Gneiting, Tilmann

Heidelberg Institute for Theoretical Studies

Session: Tilmann Gneiting, IMS Wald Award and Lecture

Talk: *Assessing Monotone Dependence: Area Under the Curve Meets Rank Correlation*

The assessment of monotone dependence between two random variables is a classical problem in statistics and a gamut of application domains. Consequently, researchers have sought measures of association that are invariant under strictly increasing transformations of the margins, with the extant literature being splintered. Rank correlation coefficients, such as Spearman's Rho and Kendall's Tau, have been studied in the statistical literature, mostly under the assumption of continuous margins. In the case of a dichotomous outcome, receiver operating characteristic (ROC) analysis and the asymmetric area under the ROC curve (AUC) measure are used to assess monotone dependence of a binary outcome on a covariate. Here we unify and extend thus far disconnected strands of literature, by developing common population level theory, estimators, and tests that bridge continuous and dichotomous settings and apply to all linearly ordered outcomes. In particular, we introduce the asymmetric grade correlation (AGC) and coefficient of monotone association (CMA) measures, which correspond to Spearman's Rho in the continuous case and to AUC for a dichotomous outcome. We establish central limit theorems for their sample versions and develop associated tests. In case studies, we assess progress in data-driven weather prediction and evaluate methods of uncertainty quantification for large language models. Joint work with Eva-Maria Walz and Andreas Eberl.

Goeman, Jelle

Leiden University Medical Center

Session: Jelle Goeman, IMS Medallion Award and Lecture

Talk: *Principles and Flexibility in Multiple Testing*

Closed testing is known to underlie many multiple testing methods. Any method controlling a tail probability of a number or proportion of errors is either equivalent to a closed testing procedure or is uniformly improved by one. We extend closed testing to a novel necessary and sufficient principle for all multiple testing methods controlling any expected loss. This e-Closure principle asserts that every multiple testing method is a special case of a generalized closed testing procedure based on e-values. It applies to a large class of error rates — in particular to the false discovery rate (FDR). We show that any procedure controlling FDR is either a e-Closure method or is uniformly improved by one. By writing existing methods as special cases of this procedure, we show uniform improvements, as we demonstrate for the e-Benjamini-Hochberg and the Benjamini-Yekutieli procedures, and the self-consistent method of Su (2018). We also show that methods derived using our novel e-Closure Principle generally control their error rate not just for one rejected set, but simultaneously over many, allowing post hoc flexibility for the researcher. This flexibility is important in many application areas including neuroimaging and genomics. Moreover, we show that because all multiple testing methods for all error metrics are derived from the same procedure, researchers may even choose the error metric post hoc, allowing post hoc switching from e.g. familywise error to FDR.

McKeague, Ian

Columbia University

Session: Ian McKeague, IMS Medallion Award and Lecture

Talk: *What we can know from complex functional data*

This talk discusses a nonparametric inference framework for functional data having sample paths of bounded variation, with applications in a variety of complex settings. The main application will be to wearable device data collected in a Columbia-based study of an experimental therapy for mitochondrial disease, a group of disorders that affect the body's ability to produce energy. Specifically, we provide the first clinical application of a novel, bias--adjusted outcome measure of acceleration across a range of subjects' activities to assess nucleoside therapy for thymidine kinase 2 deficiency, an ultra--rare autosomal recessive mitochondrial disease. In addition, the talk will briefly touch on the potential (in terms of "what we can know") of having access to functional data collected along causal paths in a possibly non-smooth metric spacetime, which relaxes the smooth Lorentzian framework of Einstein's field equations to allow rougher settings.

Sen, Bodhisattva

Columbia University

Session: Bodhisattva Sen, IMS Medallion Award and Lecture

Talk: *Wasserstein-Cramer-Rao Theory of Unbiased Estimation*

The quantity of interest in the classical Cramer-Rao theory of unbiased estimation (i.e., the Cramer-Rao lower bound, exact efficiency in exponential families, and asymptotic efficiency of maximum likelihood estimation) is the variance, which represents the instability of an estimator when its value is compared to the value for an independently-sampled data set from the same distribution. In this talk we are interested in a quantity which represents the instability of an estimator when its value is compared to the value for an infinitesimal additive perturbation of the original data set; we refer to this as the "sensitivity" of an estimator. The resulting theory of sensitivity is based on the Wasserstein geometry in the same way that the classical theory of variance is based on the Fisher-Rao (equivalently, Hellinger) geometry, and this insight allows us to determine a collection of results which are analogous to the classical case: a Wasserstein-Cramer-Rao lower bound for the sensitivity of any unbiased estimator, a characterization of models in which there exist unbiased estimators achieving the lower bound exactly, and a guarantee that Wasserstein projection estimators achieve the lower bound asymptotically. We use these results to treat many statistical examples, sometimes revealing new optimality properties for existing estimators and other times revealing new estimators. This is joint work with Nicolas Garcia Trillos and Adam Jaffe.

Zhang, Cun-Hui

Rutgers University

Session: Cun-Hui Zhang, Blackwell Award and Lecture

Talk: *Empirical Bayes for Dependent Data*

Empirical Bayes is built on the idea that pooling information across related decision problems can outperform procedures that treat each problem in isolation. While this principle has had a profound impact, most of its development has focused on independent observations. In contrast, modern applications often involve measuring many attributes on the same sample, leading to compound decision problems with dependent data. In such settings, the benefit of pooling is clear, but the challenge is how to do so with limited knowledge of the dependence structure. A key observation is that the fundamental theorem connecting compound decision problems to empirical Bayes does not require independence. Building on this, we develop a nonparametric marginal likelihood approach under dependence. The effect of dependence is captured by the largest eigenvalue of the correlation matrix, which acts as a discount factor on the effective number of problems that can be pooled. The required dependence conditions are no stronger than for linear estimators. For estimating the density, score function, component means, and compound risk of the mean estimator, the proposed methods achieve nearly parametric rates of convergence to the oracle and are nearly minimax optimal in rate. They also provide consistent estimation of the prior, posterior, and credible intervals and adapt to the atomic sparsity of the oracle prior.

Akhavan, Arya

University of Oxford

Session: Unified Perspectives on Regularization in Modern Statistics

Talk: *Sampling-Based Regularisation for Adversarial Linear Bandits*

We study the design of sampling-based regularisation in sequential decision-making. In the online learning setting, we show that the exploration-exploitation trade-off can be optimally controlled by sampling from carefully designed probability distributions, providing a statistical alternative to explicit penalty functions. Focusing on adversarial linear bandits, we develop a unified framework that connects Follow-the-Regularized-Leader (FTRL) and Follow-the-Perturbed-Leader (FTPL), extending their known equivalence beyond the full-information setting. Within this framework, we introduce self-concordant perturbations, a family of distributions that mirror the role of self-concordant barriers used in the FTRL-based SCRiBLe algorithm. This yields a new FTPL algorithm that combines self-concordant regularisation with efficient stochastic exploration. The resulting method achieves regret $O(dn\ln n)$ on both the d-dimensional hypercube and the Euclidean ball. On the Euclidean ball, this matches the best known guarantees for self-concordant FTRL methods, while on the hypercube it improves them by a factor of d, matching the optimal rate up to logarithmic factors.

Amorino, Chiara

Universitat Pompeu Fabra

Session: Recent advances in statistics for stochastic processes

Talk: *Parameter estimation for graphon-interacting particle systems from discrete observations*

In this paper, we address the problem of joint parameter estimation for the drift and diffusion coefficients of heterogeneously interacting diffusive particle systems and their associated large population limits. Unlike the homogeneous setting, the interaction here is of mean-field type with weights characterized by an underlying graphon. This heterogeneity introduces significant analytical complexity: while homogeneous particles converge to an i.i.d. limit, our system leads to particles that are independent but not identically distributed in the limit. Starting from discrete observations of the system over a fixed interval $[0, T]$, we propose a contrast function based on a pseudo-likelihood approach. We establish that the resulting estimators are consistent as the discretization step $\Delta_n \to 0$ and the number of particles $N \to \infty$. Furthermore, we prove asymptotic normality under the additional regime $\Delta_n N \to 0$, demonstrating that the graphon-based structure can be successfully navigated despite the lack of identical distribution in the limit.

Ansari, Jonathan

University of Salzburg

Session: Measures of Statistical Association

Talk: *An ordering for the strength of functional dependence*

We introduce a new dependence order—the conditional convex order—whose minimal and maximal elements characterize independence and perfect dependence. Moreover, it characterizes conditional independence, satisfies information monotonicity, and exhibits several invariance properties. Consequently, it is an ordering for the strength of functional dependence of a random variable Y on a random vector X . As we show, various recently studied dependence measures—including Chatterjee's rank correlation, Wasserstein correlations, and rearranged dependence measures—are increasing in this order and inherit their fundamental properties from it. We characterize the conditional convex order by the Schur order and by the concordance order, and we verify it in settings such as additive error models, the multivariate normal distribution, and various copula-based models. Our results offer a unified perspective on the behavior of dependence measures across statistical models.

Aragam, Bryon

University of Chicago

Session: Causality Meets AI: Generative Structure, LLM‚ÄëAssisted Trials, and Spatial Heterogeneity

Talk: *Statistical foundations of representation learning in generative models*

One of the key paradigm shifts in statistical machine learning over the past decade has been the transition from handcrafted features to automated, data-driven representation learning. A crucial step in this pipeline is to identify latent and learn structured representations from data. In many applications, meaningful concepts are not directly observed, and must be learned from data, often using flexible, nonparametric models such as deep generative models. These settings present new statistical and computational challenges that will be focus of this talk. We will re-visit the statistical foundations of nonparametric latent variable models as a lens into the problem of identifying representations in deep generative models. We discuss our recent work on developing methods for identifying and learning causal representations from data with rigourous guarantees, and discuss how even basic statistical properties are surprisingly subtle. Along the way, we will explore the connections between deep generative models, nonparametric latent variable models, and causal graphical models.

Arlotto, Alessandro

Duke University

Session: Advances in Learning, Optimization, and Generalization in Modern Stochastic Systems

Talk: *Reoptimization in the dynamic and stochastic knapsack problem: from near-optimality to distributional equivalence*

We study a dynamic and stochastic knapsack problem in which a decision maker is sequentially presented with n items with unitary rewards and independent weights that are drawn from a known continuous distribution. The decision maker seeks to maximize the expected reward she collects by including items in a knapsack while satisfying a capacity constraint, and while making terminal decisions as soon as each item weight is revealed. We propose a reoptimized heuristic and compare its total rewards with that of the optimal dynamic programming policy. We show that the two total rewards have the same asymptotic mean, the same asymptotic variance, and the same limiting distribution. In contrast, we also note that other asymptotically optimal heuristics that have been considered in the literature have different (larger) higher moments and different limiting distributions. (Joint work with Xinchang Xie and Yun-Tung Kuo.)

Baldasso, Rangel

PUC-Rio

Session: Particle Systems in Random Environments

Talk: *Percolation on Hierarchical Lattices*

In this talk we examine families of hierarchical lattices, with focus on its fine percolative properties. Thanks to the self-similarity of the graphs in these sequences, several properties of the model can be examined, including: critical exponents, scaling relations, and noise sensitivity. Furthermore, for these models, we obtain explicit expressions for several critical exponents of percolation. The results are very sharp and proofs employ a combination of tools from probability and dynamical systems. Based on a joint work with Alves, Moreira, and Teixeira.

Bao, Zhigang

The University of Hong Kong

Session: Random Matrices

Talk: *Numerical Radius of Non-Hermitian Random Matrices*

For a square matrix, the range of its Rayleigh quotients is known as the numerical range, which is a compact and convex set by the Toeplitz–Hausdorff theorem. The largest value in this convex set is known as the numerical radius, which is often used to study the convergence rate of iterative methods for solving linear systems. In this talk, we will introduce a recent result on the asymptotic behavior of the numerical radius of a large-dimensional, complex, non-Hermitian random matrix and its elliptic variants. For the former, remarkably, the radius can be represented as the extremum of a stationary Airy-like process, which undergoes a correlation-decorrelation transition from a small to a large time scale. Based on this transition, we obtain the precise first and second order terms of the numerical radius. In the elliptic case, we show that the fluctuation of the numerical radius reduces to the maximum of two independent Tracy-Widom variables. Based on joint work with Giorgio Cipolloni.

