geometric sequence except possibly the highest temperature. In the same setting we identify two different sources of variance reduction, and show how their competition determines the optimal highest temperature. In the general multi-well setting we prove that the same geometric sequence of temperature ratios as in the two-well case is always nearly optimal, with a performance gap that decays geometrically in the number of temperatures.

CTS18: Room: G05 Statistical Tests Chair: Dimitrios Bagkavos

Distance correlation, Hilbert–Schmidt and global tests: A look into independence testing in discrete spaces from three perspectives
Speaker: **Fernando Castro-Prado, University of Santiago de Compostela**
Distance covariance is an association measure that can be defined when the marginal spaces have a semimetric structure. It has been

proved to be equivalent to the Hilbert–Schmidt independence criterion, by mapping strong negative type distances into characteristic kernels. Both concepts, in turn, can be linked to the global tests of Gaussian process regression through feature maps. We present some novel tests for the association of ternary covariates with quantitative responses (motivated by a relevant problem in human genetics), using this unified approach. We show that certain versions of distance covariance correspond to locally most powerful tests for specific statistical models leading to insights regarding which situations these tests perform well, which are of paramount interest in the application fields. Closed form expressions for the distributions of the test statistics and corresponding estimators are obtained from the spectral decomposition of covariance operators. The performance of the approach is briefly illustrated via simulations and real data. Some extensions and future work are also to be discussed. This is joint work with Dominic Edelmann (German Cancer Research Centre) and Jelle J Goeman (Leiden University Medical Center).

Testing the Goodness-of-Fit of Maximum Likelihood Estimates of Normal Mixture Densities via Kernel Smoothing Techniques

Speaker: **Dimitrios Bagkavos, University of Ioannina**

A novel goodness-of-fit test for assessing the validity of normal mixture densities that result by a parametric estimation procedure, such as maximum likelihood, is introduced. The theoretical results established include analytic quantification of the test statistic's size and power functions under (composite) fixed and local alternatives. These are used to derive closed-form expressions for a bandwidth rule which optimizes the test's power while keeping the size constant at a given significance level, and a cut-off point suitable for finite sample implementations of the test. An extensive simulation study compares the performance of the new test to well-established tests in the literature and demonstrates the superiority of the former in all examples considered. Finally, its practical usefulness is demonstrated in the analysis of a real world dataset.

CTS19: Room: G07	**Some Topics in Modern Statistics**	**Chair: Sofia Villar**

Stochastic gradient descent for estimation and inference in spatial autoregressive models

Speaker: **Ji Meng Loh, New Jersey Institute of Technology**

We consider using stochastic gradient descent (SGD) procedure to fit spatial auto-regressive models to lattice data, incorporating a recently developed perturbation method to obtain standard errors in addition to model parameter estimates. We obtain the SGD update equations for fitting an SAR model to data and present results of a simulation study to examine the empirical coverage of confidence intervals constructed using the perturbation procedure. We will also discuss on-going challenges as well as application of the procedure to spatial auto-regressive quantile models.

Nonparametric, Tuning-free Estimation of S-shaped Functions

Speaker: **Yining Chen, London School of Economics and Political Science**

We consider the nonparametric estimation of an S-shaped regression function. The least squares estimator provides a very natural, tuning-free approach, but results in a non-convex optimisation problem, since the inflection point is unknown. We show that the estimator may nevertheless be regarded as a projection onto a finite union of convex cones, which allows us to propose a mixed primal-dual bases algorithm for its efficient, sequential computation. Our main theoretical results provide sharp oracle inequalities that yield worst-case and adaptive risk bounds for the estimation of the regression function, as well as a rate of convergence for the estimation of the inflection point. These results reveal that the estimator achieves the minimax or optimal rate of convergence for both the estimation of the regression function and its inflection point (up to a logarithmic factor in the latter case). Simulations and a real data application also confirm the desirable finite-sample properties of the estimator. This is joint work with Raymond Carroll, Oliver Feng, Qiyang Han and Richard Samworth.