Bartl, Daniel

National University of Singapore

Session: sub-Gaussian properties and robustness

Talk: *On the role of Rademacher complexities in statistical learning*

We study the problem of learning with respect to the squared loss over a convex class of functions. It has long been believed that the sample complexity in this setting is governed by localized Rademacher complexities. We show that, assuming access to coarse information on the covariance structure of the model class, the sample complexity is instead controlled by a localized complexity associated with the limiting Gaussian process. In heavy-tailed regimes, this quantity can be significantly smaller than the Rademacher complexity.

Belloni, Alexandre

Duke and Amazon

Session: Inference, Learning, and Decision-Making in Complex Systems

Talk: *NEIGHBORHOOD ADAPTIVE ESTIMATORS FOR CAUSAL INFERENCE UNDER NETWORK INTERFERENCE*

Estimating causal effects has become an integral part of most applied fields. In this work we consider the violation of the classical no-interference assumption with units connected by a network. For tractability, we consider a known network that describes how interference may spread. Unlike previous

work the radius (and intensity) of the interference experienced by a unit is unknown and can depend on different (local) sub-networks and the assigned treatments. We study estimators for the average direct treatment effect on the treated in such a setting under additive treatment effects. We establish rates

of convergence and distributional results. The proposed estimators considers all possible radii for each (local) treatment assignment pattern. In contrast to previous work, we approximate the relevant network interference patterns that lead to good estimates of the interference. To handle feature engineering, a key innovation is to propose the use of synthetic treatments to decouple the

dependence. We provide simulations, an empirical illustration and insights for the general study of interference.

Ben-Hamou, Anna

Sorbonne University

Session: Particle Systems in Random Environments

Talk: *Cutoff for independent random walks on the circle conditioned not to intersect*

In this talk, we will consider a class of Markov processes on the discrete circle which has been introduced by König, O'Connell and Roch. These processes describe movements of exchangeable interacting particles and are discrete analogues of the unitary Dyson Brownian motion: a random number of particles jump together either to the left or to the right, with trajectories conditioned to never intersect. We provide asymptotic mixing times for stochastic processes in this class as the number of particles goes to infinity, under a sub-Gaussian assumption on the random number of particles moving at each step. As a consequence, we prove that a cutoff phenomenon holds independently of the transition probabilities, subject only to the sub-Gaussian assumption and a minimal aperiodicity hypothesis.

Blath, Jochen

Goethe Universität Frankfurt

Session: Interdisciplinary perspectives at the interface between probability theory, ecology and evolution

Talk: *Emergence of host dormancy in the presence of a persistent virus epidemic*

We study a stochastic individual-based model for a microbial population challenged by a persistent virus epidemic. We assume that the resident microbial host population and the virus population are in stable coexistence upon arrival of a single new “mutant” host individual that is capable of switching into a reversible state of dormancy upon contact with virions as a means of avoiding infection. At the same time, we assume that this new dormancy trait comes with a cost, namely a reduced individual reproduction rate while in the active state. We prove that the mutants can nevertheless invade the resident population with strictly positive probability in the large population limit.

Given the reduced reproductive rate, such an invasion would be impossible in the absence of the virus epidemic. We explicitly characterize the parameter regime where this emergence of a host dormancy trait is possible, determine the success probability of a single invader and the typical amount of time it takes the successful mutants to reach a macroscopic population size. We conclude with an investigation of the fate of the population after the successful emergence of a dormancy trait. Here, either both host types and the virus will reach coexistence (the most interesting scenario), or the mutant hosts will drive the resident hosts to extinction.

This is joint work with András Tóbías, Budapest.

Bradic, Jelena

Cornell University

Session: Inference Beyond the Textbook: Robustness, Adaptivity, and Cross-Fitting

Talk: *Surrogate-powered Causal Inference*

Missing information is intrinsic to causal inference with survival outcomes: Experiments are con-

strained by budgets and timelines; observational studies face drop-out. In addition to the unobserved

counterfactual outcome, the factual event time is subject to right censoring. This degrades the effective

sample size even when identification is secured by design or by assumption. We study whether an in-

termediate post-treatment event, e.g., recurrence or progression, can recover the power for the primary

outcome estimand that is lost to censoring. With full-data and absent semiparametric restrictions, surrogate outcomes cannot deliver a first-order efficiency gain. With missing outcomes, surrogates and baseline covariates become informative for marginal causal parameters that, absent missing data, do not depend on the surrogate. The size of the gain depends on: (i) the balance of the post-treatment stratification induced by the surrogate, (ii) the prognostic separation of the strata, (iii) the temporal proximity of the stratification with the censoring.

Buchweitz, Erez

UC Berkeley

Session: Proper scoring rules

Talk: *Calibration and minimizing loss*

Are calibration and minimizing loss compatible goals? Is calibration desirable for minimizing loss? Does minimizing loss lead to calibration?

We study these questions via tradeoff and recalibration arguments and show that, all other things equal, the answer to all three questions is yes for various calibration modes and proper losses. But, in general, miscalibrated forecasts can be superior to calibrated ones, even across all proper losses at the same time. Furthermore, excluding auto-calibration, the answer can be loss-dependent. We introduce a new family of calibration modes that generalize marginal calibration in a unique way for each proper loss and arise naturally from minimizing loss. This family enjoys explicit tradeoffs that are analogous to the bias-variance tradeoff in regression when bias is identified with miscalibration.

Bunea, Florentina

Cornell University

Session: Statistical theory and algorithms for AI

Talk: *BabyTransformer CODL learns latent domains*

Token-level contextual embeddings for text corpora are now readily available for both natural language and genetic sequences. We take a top-down perspective and introduce CODL, a continuous topic model that leverages these embeddings to uncover the latent domains represented in a corpus. The model is a very large p-mixture of high-dimensional Gaussian distributions, where the mixture weights themselves are mixtures of K Softmax probability vectors supported on the p Gaussian means.

By first learning the many latent Gaussian means from the tokenized corpus, we construct a corpus-specific vocabulary. This vocabulary forms the support of a discrete K-Softmax mixture ensemble whose atoms encode the corpus' latent domains in the embedding space. These learned softmax atoms can then be mapped back to natural language—via the corpus-specific vocabulary—enabling interpretable topic discovery.

A central component of our procedure is a new method for learning the latent domains

from data, prior to full model fitting. We show that this reduces to detecting whether a Softmax mixture is primarily concentrated on a single leading component. To achieve this, we introduce a detection criterion that directly estimates functionals of a Softmax mixing measure, bypassing the challenging task of estimating the mixing measure itself in high dimensions.

We provide end-to-end theoretical guarantees for EM-based atom estimates, and illustrate their interpretability by data examples.

Callegaro, Alice

Technical university of Munich

Session: Random spatial systems

Talk: *Linear spreading speed of a branching annihilating random walk*

We study a branching annihilating random walk in which particles evolve on the lattice in discrete generations. Each particle produces a Poissonian number of offspring which independently move to a uniformly chosen site within a fixed distance from their parent's position. Whenever a site is occupied by at least two particles, all the particles at that site are annihilated. This can be thought of as a very strong form of local competition and implies that the system is not monotone. For certain ranges of the parameters of the model, we show that the system dies out almost surely or, on the other hand, survives with positive probability. In an even more restricted parameter range, we strengthen the survival results to complete convergence with a non-trivial invariant measure, and in the same regime we show existence of a linear spreading speed for the model on the integer line. A central tool in the proofs is comparison with oriented percolation on a coarse-grained space-time lattice, using carefully tuned density profiles which expand in time and are reminiscent of discrete travelling wave solutions. Our results extend to (possibly non-monotone) models for which comparison with oriented percolation and other coupling techniques are possible. Based on joint works with Matthias Birkner (Mainz), Jiří Černý (Basel), Nina Gantert (TU Munich) and Pascal Oswald.

Centofanti, Fabio

KU Leuven

Session: Robust Statistics for Matrix and Tensor Data

Talk: *Casewise and Cellwise Robust Multilinear Principal Component Analysis*

Multilinear Principal Component Analysis (MPCA) is an important tool for analyzing tensor data. It performs dimension reduction similar to PCA for multivariate data. However, standard MPCA is sensitive to outliers. It is highly influenced by observations deviating from the bulk of the data, called casewise outliers, as well as by individual outlying cells in the tensors, so-called cellwise outliers. This latter type of outlier is highly likely to occur in tensor data, as tensors typically consist of many cells. This talk introduces a novel robust MPCA method that can handle both types of outliers simultaneously, and can cope with missing values as well. This method uses a single loss function to reduce the influence of both casewise and cellwise outliers. The solution that minimizes this loss function is computed using an iteratively reweighted least squares algorithm with a robust initialization. Graphical diagnostic tools are also proposed to identify the different types of outliers that have been found by the new robust MPCA method. The performance of the method and associated graphical displays is assessed through simulations and illustrated on two real datasets.

Chan, Kin Wai

The Chinese University of Hong Kong

Session: Recent developments in change-point analysis

Talk: *Doubly self-normalized change point test under local stationarity via sparse subsampling and stationary pooling*

This project studies change point tests for locally stationary time series based on sparse subsampling and stationary pooling. First, studentized data are sparsely subsampled in such a way that all selected observations are lagged by a divergent amount. Consequently, a studentized sparse cumulative sum (CUSUM) process is constructed for testing change points. The resulting test is proven to be asymptotically pivotal. This approach automatically transforms a change point test designed for stationary conditions into one suitable for local stationarity, while bypassing change point-sensitive bootstrapping or computationally intensive simulations. Second, the p-values computed from different studentized sparse CUSUM processes initiated at various time points are shown to be approximately stationary and satisfy a functional central limit theorem. A method called stationary pooling is proposed to efficiently aggregate these approximately stationary p-values to enhance statistical power. Stationary pooling is shown to retrieve the change point detection power as in the stationary setting. Building on sparse subsampling and stationary pooling, we propose a doubly self-normalized test that employs self-normalization twice for two layers of aggregation: across sparsifying lags and along consecutive initial time points. Simulations demonstrate improvements in power, size accuracy, and computational efficiency.

Chandak, Rajita

École Polytechnique Fédérale de Lausanne (EPFL)

Session: New Researchers Group Session

Talk: *Convergence of the EM algorithm beyond classical Gaussian regimes*

The EM algorithm has been widely studied theoretically for finite Gaussian mixture models. However, the practical use cases of the EM algorithm span beyond these specialized theoretical models. In this talk we will cover recent advances in theory of the EM algorithm beyond the classical setting in two primary regimes. First, we will show the advantages of generalizing the EM algorithm to the Federated Learning regime wherein data heterogeneity poses a significant challenge to the applicability of classical statistical methods. We characterize the convergence rate of the EM algorithm under all regimes of data heterogeneity under finite mixtures of linear regressions. We show that rather than being a bottleneck, in some cases, data heterogeneity can suprisingly accelerate the convergence of iterative federated algorithms. Secondly, we delve briefly into the behaviour of the EM algorithm under mixtures of non-Gaussian densities. We show in particular that even with slower tail decay, the convergence rates of the EM algorithm for mixture of densities or linear regressions do not suffer significantly, providing the foundations for extending many existing results to a larger class of mixture models.

Chen, Yaqing

Rutgers University

Session: Recent advances in functional data analysis and metric statistics

Talk: *Modeling Amplitude and Phase Variation of Multivariate Random Processes in Metric Spaces*

For real-valued functional data, it is well known that failing to distinguish between amplitude variation and phase variation can distort subsequent statistical analysis, and extensive work has been devoted to developing time-warping methods to address this issue. However, much less is known about how to characterize and handle these two sources of variability when they co-exist in random processes taking values in general metric spaces, which lack inherent linear structure, particularly in the multivariate setting. In this paper, we formalize the concepts of amplitude and phase variation for multivariate random processes in geodesic spaces and propose a latent deformation model for jointly analyzing both types of variation. We establish the asymptotic convergence rates for the model estimators and demonstrate the versatility of the proposed method through simulation studies on positive semi-definite (PSD) matrix-valued data and network-valued data, as well as a real-world application to annual sex-specific age-at-death distributions across countries, providing new insights into global longevity dynamics.

Chen, Yong

University of Pennsylvania

Session: Statistical inference for data science and AI

Talk: *MOSAiC: a unified framework for lossless, one-shot, federated learning algorithms*

Multi-site studies are now central to biomedical research, but inference across sites remains challenging due to regulatory limits on sharing individual-level data, heterogeneous data distributions, sparse events in small centers, and the logistical burden of multi-round communication. In this talk, I introduce MOSAiC (Multi-site One-Shot Aggregation of Compressed Risk Functions), a unified framework for modern distributed research networks. Notably, MOSAiC reframes distributed learning as a mathematical problem of compressing and aggregating risk functions, leveraging recent advances in tensor networks—state-of-the-art tools for high-dimensional function approximation in scientific computing. Our MOSAiC enjoys four desirable properties that have never been achieved by any of the existing federated algorithms (except linear regressions): one-shot communication, lossless recovery of pooled-data estimates, inclusiveness of all sites irrespective of size or event rarity, and analytic submodel exploitability without re-querying partners. I will illustrate MOSAiC's validity and efficiency through applications in drug relabeling, drug repurposing, and post-market safety surveillance.

Chen, Yuansi

ETH Zurich

Session: Stochastic Analysis and Applications

Talk: *Recent progress on Talagrand's convolution conjecture*

In 1989, Talagrand conjectured that under the heat semigroup on the Boolean hypercube, any nonnegative function on the cube exhibits a uniform tail bound that is better than that by Markov's

inequality. We discuss recent progress on the conjecture and the reverse heat process used in the proof.

Cheng, Xiuyuan

Duke University

Session: Generative Models for Statistical Inference

Talk: *Worst-case generation via minimax optimization in Wasserstein space*

Worst-case generation plays a critical role in evaluating robustness and stress-testing systems under distribution shifts, in applications ranging from machine learning models to power grids and medical prediction systems. We develop a generative modeling framework based on min-max optimization over continuous probability distributions, namely the Wasserstein space. Unlike traditional discrete distributionally robust optimization approaches, which often suffer from scalability issues, limited generalization, and costly worst-case inference, our framework exploits Brenier's theorem to characterize the least favorable distribution as the pushforward of a transport map from a continuous reference measure. Based on the min-max formulation, we propose a Gradient Descent Ascent (GDA)-type scheme that updates the decision model and the transport map alternately, establishing global convergence guarantees without requiring convexity-concavity type assumptions. We also propose to parameterize the transport map using a neural network that can be trained simultaneously with the GDA iterations by matching the transported training samples, thereby achieving a simulation-free approach. The efficiency of the proposed method as a worst-case generator is validated by numerical experiments on synthetic and image data.

Chevyrev, Ilya

SISSA (International School for Advanced Studies)

Session: Three perspectives on singular stochastic PDEs

Talk: *Large field problem in coercive SPDE via scaling*

In this talk, I will show an approach to deriving a priori bounds for coercive SPDEs based on scaling. The basic idea is to first show bounds for the equation with a small noise and then rescale the bounds to a global scale. While many equations that the approach can handle have been treated recently with other methods, its advantages are that it is quite simple and allows one to state a single result that is applicable to a variety of equations, such as rough differential equations and parabolic/elliptic SPDEs. Based on joint work with Massimiliano Gubinelli.

Coja-Oghlan, Amin

TU Dortmund

Session: Combinatorial aspects of random matrix theory

Talk: *The random XORSAT decimation process*

We analyse the performance of Belief Propagation Guided Decimation, a physics-inspired message passing algorithm, on the random k-XORSAT problem. Specifically, we derive an explicit threshold up to which the algorithm succeeds with a strictly positive probability that we compute explicitly, but beyond which the algorithm with high probability fails to find a satisfying assignment. In addition, we analyse a thought experiment called the decimation process for which we identify a (non-) reconstruction and a condensation phase transition. The main results of the present work confirm physics predictions from [Ricci-Tersengh, Semerjian: J. Stat. Mech. 2009] that link the phase transitions of the decimation process with the performance of the algorithm, and improve over partial results from a recent article [Yung: Proc. ICALP 2024].

Joint work with Arnab Chatterjee, Amin Coja-Oghlan, Mihyun Kang, Lena Krieg, Maurice Rolvien, Gregory B. Sorkin.

Collamore, Jeffrey

University of Copenhagen

Session: Probability Foundations for Robust and Private Learning

Talk: *Large Deviations and Bahadur Efficiency for General Divergence-based Estimators*

Minimum divergence estimators have long been studied as robust alternatives to maximum likelihood methods; see, for example, Beran (1977) and Lindsay (1994). In this work we investigate, from a general perspective, the rare-event behavior of such divergence-based inferential methods. Specifically, we establish large deviation estimates for the probabilities that a divergence-based estimator deviates from the true parameter values, and we characterize the associated exponential rate of decay as the sample size tends to infinity. This decay is described via a large deviation rate function whose form depends on the choice of the divergence measure. We then apply these results to study Bahadur efficiency for general divergence estimators, thereby extending and complementing earlier work of Bahadur (1967) and Arcones (2006).

DALALYAN, Arnak

CREST / ENSAE / IP Paris

Session: Statistical theory and algorithms for AI

Talk: *Assessing the Quality of Denoising Diffusion Models in Wasserstein Distance*

Generative modeling aims to produce new random examples from an unknown target distribution, given access to a finite collection of examples. Among the leading approaches, denoising diffusion probabilistic models (DDPMs) construct such examples by mapping a Brownian motion via a diffusion process driven by an estimated score function. In this work, we first provide empirical evidence that DDPMs are robust to constant-variance noise in the score evaluations. We then establish finite-sample guarantees in Wasserstein-2 distance that exhibit two key features: (i) they characterize and quantify the robustness of DDPMs to noisy score estimates, and (ii) they achieve faster convergence rates than previously known results. Furthermore, we observe that the obtained rates match those known in the Gaussian case, implying their optimality.

Darshan, Shiva

ESSEC Business School

Session: Stochastic Analysis and Applications

Talk: *A non-equilibrium equivalence of ensembles result via propagation of chaos*

We consider an alternative formulation of non-equilibrium stochastic dynamics referred to as Norton dynamics. A reference system is perturbed by an external forcing with varying intensity so that the value of a given observable is fixed. This amounts to constraining the dynamics to a level set of this observable. For observables that are linear functions of the empirical measure of the dynamics, this can be viewed as a constraint on the empirical mean of the dynamics. This leads us to consider mean-field interacting particle systems. We show via a propagation of chaos that an equivalence of ensembles holds between the Norton formulation of non-equilibrium dynamics and the standard formulation in which the intensity of the forcing is fixed. We start by showing the well-posedness of the finite N mean--field Norton dynamics and the corresponding McKean-Vlasov dynamics. This novel non-linear diffusion has a non-linear drift term due to an oblique mean-constraint. This dynamics has a continuum of invariant probability measures which are the invariant probability measures of the McKean-Vlasov dynamics corresponding to the standard formulation of non-equilibrium dynamics. Equivalence of ensembles is then deduced from the fact that the McKean-Vlasov dynamics of the two formulations coincide when started at the same invariant probability measure.

de Heide, Rianne

University of Twente and Centrum Wiskunde en Informatica, Amsterdam

Session: E-Values: local, compound and conditional

Talk: *Local e-values for intersection hypotheses*

The recent e-closure framework (Xu et. al., 2025) shows that multiple testing procedures controlling an expected loss can be understood through families of local e-values assigned to intersection hypotheses, yielding a unifying route to valid and often uniformly improved procedures. In this talk I will show new FDR controlling tests for settings in which it was not known before that such a test would exist.

Xu, Z., Solari, A., Fischer, L., de Heide, R., Ramdas, A., and Goeman, J. J. (2025). Bringing Closure to False Discovery Rate Control: A General Principle for Multiple Testing. arXiv:2509.02517

de Roos, Dante

Centrum Wiskunde & Informatica

Session: E-Values: local, compound and conditional

Talk: *Rao-Blackwellized e-variables*

The Rao-Blackwell theorem is a cornerstone result in classical statistics which, in its earliest form, states that the mean squared error (MSE) of an (unbiased) estimator decreases after conditioning on a sufficient statistic. This conditioning has thus come to be known as "Rao-Blackwellization." In this talk we present an analogous result for e-variables, namely that any e-variable can be Rao-Blackwellized into a new e-variable that, for every concave utility function, attains greater expected concave utility than the original. The simplest version of this Rao-Blackwell theorem for e-variables holds in full generality, but its interpretation and applicability can be limited without suitable integrability assumptions. For the log-utility, we present a more sophisticated version that overcomes these limitations and is therefore fully applicable. The conceptual breadth of these Rao-Blackwell theorems for e-variables will be illustrated with an example on testing the variance in a linear regression model.

Deijfen, Maria

Stockholm University

Session: Random processes on graphs

Talk: *Surviving from the tip of a cone*

In two-type first passage percolation on the square lattice, two entities compete to capture the sites of the lattice. The entities spread between nearest neighbor sites at times specified by random passage times associated with the edges. We consider the case when both types have the same passage time distribution, with one type starting at the origin and the other from an infinite cone with tip at the origin. Can the type starting at the origin grow unboundedly? We give a partial answer to this question in terms of the slope of the cone.

Demler, Olga

ETH, Zurich / BWH, Harvard Medical School, Boston, USA

Session: Recent Developments in Head-to-Head Comparisons and Stochastic Preference Modeling

Talk: *Pairwise Comparison Metrics: AUC, Win Ratio, and the Problem of Non-Transitivity*

The area under the receiver operating characteristic curve (AUC) is a standard measure of discrimination for risk prediction models. Despite its appealing probabilistic interpretation, its limitations are often overlooked. In this talk, we examine several challenges in the use of AUC, including its sensitivity to study design, the distribution of predictors, and the presence of competing events. We also highlight pitfalls in external validation, where AUC may not be directly comparable across populations due to differences in case-mix. These considerations underscore that AUC is not an intrinsic property of a model, but depends on the underlying data-generating mechanism.

We further investigate AUC and the Win Ratio as members of a pairwise comparison group of statistics. Under certain conditions, the Win Ratio can be expressed as a monotonic transformation of AUC. However, as a head-to-head comparison measure, it may exhibit non-transitivity, a phenomenon related to classical results of Trybuła. We discuss the implications of this property for the interpretation of treatment effects, particularly in randomized trials and meta-analysis.

Dmitriev, Daniil

University of Pennsylvania

Session: Bridging Statistics and Modern AI: Foundations for Deep Learning and Generative Models

Talk: *Efficient sampling with discrete diffusion models: Sharp and adaptive guarantees*

Diffusion models over discrete spaces have recently shown striking empirical success, yet their theoretical foundations remain incomplete. In this talk, we discuss the sampling efficiency of score-based discrete diffusion models under a continuous-time Markov chain (CTMC) formulation, with a focus on tau-leaping-based samplers. We establish sharp convergence guarantees for attaining small Kullback-Leibler (KL) divergence for both uniform and masking noising processes. For uniform discrete diffusion, we show that the tau-leaping algorithm achieves an almost linear iteration complexity in the ambient dimension of the target distribution, improving the quadratic dependency from the prior work and eliminating linear dependency on the vocabulary size. Moreover, we establish a matching algorithmic lower bound showing that linear dependence on the ambient dimension is unavoidable in general. For masking discrete diffusion, we introduce a modified tau-leaping sampler whose convergence rate is governed by an intrinsic information-theoretic quantity, termed the effective total correlation, which can be sublinear or even constant for structured data. As a consequence, the sampler provably adapts to low-dimensional structure without prior knowledge or algorithmic modification, yielding sublinear convergence rates for various practical examples (such as hidden Markov models, image data, and random graphs). Joint work with Zhihan Huang and Yuting Wei.