Design-based Inference for Bandit Algorithms

Speaker: **Sofia Villar, University of Cambridge**

The use of bandit algorithms to conduct response-adaptive randomised experiments may improve performance in terms of regret (or rewards) but it poses major problems for traditional statistical inference (e.g., biased estimators, inflated type-I error and reduced power). Recent work to address these problems typically work by imposing restrictions on the "exploitative" nature of the bandit algorithm, thus trading-off regret and inference goals. Additionally, these methods also require large sample sizes to ensure asymptotic guarantees. However, in many contexts bandit algorithms could be best used in exploratory experiments that are also tightly constrained in its size or duration (e.g. pilot studies for clinical trials). Increasing power in such (small) experiments, without having to limit the adaptive nature of the algorithm, could allow for more promising interventions to reach a larger confirmatory experimental phase. In this talk, I will present a novel hypothesis test approach particularly tailored for highly "exploitative" adaptive bandit algorithms. Such design-based tests are centred on the allocation probabilities of the underlying bandit algorithm, and do not require constraining its adaptive nature or a minimum experimental size. I will illustrate the finite-sample performance of the test for two different bandit algorithms, the CARA Forward Looking Gittins Index (introduced by Villar & Rosenberger, 2017) and Thompson Sampling (first introduced by Thompson, 1933), using both extensive simulations and a real-world field experiment.

CTS20: Room: G16 Epidemic Dynamics and Environmental Study Chair: Stefan Heyder

An Endogenous Approach to Explore Epidemic Dynamics on a Network
Speaker: **Abhinandan Dalal, University of Pennsylvania**
In this talk, we build a framework to study the propagation of an epidemic in a population with observed network interactions. Most existing and recent literature has concentrated on exploring epidemic dynamics on the network using ad hoc assumptions on the interactions between the individuals. We build a systematic game theoretic framework to explain an agent's behaviour as an endogenous rational choice. We also wish to study the dynamics of the epidemic and its propagation, along with parameters for initial behavior of the epidemic, like initial reproduction number and initial growth rate etc. Alongside, we also conduct simulations to study the growth of epidemics under various random graph models, in conjunction with the rational behavior dictated by the game theoretic models, and see how the epidemic dynamics differ from one model to the other.

Learning Torus PCA-based Classification for Multiscale RNA Correction with Application to SARS-CoV-2
Speaker: **Stephan Huckeman, University of Gottingen**
Abstract: Reconstructions of structure of biomolecules, for instance via X-ray crystallography or cryo-EM frequently contain clashes of atomic centers which are not chemically permissible. Methods to correct these clashes are usually based on simulations approximating biophysical chemistry, making them computationally expensive and often not correcting all clashes. Using RNA data, we propose fast, data-driven multiscale learning reconstructions from clash free RNA benchmark data. Multiscale here is based on two levels of shape analysis for RNA geometry. On the microscopic (suite) scale we represent dihedral angles on a suitably stratified sphere, allowing, among others, for circular mode hunting post torus PCA. On the mesoscopic (corresponding to the size of a half helix turn) scale we employ constrained stratified size-and-shape spaces. Our reconstruction proposals exploit their interdependence. While this method is general for RNA we illustrate its power by application to the RNA example of SARS-CoV-2. (Joint work with Benjamin Eltzner, Kanti V. Mardia and Henrik Wiechers.)

Regional Estimates of Reproduction Numbers with Application to COVID-19
Speaker: **Stefan Heyder, TU Ilmenau**
A key statistic for monitoring epidemics is the reproduction number which measures the amount of secondary cases produced by a single infectious individual. While estimates of this quantity are readily available on the national level, subnational estimates, e.g. on the county level, pose more difficulties as there are only few incidences. However, as countermeasures to the pandemic are usually enforced on the subnational level, such estimates are of great interest to assess the efficacy of the measures taken, and to guide future policy. We present a novel extension of the well established estimator C.Fraser (2007) of the country level reproduction number to the county level by applying techniques from small-area estimation. This new estimator yields sensible estimates of reproduction numbers both on the country and county level. It can handle low and highly variable case counts on the county level, and may be used to distinguish local outbreaks from more widespread ones. We demonstrate the capabilities of our novel estimator by a simulation study and by applying the estimator to German case data on COVID-19. This is joint work with Jan Pablo Burgard (University of Trier), Thomas Hotz (TU Ilmenau) and Tyll Krüger (Wroclaw University of Science and Technology).