Dong, Jing

Columbia University

Session: Inference, Learning, and Decision-Making in Complex Systems

Talk: *Right-Sizing Communication and Recommendation Set Size in AI-Assisted Search*

We model the interaction between a user and an AI-driven recommendation system. The user initiates the process by conveying preference information through a costly and noisy message. The AI assistant, acting as a Bayesian agent, interprets the user's message to form a posterior belief about their preferences and make product recommendations. In particular, it determines how many recommendations to present to maximize the user's expected utility from their final choice, while accounting for the search cost. We use a mutual information-based cost to model the two distinct costs incurred by the user: a communication cost, which increases with the precision of the preference message, and a search cost, which increases with the size of the recommendation set. We study the high-dimensional setting and characterize how optimal message precision and recommendation set size depend on the cost parameters, under two distinct distributions from which recommendations can be sampled from the product universe: Bayes' posterior belief and an optimized tilted distribution. Under Bayes' posterior sampling, we identify a hybrid regime in which an efficient interaction policy requires jointly optimizing the amount of information conveyed by the user and the number of recommendations provided by the AI assistant. In the tilted sampling, our results show that the optimal interaction policy uses only one of communication and search, favoring whichever of them is less costly.

Drogin, Reuben

Yale University

Session: Random Matrices

Talk: *Localization in Quantum and Classical Systems*

Random band matrices are Hermitian matrices with random entries supported in a band of width W around the diagonal. The eigenfunctions of such matrices are expected to decay exponentially at the scale W^2 in dimension one, and exp(CW^2) in dimension two. Remarkably, the same scaling is expected for the cycle lengths in various models of random permutations where points are typically displaced by distances of order W. In this talk I will discuss some recent progress on these problems and some open questions.

Drton, Mathias

Technical University of Munich

Session: Advances in Unsueprvised Learning & Causal Discovery

Talk: *Parameter identification in linear non-Gaussian causal models under general confounding*

Linear non-Gaussian causal models postulate that each random variable is a linear function of parent variables and non-Gaussian exogenous error terms. We study identification of the linear coefficients when such models contain latent variables. Our focus is on the commonly studied acyclic setting, where each model corresponds to a directed acyclic graph (DAG). For this case, prior literature has demonstrated that connections to overcomplete independent component analysis yield effective criteria to decide parameter identifiability in latent variable models. However, this connection is based on the assumption that the observed variables linearly depend on the latent variables. Departing from this assumption, we treat models that allow for arbitrary non-linear latent confounding. Our main result is a graphical criterion that is necessary and sufficient for deciding the generic identifiability of direct causal effects. Moreover, we provide an algorithmic implementation of the criterion with a run time that is polynomial in the number of observed variables. Finally, we report on estimation heuristics based on the identification result and explore a generalization to models with feedback loops.

Du, Hang

MIT

Session: Statistical Inference on Preferential Attachment Networks

Talk: *A Proof of The Changepoint Detection Threshold Conjecture in Preferential Attachment Models*

We investigate the problem of detecting and estimating a changepoint in the attachment function of a network evolving according to a preferential attachment model on n vertices, using only a single final snapshot of the network. Bet et al. show that a simple test based on thresholding the number of vertices with minimum degrees can detect the changepoint when the change occurs at time $n-\Omega(\sqrt{n})$. They further make the striking conjecture that detection becomes impossible for any test if the change occurs at time $n-o(\sqrt{n}).$ Kaddouri et al. make a step forward by proving the detection is impossible if the change occurs at time $n-o(n^{1/3}).$ In this paper, we resolve the conjecture affirmatively, proving that detection is indeed impossible if the change occurs at time $n-o(\sqrt{n}).$ Furthermore, we establish that estimating the changepoint with an error smaller than $o(\sqrt{n})$ is also impossible, thereby confirming that the estimator proposed in Bhamidi et al. is order-optimal.

Du, Jin-Hong

University of Hong Kong

Session: IMS Lawrence D. Brown Ph.D. Student Award

Talk: *Seeing through correlations: Disentangled feature importance*

Quantifying feature importance with valid statistical uncertainty is central to interpretable machine learning, yet classical model-agnostic methods often fail under feature correlation, producing unreliable attributions and compromising statistical inference. Existing approaches—such as Shapley values and leave-one-covariate-out—are designed for feature selection and vulnerable to correlation distortion, limiting their robustness on model interpretation. We introduce Disentangled Feature Importance (DFI), a model-agnostic framework that resolves these limitations by combining principled statistical inference with computational flexibility. DFI leverages entropic optimal transport to learn flexible disentanglement maps and provide an interpretable pathway for understanding how importance is attributed through the data's correlation structure. The framework generalizes to flow matching and differentiable loss functions, enabling statistically valid importance assessment for black-box predictors in both regression and classification. We establish statistical inference theory, which enables valid confidence intervals and hypothesis testing with Type I error control. Empirical results on synthetic and biomedical datasets show that DFI delivers substantially higher statistical power than removal-based and conditional permutation methods, while maintaining robust, interpretable attributions under severe feature interdependence.

Duan, Congyuan

Hong Kong University of Science and Technology

Session: Advances in Stochastic Gradient Descent: Algorithms and Asymptotics

Talk: *Online policy learning and inference by matrix completion*

Is it possible to make online decisions when personalized covariates are unavailable? We take a collaborative-filtering approach for decision-making based on collective preferences. By assuming low-dimensional latent features, we formulate the covariate-free decision-making problem as a matrix completion bandit. We propose a policy learning procedure that combines an eps-greedy policy for decision-making with an online gradient descent algorithm for bandit parameter estimation. Our novel two-phase design balances policy learning accuracy and regret performance. For policy inference, we develop an online debiasing method based on inverse propensity weighting and establish its asymptotic normality. Our methods are applied to data from the San Francisco parking pricing project, revealing intriguing discoveries and outperforming the benchmark policy.

Dubey, Paromita

University of Southern California

Session: Inference Without Borders: Methods for Random Objects Beyond Euclidean Data

Talk: *Minimax Optimal Conditional Two Sample Test via Distance Profiles*

We study the problem of comparing the conditional distributions of two random variables X and Y given a confounding variable Z, where all variables take values in general metric spaces. We introduce a new discrepancy measure that quantifies conditional distributional differences through joint distance profiles and develop a consistent estimator of this quantity. Building on this estimator, we construct a conditional two-sample test that avoids local smoothing and, therefore, remains simple, broadly applicable, and powerful against general alternatives. We further propose an algorithm to calibrate our test, which is both Pitman efficient and minimax-rate optimal against a large class of alternatives. Through simulations and real data illustrations—including a comparison of the distributions of outputs produced by different large language models using the same prompts—we demonstrate that the proposed method outperforms existing approaches.

Dukes, Oliver

Ghent University

Session: Royal Statistical Society Session

Talk: *Optimal binning for treatment effect heterogeneity*

Recent work has focused on nonparametric estimation of conditional average treatment effects (CATE) and learning optimal treatment regimes. However, inference has remained relatively unexplored, and may be challenging when the CATE depends on more than a handful of variables.

In this talk I will present a flexible workflow for inference on treatment effect heterogeneity in clinical trials and observational studies. This begins a nonparametric hypothesis test of homogeneity using a class of decision rules. By inverting the test, one can obtain the optimal decision rule with provable regret guarantees. I will describe on-going work on study design, how to adaptively refine the class of rules and how to obtain a confidence interval on the treatment effect in those that benefit from treatment.

Dutta, Ritabrata

University of Warwick

Session: Proper scoring rules

Talk: *Generalised Bayesian model selection*

Scritly proper scoring rules (SPSR), e.g. Energy score, Kernel score, are widely used as diagnostics of probabilistic forecasting. As we can derive statistical divergences from SPSRs, recently they have become popular for the purpose of inference in both frequentist (as minimum scoring rules estimator) and Bayesian framework (as generalised posterior) in the absence of tractable likelihood functions via replacement of negative log-likelihoods with SPSRs. Here we explore a generalised Bayesian model selection framework for models without tractable likelihood functions, using Bayes factor derived from generalised posteriors. To compute the generalised marginal evidence under each model, we propose a path sampling based estimator in conjunction with sequential Monte Carlo sampling scheme. We study the consistency of the proposed procedure theoretically and using simulation studies, finally illustrating their use for the choice of challenging biological models without tractable likelihood functions.

Fan, Zhou

Yale University

Session: Recent Advances in Iterative Methods for Random Optimization

Talk: *High-dimensional learning dynamics of multi-pass Stochastic Gradient Descent in multi-index models*

We study the learning dynamics of a multi-pass, mini-batch Stochastic Gradient Descent (SGD) procedure for empirical risk minimization in high-dimensional multi-index models with isotropic random data. In an asymptotic regime where the sample size n and data dimension d increase proportionally, for any sub-linear batch size kappa ~ n^alpha where alpha in [0,1), and for a commensurate "critical" scaling of the learning rate, we provide an asymptotically exact characterization of the coordinate-wise dynamics of SGD. This characterization takes the form of a system of dynamical mean-field equations, driven by a scalar Poisson jump process that represents the asymptotic limit of SGD sampling noise. We develop an analogous characterization of the Stochastic Modified Equation (SME) which provides a Gaussian diffusion approximation to SGD.

Our analyses imply that the limiting dynamics for SGD are the same for any batch size scaling alpha in [0,1), and that under a commensurate scaling of the learning rate, dynamics of SGD, SME, and gradient flow are mutually distinct, with those of SGD and SME coinciding in the special case of a linear model. We recover a known dynamical mean-field characterization of gradient flow in a limit of small learning rate, and of one-pass/online SGD in a limit of increasing sample size n/d to infinity.

Fels, Maximilian

Universität zu Köln

Session: Random spatial systems

Talk: *Gaussian Free Field with Hard Wall Constraints From Discrete Trees to Planar Domains*

The Discrete Gaussian Free Field (DGFF) serves as a canonical model for random interfaces in statistical mechanics. When conditioned to stay non-negative on a domain—an event known as the "hard wall" constraint—the field undergoes entropic repulsion, lifting away from the wall to accommodate its fluctuations.

The study of this phenomenon in two dimensions was pioneered by Bolthausen, Deuschel, and Giacomin (2001), who established the leading-order asymptotics for the probability of the hard wall event. In this talk, I will present two major recent advancements that significantly refine our understanding of this problem:

The Binary Tree (Fels, Hartung, Louidor 2024): We first discuss the case of the DGFF on the binary tree, where the hierarchical structure permits a complete derivation of the conditioned field's behavior. We will show that the hard wall constraint leads to a repulsion profile that makes the conditioned law asymptotically mutually singular with respect to the unconditioned law.

The 2D Plane (Fels, Louidor, Wu 2026): Turning to the Euclidean lattice $\mathbb{Z}^2$, I will present very recent work that improves upon the classical BDG result by capturing the subleading order of the probability asymptotics. The analysis relies on a novel orthogonal decomposition of the field into a "constant" component and a conditioned remainder, handled via multi-scale analysis and double-exponential tail bounds.

Feng, Oliver

London School of Economics

Session: The many facets of score matching

Talk: *Learning the score under shape constraints*

Score estimation has recently emerged as a key statistical challenge due to its pivotal role in generative modelling via diffusion models, and as an essential ingredient in a new approach to linear regression via convex M-estimation. Motivated by these applications, we study the minimax risk of estimating the score function of a log-concave density. On its own, this shape constraint is insufficient to guarantee a finite minimax risk with respect to a density-weighted L^2 norm, so we define subclasses capturing two fundamental aspects of the problem. First, we establish the crucial impact of tail behaviour by determining the minimax rate over log-concave densities whose score function exhibits controlled growth relative to the quantile levels. Second, we explore the interplay between smoothness and log-concavity by considering densities with a scale restriction and a Hölder assumption on the log-density. When the smoothness exponent is less than 2, the minimax rate is faster than under either constraint alone. Our upper bounds are attained by a locally adaptive, multiscale estimator constructed from a uniform confidence band for the score function. This study highlights intriguing differences between score and density estimation.

Freidling, Tobias

EPFL

Session: Post-selection inference

Talk: *Selective Inference in Adaptive Trials*

Many clinical trials follow a design with multiple stages: After each stage, the data is provisionally analysed and – based on these results – the recruitment of participants for the next stage as well as the administered treatment is chosen adaptively. For instance, we may want to exclude poorly performing drugs early or gather more samples from a certain subpopulation that shows a potentially beneficial response.

Analysing such adaptive studies is challenging as the data is used twice: (1) for selection of the design of later stages and the null hypothesis, (2) for testing the null hypothesis with the data generated under the chosen design. Since the data generating mechanism and null hypothesis are not pre-specified, classical statistical methods do not provide valid inference.

Existing solutions such as data splitting often lose efficiency or are specific to a certain design. In this presentation, we advocate for the use of selective inference methods to construct more powerful tests that control the selective type-I error. Randomization- or design-based inference is particularly amenable to this approach: We can provide finite-sample valid selective inference for a wide range of trial designs without any assumptions on the law of the outcomes and covariates. We show that our proposed method improves power compared to other valid tests, construct confidence intervals and address computational aspects.

Fritsch, Coralie

Inria

Session: Interdisciplinary perspectives at the interface between probability theory, ecology and evolution

Talk: *Niche construction emerging from fast ecological and slow evolutionary and environmental timescales in individual-based models*

I will introduce an individual-based model of niche construction based on birth-death processes of interacting (sub)species immersed in a slowly-varying environment which is influenced by the population state. Extinction and/or emergence of negligible (sub)species on long time scales can be observed. I will present a convergence result showing that the joint dynamics of the species sizes on a logarithmic scale and the environment can be approximated by an explicit dynamical system in the limit of large populations. This convergence result holds true under the assumption that no "bad events" occur during the population dynamics, corresponding to situations where several species become dominant or go extinct simultaneously. No such bad event occurs for Baire-almost all initial conditions of the population sizes and the environmental resources. I will apply this framework to study the long term co-existence of two specialist species consuming two resources, with a "joint" niche construction where each species constructs the niche of the other while depleting its resources. I will also present an example of immune escape in cancer, where the environmental variable is associated to the state of the immune system and the species are associated to three types of tumor cells.

This is joint work with Nicolas Champagnat, Cristóbal Quiñinao, Leonardo Videla and Nicolás Zalduendo-Vidal.

Fyodorov, Yan

King's College London

Session: Chaotic systems from physics

Talk: *Eigenvector Ergodization in coupled non-Hermitian Quantum Chaotic Systems: random matrix approach*

Consider two non-Hermitian quantum chaotic systems that are modelled by two independent $N\times N$ complex Ginibre matrices. Suppose the two systems interact with each other through a deterministic matrix $c{\bf 1}_N$, where$ {\bf 1}_N$ is $N\times N$ identity matrix

and c is the complex coupling parameter whose magnitude $|c|$ controls the interaction strength .

We characterize quantitatively how the eigenvectors of the whole system, initially localized in one of the individual subsystems for $|c|=0$, eventually spread over the full system with growing interaction strength. The resulting asymptotic formula describing such spread in the limit $N\to \infty$ is very explicit and provides full picture of gradual ergodization of eigenvectors as a function of the coupling parameter $|c|$ in the whole transition regime. The presentation will be based on a joint work-in-progress with Margherita Disertori.

Ganesh, Ayalvadi

University of Bristol

Session: Stochastic Networks

Talk: *Decentralised bipartite matching*

Modern data centres connect tens of thousands of processors. We model the scheduling of communications between these processors as a bipartite matching problem. The scale of the problem, and the very low latencies required, render traditional centralised algorithms for maximum matching unsuitable. We present a class of probabilistic matching algorithms that are decentralised and only require knowledge each node to know its 2-neighbourhood in the bipartite graph. We also present an analysis of the performance of this algorithm on a class of random graphs.

Ganguly, Ankan

Boston University

Session: Stochastic Networks

Talk: *Mean-Field Analysis of an Interacting Particle System on Dynamically Coevolving Graphs*

Interacting particle systems (IPS) model collections of stochastic processes called "particles" whose interactions are governed by an underlying graph. Such models have applications throughout the social and physical sciences. Because IPS are high-dimensional stochastic processes, their large-scale limits are of interest. When the interaction graph is dense, particles in the IPS are typically asymptotically independent. This is called "propagation of chaos." Under propagation of chaos, each particle converges to a mean-field process. Previous work on mean-field limits of IPS has primarily focused on models with exogenously defined graphs. We present a simple interacting particle system with a dynamic interaction graph subject to bidirectional feedback: particle evolution depends on the underlying graph, and the graph dynamics depend on particle states. Because the graph is endogenous to the model, the mean-field limit is no longer sufficient to completely describe the limiting behavior of the IPS. We introduce a sampling perspective that yields mean-field limits for the particles, graphon limits for the network, and other limits that capture the joint distribution of the particles and graph. We show that IPS has a rich asymptotic conditional structure, motivating a general conditional propagation of chaos result that extends a famous propagation of chaos theorem by Sznitman. We finish with some numerical illustrations of our results.

Gantert, Nina

Technical University of Munich

Session: Random graph models

Talk: *Some result about opinion dynamics*

We explain some results about opinion dynamics, more precisely about the averaging process on infinite graphs. Vertices have real-valued opinions and when they interact along edges, both vertices take the average value of the two endpoints of the edge.

No prerequisites required.

Based on joint work with Timo Vilkas.

Gerencsér, Máté

TU Wien

Session: Three perspectives on singular stochastic PDEs

Talk: *Variance renormalisation of parabolic stochastic PDEs*

Scaling arguments give a natural guess at the regularity condition on the noise in a stochastic PDE for a local solution theory to be possible, using the machinery of regularity structures or paracontrolled distributions. This guess of "subcriticality" is often, but not always, correct. In cases when it is not, a the blowup of the variance of certain nonlinear functionals of the noise necessitates a different, multiplicative renormalisation. We present some recent progress on this variance renormalisation, and discuss the strengths and weaknesses of different approaches such as regularity structures, Ito calculus, or stochastic sewing.

Gibbs, Isaac

University of Toronto

Session: Conformal prediction for non-exchangeable data

Talk: *Calibrated Multi-Level Quantile Forecasting*

In this talk, I will discuss an online method for calibrating quantile forecasts at multiple levels simultaneously. We say that a sequence of quantile forecasts is calibrated if its q-level predictions are greater than or equal to the target value at a q fraction of time steps, for each level q. I will introduce a lightweight procedure, multi-level quantile tracker (MultiQT), that can wrap around any baseline sequence of point or quantile predictions to produce adjusted quantile forecasts that are guaranteed to be calibrated, even against adversarial distribution shifts. Critically, this method ensures that the quantiles remain ordered, e.g., the 0.5-level quantile forecast will never be larger than the 0.6-level forecast. Moreover, the method has a no-regret guarantee, implying it will not degrade the performance of the existing forecaster (asymptotically) with respect to the quantile loss. In experiments, we find that MultiQT significantly improves the calibration of existing real-world forecasters in epidemic and energy forecasting problems, while leaving the quantile loss largely unchanged or slightly improved.

Gijbels, Irène

University of Leuven (KU Leuven)

Session: Recent advances in Dependence Modeling

Talk: *Copula-based estimation of functionals of aggregated risks*

The interest in this talk is in aggregated random variables, where the dependency between the components is modelled via copulas. We are in particular interested in a general class of functionals of such an aggregated random variable, and estimation of these functionals. The separation of dependence structure and margins allows for highly adaptable estimation procedures, accommodating parametric and nonparametric methods for both components. The generality of the framework enables its application to diverse contexts. The practical use of the developed methodology is demonstrated in real-data applications.

Gneiting, Tilmann

Heidelberg Institute for Theoretical Studies

Session: Wald Lecture 2

Talk: *Hierarchies of Calibration: Classification and Regression*

Concepts of calibration formalize the compatibility between probabilistic predictions and the respective outcomes. In a nutshell, the outcomes ought to be indistinguishable from random draws from the predictive distributions. The talk strives to review and extend notions of calibration that have been proposed for classification and regression tasks. Particular emphasis is given to hierarchical relations between the various notions, as they apply to general real-valued, continuous, nominal, and binary outcomes, respectively. Furthermore, we discuss concepts of calibration that are expressed in terms of properties or functionals of the predictive distribution, such as means, quantiles, or event probabilities. To illustrate the applied and methodological relevance of these notions, we revisit associated decompositions of proper scoring rules and consistent scoring functions into measures of miscalibration, discrimination, and uncertainty. While calibration checks apply to (out-of-sample) assessments of predictive performance, they relate closely to (in-sample) model diagnostics, and we elucidate these connections in classification and regression settings. Joint work with Johannes Resin and Lu Yang.

Gong, Ruobin

Rutgers University

Session: Statistical Inference and uncertainty quantification in the age of AI

Talk: *Simulation-Based Methods for Valid Inference from Privacy-Protected Data*

Data subject to disclosure control is becoming increasingly ubiquitous in the social sciences. While such protections are essential for safeguarding the privacy of individual data subjects, privacy-protected data poses substantial challenges for statistical inference. In particular, the mechanisms used to protect privacy often occlude the underlying confidential data in highly intractable ways. This talk discusses a set of approaches that leverage simulation-based methods to enable valid statistical inference in these settings. These methods are applicable to data swapping, an important technique in traditional statistical disclosure control, as well as to modern mechanisms that satisfy differential privacy. These methods sketch a flexible framework for principled inference when the true data generating process is obscured by probabilistic privacy protection.

Gong, Yan

MBZUAI

Session: Causality Meets AI: Generative Structure, LLM‚ÄëAssisted Trials, and Spatial Heterogeneity

Talk: *Causal Spatial Quantile Regression*

Treatment effects in a wide range of economic, environmental, and epidemiological applications often vary across space, and understanding the heterogeneity of causal effects across space and outcome quantiles is a critical challenge in spatial causal inference. To effectively capture spatial heterogeneity in distributional treatment effects, we propose a novel semiparametric neural network-based causal framework leveraging deep spatial quantile regression and then construct a plug-in estimator for spatial quantile treatment effects (SQTE). This framework incorporates an efficient adjustment procedure to mitigate the impact of spatial hidden confounders. Extensive simulations across various scenarios demonstrate that our methodology can accurately estimate SQTE, even with the presence of spatial hidden confounders. Additionally, the spatial confounding adjustment procedure effectively reduces neighborhood spatial patterns in the residuals. We apply this method to assess the spatially varying quantile treatment effects of maternal smoking on newborn birth weight in North Carolina, United States. Our findings consistently show negative effects across all birth weight quantiles, with particularly severe impacts observed in the lower quantile regions.

Gorodetsky, Ofir

Technion - Israel Institute of Technology

Session: Recent advances in probabilistic number theory

Talk: *The distribution of partial sums of the Steinhaus function*

The Steinhaus function is a random, completely multiplicative function on the positive integers, whose values on primes are iid random variables uniformly distributed on the complex unit circle. Its study is motivated by the study of classical (deterministic) multiplicative functions such as the Möbius function and Dirichlet characters.

We shall describe recent joint work with Mo Dick Wong, where the limiting distribution of the partial sums of the Steinhaus function was determined. The limiting distribution is Gaussian with random variance; the variance is given by the total mass of a random measure. This measure is an instance of critical multiplicative chaos.