CTS21: Room: G22 Some Topics in Statistical Learning Chair: Wicher Bergsma

Plugin Estimation of Smooth Optimal Transport Maps
Speaker: **Tudor Manole, Carnegie Mellon University**
Motivated by the growing popularity of optimal transport as a methodological tool in statistics and machine learning, we study the question of efficiently estimating the optimal transport map between two unknown multivariate distributions. The minimax rate for this problem was recently derived by Hutter and Rigollet (2021), but was only shown to be achievable by a computationally intractable estimator. We prove that various plugin estimators—which are defined as the optimal transport map between nonparametric estimators of the two distributions—are also minimax optimal. Such estimators are simple to compute using standard numerical solvers. Our proofs rely on new stability arguments for the quadratic optimal transport problem. As a byproduct of these results, we also derive central limit theorems and convergence rates for plugin estimators of the quadratic Wasserstein distance, which are of independent interest. This is based on joint work with Sivaraman Balakrishnan, Jonathan Niles-Weed, and Larry Wasserman.

On Quantiles, Continuity, and Robustness
Speaker: **Riccardo Passeggeri, Imperial College London**
In this talk I will present novel properties of the quantiles and provide a robustification procedure based on them. I will consider both the geometric and the component-wise quantiles in both finite and infinite dimensions. I will show their existence, uniqueness, and continuity. Building on these results, I will introduce a robustification procedure for almost all estimators, which extends the celebrated median-of-means (see Lugosi and Mendelson J. Eur. Math. Soc. 22 (3), 925-965, 2020). The robustification procedure applies even to settings where the contaminated data are of any nature and their number is of order n^c, where c in $(0,1)$.

Estimation and Selection of Additive Interaction Models using I-priors
Speaker: **Wicher Bergsma, London School of Economics and Political Science**
Additive models with interactions have been considered extensively in the literature and have also been referred to as interaction

spline models. Two problems have hampered their application. Firstly, estimating the models can be difficult due to potentially many tuning (or smoothing) parameters. Secondly, model selection may be difficult due to potentially multiple models needing to be compared. In this talk we introduce a novel approach to estimating and comparing additive models with and without interaction effects. Firstly, we extend the I-prior methodology (Bergsma, 2020) to multiple covariates, each of which may be multidimensional. The I-prior is an objective, proper prior for a regression function based on its Fisher information, and the regression function is then estimated by its posterior mode. In general, the I-prior methodology has some theoretical advantages over competing methods such as Gaussian process regression and Tikhonov regularization. In the present case, it has a particular computational advantage, in that it permits an EM algorithm with simple E and M steps to find the maximum marginal likelihood estimators of the scale parameters (also known as tuning parameters), making their estimation easier than for competing methods. Secondly, we propose a parsimonious specification of models with interactions. An empirical Bayes model selection procedure can then be used, selecting the model with the highest marginal likelihood. Simulations show competitive performance with the lasso and other methods, while our approach is much more generally applicable to arbitrary additive interaction models. This is joint work with Haziq Jamil (Universiti Brunei Darussalam).