In the talk we shall discuss the motivation and history behind the problem, and the intuition behind the result.

Grünwald, Peter

CWI and Leiden University

Session: E-Values: local, compound and conditional

Talk: *E-Values obtained by Conditioning on a Sufficient Statistic are Essentially Optimal*

Whenever a composite null hypothesis admits a sufficient statistic T, the conditional likelihood ratio, in which both the null and a representative of the alternative are conditioned on T provides an e-variable. We provide a precise analysis of such conditional e-variables for the case of exponential family nulls and alternatives. If T is also sufficient for the alternative, as happens in applications such as 2-sample testing, this conditional e-variable performs exceedingly well: it is asymptotically indistinguishable from the (much harder to compute!) growth (i.e. log)-optimal e-variable under any prior W1 on the alternative with compact support: for any such prior W1, the corresponding reverse information prior on the null W* converges to the same W1 (they induce the same densities on the mean-value parameters), an intriguing result in its own right. Further, the likelihood ratio of the Bayes marginal distributions with priors W1 and W* converges to the conditional e-variable. As such, the conditional e-variable asymptotically achieves *uniform growth optimality* - a novel concept that is stronger than both absolute and relative growth optimality. We also briefly highlight how to extend the definition and analysis if the alternative is of larger dimensionality than the null (this includes GLMs) and in more general cases in which the null is not an exponential family, yet simple sufficient statistics exist.

Joint work with Dante de Roos, Sebastian Arnold and Yunda Hao.

Gui, Yu

University of Pennsylvania

Session: IMS Lawrence D. Brown Ph.D. Student Award

Talk: *Distributionally robust risk evaluation with an isotonic constraint*

Statistical learning under distribution shift is challenging when neither prior knowledge nor fully accessible data from the target distribution is available. Distributionally robust learning (DRL) aims to control the worst-case statistical performance within an uncertainty set of candidate distributions, but how to properly specify the set remains challenging. To enable distributional robustness without being overly conservative, in this paper we propose a shape-constrained approach to DRL, which incorporates prior information about the way in which the unknown target distribution differs from its estimate. More specifically, we assume the unknown density ratio between the target distribution and its estimate is isotonic with respect to some partial order. At the population level, we provide a solution to the shape-constrained optimization problem that does not involve the isotonic constraint. At the sample level, we provide consistency results for an empirical estimator of the target in a range of different settings. Empirical studies on both synthetic and real data examples demonstrate the improved accuracy of the proposed shape-constrained approach.

GUILLIN, Arnaud

Université Clermont Auvergne

Session: Stochastic Analysis and Applications

Talk: *Mean field particles system : propagation of chaos and concentration*

We will review recent progress for sharp propagation of chaos for mean field particles system, in particular showing new results stated in terms of Fisher information. We will also prove concentration inequalities via autorefinement of Poincaré inequalities, in particular for mean field particles system.

In collaboration with Jules Grass, Christophe Poquet, Boris Nectoux and Liming Wu.

Guo, Zijian

Zhejiang University

Session: Modern Optimization for Statistical Learning

Talk: *Statistical Analysis of Conditional Group Distributionally Robust Optimization with Cross-Entropy Loss*

In multi-source learning with discrete labels, distributional heterogeneity across domains poses a central challenge to developing predictive models that transfer reliably to unseen domains. We study multi-source unsupervised domain adaptation, where labeled data are available from multiple source domains and only unlabeled data are observed from the target domain. To address potential distribution shifts, we propose a novel Conditional Group Distributionally Robust Optimization (CG-DRO) framework that learns a classifier by minimizing the worst-case cross-entropy loss over the convex combinations of the conditional outcome distributions from sources domains. We develop an efficient Mirror Prox algorithm for solving the minimax problem and employ a double machine learning procedure to estimate the risk function, ensuring that errors in nuisance estimation contribute only at higher-order rates.

We establish fast statistical convergence rates for the empirical CG-DRO estimator by constructing two surrogate minimax optimization problems that serve as theoretical bridges. A distinguishing challenge for CG-DRO is the emergence of nonstandard asymptotics: the empirical CG-DRO estimator may fail to converge to a standard limiting distribution due to boundary effects and system instability. To address this, we introduce a perturbation-based inference procedure that enables uniformly valid inference, including confidence interval construction and hypothe

Gurvich, Itai

Northwestern University

Session: Inference, Learning, and Decision-Making in Complex Systems

Talk: *Goggin's Corrected Kalman Filter: Guarantees and Filtering Regimes*

We revisit a nonlinear filter for linear state-space models with {\em non-Gaussian} signal and observation noise, introduced by Eimear Goggin in 1992. Goggin showed that applying the Kalman filter (KF) to score-transformed observations is asymptotically optimal in a specific signal-to-noise ratio (SNR) regime. We provide non-asymptotic convergence rates across a broad range of SNR regimes, recovering Goggin's setting as a special case, with bounds explicit in the SNR.

Our analysis combines two ingredients: a posterior Cramér-Rao lower bound for filtering and convergence-rate bounds in the Fisher information central limit theorem. We also map the parameters space into filtering regimes, identify degenerate regimes-where simple filters are nearly optimal-and isolate a balanced regime, where Goggin's filter has the most value.

Győrffy-Kerekes, Anna

ETHZ and Max Planck ETH Center for Learning Systems

Session: Recent Developments in Head-to-Head Comparisons and Stochastic Preference Modeling

Talk: *Non-transitivity in Pairwise Comparisons*

Head-to-head comparison metrics, such as win proportions (WP), win ratios (WR), and hazard ratios (HR), are fundamental for evaluating efficacy in randomized controlled trials (RCTs). Despite their widespread use, certain structural vulnerabilities remain underappreciated. While prior research has identified "Oakes-type intransitivity"—where specific realizations of event times create loops of "wins" among participant triplets—other forms of intransitivity require further investigation.

We examine "Efron-type non-transitivity," a distinct phenomenon where long-run comparisons violate logical transitivity (e.g., treatment A outperforms B, B outperforms C, yet C outperforms A). Unlike Oakes-type intransitivity, Efron-type non-transitivity is intrinsic to the underlying random variables and cannot be resolved by increasing sample sizes or extending observation periods. Because mathematical transitivity is only guaranteed for univariate statistics (as shown by Lumley and Guillain), pairwise head-to-head comparisons are inherently susceptible to this paradox, thereby challenging their universal reliability as definitive measures of efficacy.

Our work explores the scope of Efron-type non-transitivity in clinical metrics. Specifically, we aim to identify the exact statistical conditions necessary to preclude non-transitivity for WR, WP, and HR, thereby clarifying their appropriate utility in RCTs.

Han, Fang

University of Washington, Seattle

Session: Measures of Statistical Association

Talk: *Conditional independence testing via Azadkia-Chatterjee's graph correlation*

Since Azadkia and Chatterjee introduced their nearest-neighbor-based graph correlation as a measure of conditional dependence, an important open question has been how to use it for statistically valid testing of conditional independence. In this talk, I will present, for the first time, a complete roadmap for accomplishing this. The key ingredients are (1) joint asymptotic normality of the graph correlation under arbitrary dependence, (2) an exact characterization of the limiting variance, and (3) a careful bias analysis and correction building on Azadkia, Chen, and Han (2025).

Han, Qiyang

Rutgers University

Session: Recent Advances in Iterative Methods for Random Optimization

Talk: *Algorithmic inference via nonconvex gradient descent*

Conventional statistical inference methods are typically developed for models simple enough to admit tractable estimators through carefully designed iterative algorithms. In contrast, modern learning architectures are enormously complex, yet are trained by simple gradient-descent-type algorithms, often without any provable guarantee of algorithmic convergence to global/local optima.

Can we reconcile classical inference principles with these highly complicated, modern learning paradigms? In this talk, we will present a new inference framework addressing this question, by showing that valid statistical inference can be performed along the entire gradient descent trajectory, iteration by iteration, without requiring convexity of the loss landscape or convergence of the algorithm.

To illustrate this concept, we begin with a single-index regression model and demonstrate how gradient descent iterates can be "debiased", at each iteration, to yield valid confidence intervals for the underlying signal and consistent estimates of generalization errors. We then extend this paradigm to the much more challenging setting of learning with general multi-layer neural networks in their full complexity, where the loss landscape can be arbitrarily complex. Crucially, the proposed method remains valid without requiring either algorithmic convergence or oracle knowledge of the unknowns, and may therefore inform practical decisions such as early stopping and hyperparameter tuning.

Hannig, Jan

University of North Carolina at Chapel Hill

Session: From Simulators to Generative Models: Uncertainty Quantification and Model Interpretability

Talk: *Generative Fiducial Models*

While generalized fiducial inference (GFI) and its variants have yielded many theoretical and practical results to parametric inference and uncertainty quantification, applying it to generative models remains challenging. We identify three key issues misspecification, metric choices, and over-parameterization hinder the direct application of the GFI to generative models. In this paper, we propose a novel method based on the framework of generalized fiducial inference, designed to construct distributional estimates over the parameter space given observed data, while also enabling uncertainty quantification for generative models. We employ a truncation-based approach and further provide a theoretical analysis of its behavior under varying truncation parameters. Both theoretical results and empirical evidence suggest that, with an appropriately chosen truncation parameter, the truncated distribution derived from generalized fiducial inference achieves valid coverage of the true parameter and leads to improved generalization performance.

Joint Work with Zijie Tian, T. C. M. Lee (UC Davis)

Harel, Matan

Northeastern University

Session: Statistical mechanics

Talk: *Planar percolation and the loop O(n) model*

Consider a tail trivial, positively associated site percolation process such that the set of open vertices is stochastically dominated by the set of closed ones. We show that, for any planar graph G, such a process must contain zero or infinitely many infinite connected components. The assumptions cover Bernoulli site percolation at parameter p less than or equal to one half, resolving a conjecture of Benjamini and Schramm. As a corollary, we prove that p_c is greater than or equal to 1/2 for any unimodular, invariantly amenable planar graphs. We will then apply this percolation statement to the loop O(n) model on the hexagonal lattice, and show that, whenever n is between 1 and 2 and x is between 1/sqrt(2) and 1, the model exhibits infinitely many loops surrounding every face of the lattice, giving strong evidence for conformally invariant behavior in the scaling limit (as conjectured by Nienhuis).

This is joint work with Alexander Glazman (University of Innsbruck) and Nathan Zelesko (Northeastern University).

Harrar, Solomon

University of Kentucky

Session: Statistical Inference and uncertainty quantification in the age of AI

Talk: *Analysis of Multiple Outcomes in Contaminated Trials Reinforced With Validation Data*

This paper is concerned with estimation and testing for treatment effects with multivariate outcomes. It primarily focuses on the situation where imperfect diagnostic tools are used to classify subjects into different groups. Oftentimes, there are more expensive and/or invasive diagnostic tools to accurately determine the subjects' status or conditions, yielding partially validated data on a smaller number of subjects. We propose moment-based approaches for estimating and testing treatment effects. We compare our methods with maximum likelihood approach using the EM algorithm, which requires strong assumptions and bears computational burden, and with traditional methods, which ignore the diagnostic tool's imperfection. The proposed methods show advantages in terms of coverage probability, computations efficiency, and robustness. The application of the methods is illustrated with gene-expression data from the Genes-environments & Admixture in Latino Americans (GALA) II study of asthma in Hispanic/Latino children.

Heydenreich, Markus

University of Augsburg

Session: Statistical mechanics

Talk: *Critical behaviour for weight-dependent random connection models*

The random connection model is a versatile model encompassing various variants of continuum percolation: vertices are given as point process in a suitable metric space, and edges are drawn independently with probability depending on the distance of the endpoints and possible further vertex parameters known as weights. We introduce the model and present results for the critical percolation phase transition in high-dimensional Eudlidean space and in hypberbolic space.

Hoff, Peter

Duke University

Session: Modern Approaches to Inference and Estimation

Talk: *Universal prediction with extended ranks*

The accuracy of predictions derived from a regression model depends on the extent to which the model is able to mimic certain aspects of the conditional distribution of an outcome given its features. Primary among these are the distributional support of the outcome variable, and the relationship between the mean and variance of the conditional distribution. In this talk we develop a single inferential framework that can represent these aspects for a wide range of conditional distributions. Our approach is based on an unknown but arbitrary non-decreasing transformation of a linear regression model, using a pseudo-likelihood based on a partial ordering of the observed outcomes. This transformation model can accommodate any ordinal data type, including continuous and discrete ordered data, and can represent a wide range of mean-variance relationships. Theoretical results show that the pseudo-likelihood is asymptotically efficient relative to an appropriate full likelihood at the two extremes of ordinal data – continuous and binary data. Parameter estimates and predictive inference are available via a simple Gibbs sampling algorithm and conformal calibration.

Hörmann, Siegfried

Graz University of Technology

Session: Measures of Statistical Association

Talk: *Measuring dependence between a categorical response and a functional covariate*

We suggest a dependence coefficient between a categorical variable and some general variable taking values in a metric space. We derive important theoretical properties and study the large sample behaviour of our suggested estimator. Moreover, we develop an independence test which has an asymptotic χ^2-distribution if the variables are independent and prove that this test is consistent against any violation of independence. The test is also applicable to the classical~K-sample problem with possibly high- or infinite-dimensional distributions. We discuss some extensions, including a variant of the coefficient for measuring conditional dependence.

Huang, Jian

The Hong Kong Polytechnic University

Session: Statistical inference for data sceince and AI

Talk: *Continuous Normalizing Flows for Statistical Generative Learning*

Continuous Normalizing Flows (CNFs) have emerged as a powerful class of generative models, distinguished by their capacity for their high-fidelity sample generation. By leveraging neural ordinary differential equations, CNFs construct a continuous-time, invertible mapping that smoothly transforms a tractable base distribution, such as standard Gaussian, into a complex, high-dimensional target data distribution. In this talk, we will explore the theoretical foundations underpinning CNFs. Building on these principles, we will examine the versatility of this framework across a broad spectrum of modern statistical and machine learning tasks. Specifically, we will highlight how the exact invertibility and tractable density evaluation inherent to CNFs can be uniquely leveraged to characterize conditional independence, advance counterfactual estimation in causal inference, provide rigorous uncertainty quantification in conformal prediction, and enable dynamic trajectory modeling in motion generation.

Ignatiadis, Nikolaos

University of Chicago

Session: Empirical Bayes for approximate oracle inference

Talk: *How does limma-trend work? An empirical partially Bayes perspective*

In high-throughput biology, it is common to fit thousands of linear regressions---one per gene, protein, or other unit---with very few samples per unit. Limma-trend, one of the most widely used methods in this setting, improves power by shrinking variance estimates toward an estimated mean-variance trend before computing p-values and applying the Benjamini-Hochberg procedure to control the false discovery rate (FDR). In this work, we view limma-trend through the lens of empirical partially Bayes inference, a paradigm in which a prior is posited and estimated for the nuisance parameters while parameters of interest remain fixed. From this perspective, we derive formal statistical justification for a nonparametric generalization of limma-trend. We show that this approach asymptotically controls the FDR even when the mean-variance trend is misspecified or inconsistently estimated. We also develop a new empirical partially Bayes procedure that gains power by learning the full conditional distribution of variances rather than just a trend, and illustrate our results on RNA-seq, proteomics, and ChIP-seq data.

Janz, David

University of Oxford

Session: Interactive decision making

Talk: *Maximum information gain in kernel ridge regression*

We establish matching lower and upper bounds for the maximum information gain (log-determinant of the regularised Gram matrix) for kernel ridge regression with covariates in the unit hypercube for families of kernels including the Matérn and squared exponential. Our results are based on elementary methods and avoid the strong and unverified assumption of uniformly bounded Mercer eigenfunctions used in earlier work.

Jarai, Antal A.

University of Bath

Session: Random spatial systems

Talk: *On the number of particles in Activated Random Walk and the Abelian sandpile*.

We consider the number of particles in two different models of self-organised criticality: the Activated Random Walk (ARW) and the Abelian sandpile model (ASM). For the driven-dissipative ARW on the complete graph with N vertices and one sink and sleep rate lambda, we show that the stationary number of particles is

rho(lambda) N + a(lambda) sqrt(N log N) (1 + o(1)),

for simple explicit functions rho and a.

For the driven-dissipative ASM on a lattice approximation of a 2D domain D, with lattice scaled by epsilon, we show that the stationary probability that a site at z in D is vacant scales as

p(0) + c f(z) epsilon^2 + o(epsilon^2),

where p(0) is the vacancy probability on the infinite lattice, c>0 is a lattice dependent constant, and f(z)>0 is a conformally covariant function on D of scale dimension 2. We also show that the symmetry assumptions we impose on the lattice cannot be dropped, that is, without them the formula does not necessarily hold.

(Joint works with C. Moench & L. Taggi, and M.W. Elvidge, respectively)

Jiang, Binyan

The Hong Kong Polytechnic University

Session: Advances in Matrix Methods

Talk: *Dynamic Networks with Node Heterogeneity and Homophily*

The goal of this work is to model node heterogeneity and link homophily for dynamic networks. The proposed framework brings new insights on how networks evolve over time. It also provides more sophisticated tools for the prediction of future networks with statistical guarantees. The new model accounts for the link homophily associated with both observed traits and latent traits. The joint modeling of node heterogeneity and both observed and latent homophily effects also poses the significant challenge in statistical inference, resulted from the large number of confounding parameters in the model. To overcome this, we propose a novel normalized squared loss, paving the way for estimating parameters stably in a high-dimensional setting. We provide a rigorous theoretical analysis of the estimation method, and demonstrate its effectiveness through extensive simulations and the illustration with some real-world network data.

Kaakai, Sarah

LAGA, Université Sorbonne Paris Nord

Session: Interdisciplinary perspectives at the interface between probability theory and cell biology

Talk: *Nonparametric hazard rate estimation with associated kernels: theory and application to a two-phase aging model*

In this talk, I will present a general theoretical framework for the nonparametric estimation of the intensity of random events, using so-called associated kernels, whose shape adapts locally to the estimation point. One of the main advantages of these kernels is that they correct the well-known boundary bias of classical kernel estimators. I will present several results on the convergence of the associated kernel hazard rate estimator, as well as theoretical guarantees on a minimax bandwidth selection method in the spirit of Goldenshluger and Lepski (2011).

I will then apply these results to a two-phase aging model, using experimental data on Drosophila, obtained via the so-called "Smurf" phenotype. This phenotype, which measures intestinal permeability, is an important marker of aging in several species. I will show how the estimation of transition and mortality rates using associated kernels sheds new light on the biological mechanisms underlying the aging process, and lay the groundwork for the study of aging in natural populations.

Joint work with L. Breuil, M. Doumic and M. Rera.

Kepplinger, David

George Mason University

Session: Probability Foundations for Robust and Private Learning

Talk: *Robust and Efficient Domain Adaptation via ϕ-Divergences*

Machine learning models often perform unreliably when the distribution of the training (source) data differs from the test (target) data. Domain adaptation (DA) addresses this challenge by augmenting the training process to transfer knowledge across mismatched domains. However, most conventional DA approaches rely on strict, unverifiable assumptions—such as label or covariate shift—and fail to account for distributional irregularities like outliers and model misspecification, which can severely impact performance.

We propose a principled and robust DA framework based on Distributionally Robust Optimization (DRO) using ϕ-divergences. Our framework bypasses restrictive assumptions and naturally handles misspecification and outliers in both the source and target domains. We establish strong theoretical guarantees on the learned model's performance, deriving generalization bounds under mild conditions. We further show how varying levels of information from the target domain (e.g., no data, or labeled/unlabeled data) seamlessly integrate into our framework and theory. Finally, we demonstrate the advantages of our method through numerical experiments and real-world applications.

Kerriou, Céline

University of Cologne

Session: Random graph models

Talk: *Temporal connectivity of random geometric graphs*

A temporal random geometric graph is a random geometric graph in which all edges are endowed with a uniformly random time-stamp, representing the time of interaction between vertices. In such graphs, paths with increasing time stamps indicate the propagation of information. We determine a threshold for the existence of monotone increasing paths between all pairs of vertices in temporal random geometric graphs. The results reveal that temporal connectivity appears at a significantly larger edge density than simple connectivity of the underlying random geometric graph. This is in contrast with Erdős-Rényi random graphs in which the thresholds for temporal connectivity and simple connectivity are of the same order of magnitude. This talk is based on joint work with Anna Brandenberger, Serte Donderwinkel, Gábor Lugosi and Rivka Mitchell.

Khamaru, Koulik

Assistant Professor

Session: Bridging Statistics and Modern AI: Foundations for Deep Learning and Generative Models

Talk: *Avoiding the Price of Adaptivity: Inference in Linear Contextual Bandits via Stability*

Statistical inference under adaptive data collection is challenging due to the non-i.i.d. nature of the data. In particular, classical least-squares inference can fail under contextual bandit sampling, and constructing valid confidence intervals for linear functionals typically requires an unavoidable inflation of order square-root of d log T---where d is the dimension and T is the number of samples---the so-called price of adaptivity. The Lai--Wei stability condition, which requires the empirical feature covariance to concentrate around a deterministic limit, is known to circumvent this limitation. When it holds, the OLS estimator satisfies a CLT and standard Wald-type confidence intervals remain asymptotically valid without the need to inflate the confidence intervals.

In this talk, we will discuss a regularized EXP4 algorithm for linear contextual bandits that provably satisfies the Lai--Wei stability condition, thereby enabling valid Wald-type inference with quantitative CLT rates. We will also show that the same algorithm achieves minimax-optimal regret up to logarithmic factors, demonstrating that stability and statistical efficiency can coexist. As an application, we will discuss confidence intervals for the conditional average treatment effect (CATE) under adaptively collected data, and present simulations illustrating the empirical normality of the resulting estimators.

Kim, Sehwan

Ewha Womans University

Session: From Simulators to Generative Models: Uncertainty Quantification and Model Interpretability

Talk: *Extended Fiducial Inference for Individual Treatment Effects via Deep Neural Networks*

Fiducial inference was long regarded as a failure, yet Fisher's original goal—quantifying parameter uncertainty directly from observations—remains fundamental in statistics. We develop extended fiducial inference (EFI), a scalable framework that revisits this goal using modern statistical computing. EFI jointly imputes latent random errors via stochastic gradient Markov chain Monte Carlo and estimates the inverse map from observations to parameters using a sparse deep neural network (DNN). The consistency of the sparse DNN ensures proper propagation of observational uncertainty to parameter uncertainty, yielding downstream inference. Compared with frequentist and Bayesian approaches, EFI is more robust in parameter estimation, especially under outliers, and avoids reference distributions in hypothesis testing, moving toward automated inference. EFI also provides a framework for semi-supervised learning. As an application, we propose a Double Neural Network (Double-NN) method for individual treatment effect estimation within EFI. Separate DNNs model treatment and control response functions, while an additional network estimates their parameters. The method is broadly applicable by universal approximation theory and empirically outperforms conformal quantile regression. Theoretically, it enables uncertainty quantification in models whose complexity grows with sample size, and provides a rigorous framework for uncertainty quantification of DNNs under neural scaling laws.

Klippel, Andreas

TU Darmstadt

Session: Random walks and interacting particles on random graphs

Talk: *Long-Range Order in the Monomer Double-Dimer Model with Long-Range Interactions*

The dimer model and its associated double-dimer model are fundamental objects in probability theory, statistical mechanics, and combinatorics. While their behavior in planar settings is by now well understood, much less is known in higher dimensions and in the presence of a positive density of monomers, leading to the so-called monomer double-dimer model.