CTS22: Room: G26	**Complex Data Analysis**	**Chair: Aritra Halder**

Wasserstein Distributionally Robust Estimation with Wasserstein Barycenters
Speaker: **Tim Tsz-Kit Lau, Northwestern University**
In many applications in statistics and machine learning, the availability of data samples from multiple possibly heterogeneous sources has become increasingly prevalent. On the other hand, in distributionally robust optimization, we seek data-driven decisions which perform well under the most adverse distribution from a nominal distribution constructed from data samples within a certain discrepancy of probability distributions. However, it remains unclear how to achieve such distributional robustness in model learning and estimation when data samples from multiple sources are available. In this work, we propose constructing the nominal distribution in optimal transport-based distributionally robust optimization problems through the notion of Wasserstein barycenter as an aggregation of data samples from multiple sources. Under specific choices of the loss function, the proposed formulation admits a tractable reformulation as a finite convex program, with powerful finite-sample and asymptotic guarantees. As an illustrative example, we demonstrate with the problem of distributionally robust sparse inverse covariance matrix estimation for zero-mean Gaussian random vectors that our proposed scheme outperforms other widely used estimators in both the low- and high-dimensional regimes.

Curvature Processes: Directional Concavity in Gaussian Random Fields
Speaker: **Aritra Halder, University of Virginia**
Spatial process models are widely used for modeling point-referenced variables arising from diverse scientific domains. Analyzing the resulting random surface provides deeper insights into the nature of latent dependence within the studied response. We develop Bayesian modeling and inference for rapid changes on the response surface to assess directional curvature along a given trajectory. Such trajectories or curves of rapid change, often referred to as *wombling* boundaries, occur in geographic space in the form of rivers in a flood plain, roads, mountains or plateaus or other topographic features lead to high gradients on the response surface. We demonstrate fully model based Bayesian inference on directional curvature processes to analyze differential behavior in responses along wombling boundaries. We illustrate our methodology with a number of simulated experiments followed by multiple applications featuring the Boston Housing data; Meuse river data; and temperature data from the Northeastern United States.

CTS24: Room: G21A	**Spatial Data and Outliers**	**Chair: Gabriel Wallin**

Additive Gaussian Process Models for Spatial and Spatio-temporal Analysis
Speaker: **Sahoko Ishida, London School of Economics and Political Science**
Regression with Gaussian Process (GP) prior is a powerful statistical tool for modelling a wide variety of data with both Gaussian and non-Gaussian likelihood. In the spatial statistics community, GP regression, also known as Kriging, has a long-standing history. It has been proven useful since its introduction, due to its capability of modelling autocorrelation of spatial and spatio-temporal data. Other than space and time, real-life applications often contain additional information with different characteristics. In applied research, interests often lie in exploring whether there exists a space-time interaction or investigating relationships with covariates and the outcome while controlling for space and time effect. Additive GP regression allows to model such flexible relationships by exploiting the structure of the GP covariance function (kernel) by adding and multiplying different kernels for different types of covariates. This has been only partially adapted in spatial and spatio-temporal analysis. In this study, we use ANOVA decomposition of kernels and introduce a unified approach to model spatio-temporal data, using the full flexibility of additive GP models. Not only does this permit modelling of main effects and interactions of space and time, but furthermore to include covariates, and let the effects of the covariates vary with time and space. We consider various types of outcomes including, continuous, categorical and counts. By exploiting kernels for graphs and networks, we show that areal data can be modelled in the same manner as the data that are geo-coded using coordinates. For model estimation, we have implemented both MCMC algorithm and analytical approximations including Laplace approximation and variational inference. In this presentation, we demonstrate the proposed methods using empirical data.

Blind Source Separation over Space

Speaker: **Bo Zhang, University of Science and Technology of China**

We propose a new estimation method for the blind source separation model of Bachoc et al. (2020). The new estimation is based on an eigenanalysis of a positive definite matrix defined in terms of multiple spatial local covariance matrices, and, therefore, can handle moderately high-dimensional random fields. The consistency of the estimated mixing matrix is established with explicit error rates even when the eigen-gap decays to 0 slowly. The proposed method is illustrated via both simulation and a real data example.

Two-way Outlier Detection for Item Response Data

Speaker: **Gabriel Wallin, London School of Economics and Political Science**

Outliers in multivariate data are commonly encountered. An example of such is cheating in standardised educational tests, where cheaters and leaked test items are regarded as outliers that need to be detected. In contrast to regression analysis, outliers for multivariate data are not as well-defined and harder to detect. Detection of cheaters and compromised items has received much attention in recent years, however, most of the existing methods target either respondent or item outlier detection, but not both. We therefore propose a flexible latent variable model based on substantial knowledge and historical data as a baseline model and define outliers as observations that deviate from the baseline model. Our model simultaneously classifies both respondent and item outliers and can incorporate, but doesn't require any, prior information. The two outlier terms are assumed to interact with each other and to have a sparse structure. The simultaneous classification is conducted in two steps. First, an initial estimate of the model parameters based on joint maximum likelihood is retrieved. In a second, post-estimation learning step, we seek a sparse representation of the outlier component of the model by rotating the parameter solution using a sparse-pursuit transformation and hard-thresholding. An algorithm based on iterative reweighted least squares is developed to find the rotation matrix which yields a sparse solution. Our simulation results are much promising, showing a very high rate of true positive and true negative classifications for both respondent and item outliers. Our method will furthermore be illustrated using empirical data and theoretically justified under a double asymptotics regime to give a classification consistency result.

Thursday 30-06-2022	10:30-12:30	Contributed Talks

CTS1: Room: G21A High-Dimensional Inference Chair: Jie Li

High Dimensional PCA: A New Model Selection Criterion
Speaker: **Abhinav Chakraborty, University of Pennsylvania**
Estimating the number of dominant/large eigenvalues of the population covariance matrix based on the sample information is an important question arising in Statistics with wide applications in many areas. In the context of Principal Components Analysis (PCA), the linear combinations of the original variables having the largest amounts of variation are determined by this number. In this talk, we investigate the high dimensional asymptotic setting where the number of variables grows at the same rate as the number of observations. We work in the framework where the population covariance matrix is assumed to have the spiked structure proposed in Johnstone (2001) . In this setup, the problem of interest becomes essentially one of model selection and has attracted a lot of interest from researchers. Our focus is on the Akaike Information Criterion (AIC) which is known to be strongly consistent from the work of Bai (2018). This requires a certain "gap condition" ensuring that the dominant eigenvalues of the covariance matrix are all above a level which is strictly larger than a threshold discovered by Baik, Ben Arous and Peche (called the BBP threshold), both quantities depending on the limiting ratio of the number of variables and observations. It is well-known in the literature that, below this threshold, a spiked covariance structure becomes indistinguishable from one with no spikes. Thus the strong consistency of AIC requires in a sense some extra "signal strength" than what the BBP threshold corresponds to. In this talk, our aim is to investigate whether consistency continues to hold even if the "gap" is made smaller. In this regard, we make two novel theoretical contributions. Firstly, we show that strong consistency under arbitrarily small gap is achievable if we alter the penalty term of AIC suitably depending on the target gap. Inspired by this result, we are able to show that a further intuitive alteration of the penalty can indeed make the gap exactly zero, although we can only achieve weak consistency in this case. We compare the two newly-proposed estimators with other existing estimators in the literature via extensive simulation studies, and show, by suitably calibrating our proposals, that a significant improvement in terms of mean-squared error is achievable.

Inference in High-dimensional Logistic Regression Models with Separated Data
Speaker: **Rebecca Lewis, Imperial College London**
Existence of the maximum likelihood estimate of logistic regression coefficients requires that the observed sequence of covariate and response values are not linearly separable. Even when the maximum likelihood estimator exists, it can suffer from considerable bias when the number of independent observations is not large relative to the dimension. We propose an alternative approach to inference on the logistic regression coefficients based on a corrected ordinary least squares estimator. Consistency and asymptotic normality of this estimator is established under a high-dimensional regime in which the number p of covariates and the sample size n both tend to infinity with p < n under weak conditions on the design matrix. Validity of Wald-based inference through this route is thereby established, even when maximum likelihood is infeasible.

A Finite Sample Analysis of the Benign Overfitting Phenomenon for Ridge Function Estimation
Speaker: **Stephane Chretien, University of Lyon**
Recent extensive numerical experiments in high scale machine learning have allowed to uncover a quite counter-intuitive phase transition, as a function of the ratio between the sample size n and the number of parameters p in the model. As the number of parameters approaches the sample size, the generalisation error (a.k.a. testing error) increases, but in many cases, it starts decreasing again past the threshold $p = n$. This surprising phenomenon, brought to the theoretical community attention in Belkin et al.(2019), has been thoroughly investigated lately, more specifically for simpler models than deep neural networks, such as the linear model when the parameter is taken to be the minimum norm solution to the least-square problem, mostly in the asymptotic regime when p and n tend to infinity; see e.g. Hastie et al. (2019). In the present paper, we propose a novel finite sample analysis of non-linear models of ridge type, where we investigate the over-parameterised regime of the double descent phenomenon for both the estimation problem and the prediction problem. Our results provide a precise analysis of the distance of the best estimator from the true parameter as well as a generalisation bound which complements recent works of Bartlett et al. (2020) and Chinot and Lerasle (2020). Our analysis is based on efficient but elementary tools closely related to the continuous Newton method as advocated by Neuberger (2007). This is joint work with Emmanuel Caron.

Robust Mean and Eigenvalues Regularised Covariance Matrix Estimation
Speaker: **Wenyu Cheng, London School of Economics and Political Science**
Covariance matrix is a common tool for summarising linear relationships between variables. The topic is particularly of interest in the high-dimensional setting ($p/n \to c > 0$), where the sample covariance estimator is no longer consistent. Non-parametric eigenvalue shrinkage covariance estimator (NERCOME) offers a solution by shrinking eigenvalues non-linearly, without structural assumptions on the covariance matrix. However, it requires the 12th moment to exist, a condition hard to verify and often unsatisfied in the real world. We draw on recent developments from the robust statistics literature to robustify NERCOME against heavy-tailed densities. Specifically, we propose a set of cutomisable influence functions that generalise the famous Huber loss and Catoni loss functions. The result allows us to estimate covariances consistently as long as barely over 2 moments exist and to produce narrower bounds. A tuning-free scheme is provided to deal with practical situations where such moment exist but unknown. Crucially, robust

NERCOME keeps the overall efficiency and other properties of NERCOME, while curbing the influences of extreme observations. We challenge the robust NERCOME with other state-of-art covariance estimators through extensive simulation studies. Application in Parkinson's disease classification is provided in the end.

Laplace and Saddlepoint Approximations in High Dimensions
Speaker: **Yanbo Tang, University of Toronto**
We examine the behaviour of the Laplace and saddlepoint approximations in the high-dimensional setting, where the dimension of the model is allowed to increase with the number of observations. Approximations to the joint density, the marginal posterior density and the conditional density are considered. Our results show that under the mildest assumptions on the model, the error of the joint density approximation is asymptotically negligible if $p = o(n^{1/4})$ for the Laplace approximation and saddlepoint approximation, with improvements being possible under additional assumptions, particularly for stratified models where a scaling of $p = o(n^{1/2})$ is possible. Stronger results are also obtained for the approximation to the marginal posterior density.

Factorized Estimation of High-dimensional Nonparametric Covariance Models
Speaker: **Jie Li, London School of Economics and Political Science**
Estimation of covariate-dependent conditional covariance matrix in a high-dimensional space poses a challenge to contemporary statistical research. The existing kernel estimators may not be locally adaptive due to using a single bandwidth to explore the smoothness of all entries of the target matrix function. In this paper, we propose a novel framework to address this issue, where we factorize the target matrix into factors and estimate these factors in turn by the kernel approach. The resulting estimator is further regularized by thresholding and optimal shrinkage. Under certain mixing and sparsity conditions, we show that the proposed estimator is well-conditioned and uniformly consistent with the underlying matrix function even when the sample is dependent. Simulation studies suggest that the proposed estimator significantly outperforms its competitors in terms of integrated root-squared estimation error. We present an application to financial return data.

Author Index