We study these models on Zd-like graphs (d≥1) that allow long-range edges whose weights decay with distance. For a large class of such interactions, we prove that the monomer double-dimer model exhibits long-range order. As a consequence, monomer correlations in the dimer model remain uniformly positive, and loops in the double-dimer model become macroscopic.

In this talk, I will introduce the models and outline the main ideas of the proof. We will see that the model admits a natural correspondence with a spin system, which allows us to transfer results obtained via reflection positivity and thereby establish long-range order.

This is joint work with Lorenzo Taggi and Wei Wu.

Knowles, Antti

University of Geneva

Session: Chaotic systems from physics

Talk: *Random graphs as models of quantum disorder*

Abstract: A disordered quantum system is mathematically described by a large Hermitian random matrix. One of the most remarkable phenomena expected to occur in such systems is a localization-delocalization transition for the eigenvectors. Originally proposed in the 1950s to model conduction in semiconductors with random impurities, the phenomenon is now recognized as a general feature of wave transport in disordered media, and is one of the most influential ideas in modern condensed matter physics. A simple and natural model of such a system is given by the adjacency matrix of a random graph. I report on recent progress in analysing the phase diagram for the Erdös-Renyi model of random graphs. In particular, I explain the emergence of fully localized and fully delocalized phases, which are separated by a mobility edge. I also explain how to obtain optimal delocalization bounds using a new Bernoulli flow method. Based on joint work with Johannes Alt, Raphael Ducatez, and Joscha Henheik.

Kojadinovic, Ivan

University of Pau

Session: Recent advances in Dependence Modeling

Talk: *An algorithmic procedure for solving the generalized minimum information checkerboard copula problem*

The minimum information copula principle (see Meeuwissen and Bedford, 1997) is a maximum entropy-like approach for finding the least informative copula that satisfies a certain number of expectation constraints specified either from expert knowledge or the available limited data. In this presentation, we first propose a generalization of this principle allowing the inclusion of additional constraints fixing certain higher-order margins of the copula. We next prove that the associated optimization problem has a unique solution under a natural condition. As the latter problem is intractable in general, following the existing literature, we consider its version with all the probability measures involved in its formulation replaced by checkerboard approximations. This amounts to attempting to solve a so-called discrete I-projection linear problem. We then use the seminal results of Csiszar (1975) to derive an IPFP-like procedure for solving the latter and provide theoretical guarantees for its convergence. We conclude the presentation with numerical experiments in dimensions up to four with substantially finer discretizations than those encountered in the literature. This is joint work with Tommaso Martini, University of Torino, Italy.

Kolaczyk, Eric

McGill University

Session: Statistical Learning with Noisy Labels, Privacy Perturbations, and Latent Structures

Talk: *Privacy-Utility Tradeoffs of Noisy Stochastic Gradient Descent in High Dimensions*

The interplay between optimization and privacy has become a central theme in privacy-preserving machine learning. Noisy stochastic gradient descent (SGD) has emerged as a cornerstone algorithm, particularly in large-scale settings. These variants of gradient methods inject carefully calibrated noise into each update to achieve differential privacy, which is the gold standard for rigorous privacy guarantees. Prior works primarily provide various bounds on statistical risks (utility) and privacy loss for noisy SGD, yet the exact behavior of the process remain unclear, particularly in high-dimensional settings. This work leverages a diffusion approach to analyze noisy SGD precisely, providing a continuous-time perspective that captures both statistical risk evolution and privacy loss dynamics in high dimensions. Moreover, we study a variant of noisy SGD that does not require explicit knowledge of gradient sensitivity, unlike existing approaches that assume or enforce sensitivity through gradient clipping. Specifically, we focus on the least squares problem with l2 regularization. Joint work with Shurong Lin and Elliot Paquette.

Kolesnik, Brett

University of Warwick

Session: Random walks and interacting particles on random graphs

Talk: *Recent progress on graph bootstrap percolation*

In H-percolation, edges in the random graph G(n,p) are initially active. Other edges of the complete graph K_n become active if they are the only inactive edge in some copy of H in K_n. If all edges of K_n become active, the process H-percolates. In this talk, we will identify the critical threshold p_c(n,H) when H is the clique K_r, and we will discuss recent progress for general H. These results resolve open problems of Balogh, Bollobás, and Morris (RS&A 2012). The talk is based on joint work with Bartha, Kronenberg, and Peled, and with Makai, Nenadov, Pérez-Giménez, Prałat, and Zhukovskii.

Kolupaiev, Oleksii

Institute of Science and Technology Austria

Session: Chaotic systems from physics

Talk: *The eigenvalues of iid matrices are hyperuniform*

We consider the point process of the eigenvalues of real or complex non-Hermitian matrices X with independent, identically distributed entries. In this talk, I will present a recent result obtained in a joint work with L. Erdos and G. Cipolloni, which establishes that this point process is hyperuniform: the variance of the number of eigenvalues in a subdomain of the spectrum is much smaller than the volume of this subdomain. Our main technical novelty is a very precise computation of the covariance between the resolvents of the Hermitization of $X-z_1$, $X-z_2$, for two distinct complex parameters z_1,z_2.

Kunisky, Dmitriy

Johns Hopkins University

Session: Random matrices in machine learning

Talk: *Universality of first-order methods on random and deterministic matrices*

I will present recent results on the universal behavior of general first-order methods (GFOM), a flexible class of iterative algorithms which update a state vector by matrix-vector multiplications and entrywise nonlinearities, thereby generalizing the much-studied class of approximate message passing (AMP) algorithms. Our approach will study the dynamics of GFOM through a moment method, reducing them to the traffic distribution of a sequence of matrices, an object studied in connection with free probability that consists of the collection of limiting values of certain diagrammatic permutation-invariant polynomials in the matrix entries. I will show how this approach gives a simple diagrammatic explanation for the asymptotic Gaussianity of AMP that distinguishes these algorithms from other GFOM, and how it unifies known limit theorems for AMP and GFOM on Wigner and rotationally invariant matrices and extends them to matrices with less randomness, more complicated variance profiles, and even some fully deterministic settings, in the latter case making progress towards resolving longstanding conjectures in the statistical physics and AMP literature.

Based on joint work with Nicola Gorini, Chris Jones, and Lucas Pesenti.

Kur, Gil

ETH Zurich

Session: The many facets of score matching

Talk: *Minimum Norm Interpolation Under Gaussian Covariates*

Minimum-Norm Interpolators (MNI) in overparameterized linear models have gained attention as a tractable framework for studying interpolation phenomena that resemble empirical observations in neural networks. Most prior work on these interpolators either exploits closed-form solutions when available or relies heavily on Gaussian comparison results, such as the convex Gaussian Min-Max Theorem (CGMT). In this paper, we introduce a new perspective on MNI under Gaussian covariates by leveraging tools from high-dimensional geometry. First, we obtain a ``localized'' bound on how much the MNI shrinks the original ground truth under isotropic Gaussian covariates when the norm is in isotropic position. Next, we prove sharp rates for the Mean Squared Error (MSE) of the ℓ_1-MNI, as obtained by Wang et al. via a purely geometric proof; crucially, our approach does not rely on CGMT.

Kurisu, Daisuke

The University of Tokyo

Session: Statistical learning for spatial-temporal data

Talk: *M-Estimation for Non-stationary Spatial Data*

While non-stationary spatial data are increasingly prevalent in various scientific applications, developing robust estimation and inference methods remains theoretically challenging due to complex spatial dependencies and potentially non-smooth loss functions. In this talk, we investigate the local linear quantile estimation for non-stationary spatial random fields. As our primary theoretical contribution, we derive the spatial Bahadur representation for the proposed estimator. This foundational result provides a unified asymptotic framework that bypasses traditional restrictive conditions. Furthermore, based on this representation, we explore its applications in facilitating further statistical inferences, such as the construction of Simultaneous Confidence Bands (SCBs) and spatial hypothesis testing.

Lahiri, Soumendra

Washington University in St. Louis

Session: Statistical learning for spatial-temporal data

Talk: *A Central Limit Theorem for sums of spatial block variables and some applications*

Many statistical inference methods for spatial data are based on sums of functions of blocks of spatial variables instead of sums of functions of individual observations. Common examples include Spatial Block Bootstrap, Spatial Block Empirical Likelihood, Log-likelihood functions under different Markov Random Field models, among others. Asymptotic distribution theory for these methods is complicated by nontrivial overlaps that lead to strong dependence among the summands and is often derived on a case by case basis. This work proves a general Central Limit Theorem for sums of block variables using a version of Stein's method. allowing both pure- and mixed-increasing domain spatial asymptotic structures. Applications to some important inference problems in Spatial Statistics are also given.

Langthaler, Patrick

Paris-Lodron University Salzburg

Session: Recent Developments in Head-to-Head Comparisons and Stochastic Preference Modeling

Talk: *A nonparametric test for survival curves combining the Mann-Whitney effect and an overlap index*.

In survival analysis several methods are available for testing the identity of two survival functions. A prevalent model is that of proportional hazards, which is violated in many cases occuring in practice, such as the treatment arm conferring late or early benefits compared to the placebo arm. Here we provide a method for comparing two survival functions in a right-censoring framework, which is particularly useful for crossing survival curves, while still retaining good power in a proportional hazards setting. The method combines two stochastic functionals: the well known Mann-Whitney effect, and an overlap index which considers distributional overlap. We construct our test by adapting a joint test methodology for the two functionals to the case of right-censored data. We present essential theoretical results regarding our method as well as results from a simulation study and an application to a real dataset.

Lee, Ann

Carnegie Mellon University

Session: From Simulators to Generative Models: Uncertainty Quantification and Model Interpretability

Talk: *Trustworthy scientific inference with neural density estimators in intractable likelihood settings*

Scientific inference often involves inferring internal key parameters that determine the outcome of a complex physical phenomenon. The data themselves may come in the form of a labeled set (from e.g. a simulator or cross-matched studies) that implicitly encodes an often intractable likelihood function. We refer to inference in such a setting as "Likelihood-Free Inference'" (LFI). The application of neural density estimators and generative models to scientific LFI settings is becoming increasingly widespread. However, high-posterior density (HPD) regions derived from these density estimators do not necessarily have a high probability of including the true parameter of interest, even if the posterior is well-estimated and the labeled data have the same distribution as the target distribution. Furthermore, if the prior distribution is poorly specified, then the HPD regions could severely undercover and/or be biased, thereby leading to misleading scientific conclusions. In this talk, I will present new LFI methodology and algorithms for leveraging neural density estimators to produce confidence regions that have (i) nominal frequentist coverage for any value of the (unknown) parameter, even with just one observation (sample size n=1), and (ii) smaller average area (yielding higher constraining power) if the prior is well-specified. I will illustrate our methods on examples from astronomy and high-energy physics, and discuss where we stand and what challenges still remain.

Leung, Sun Kai

University of Oxford

Session: Recent advances in probabilistic number theory

Talk: *Sato--Tate random multiplicative function*

Random multiplicative functions have become a central theme in modern probabilistic number theory. In this talk, I will introduce the Sato--Tate random multiplicative function and outline its connections to arithmetic quantum chaos, with emphasis on the statistical behaviour of Hecke eigenforms. I will conclude with a discussion of ongoing joint work with Jad Hamdan and Mo Dick Wong.

Levin, Keith

University of Wisconsin--Madison

Session: Modeling complex network-linked data

Talk: *Testing for correlation between network structure and high-dimensional node covariates*

In many application domains, networks are observed with node-level features. In such settings, a common problem is to assess whether or not nodal covariates are correlated with the network structure itself. Here, we present four novel methods for addressing this problem. Two of these are based on a linear model relating node-level covariates to latent node-level variables that drive network structure. The other two are based on applying canonical correlation analysis to the node features and network structure, avoiding the linear modeling assumptions. We provide theoretical guarantees for all four methods when the observed network is generated according to a low-rank latent space model endowed with node-level covariates, which we allow to be high-dimensional. Our methods are computationally cheaper and require fewer modeling assumptions than previous approaches to network dependency testing. We demonstrate and compare the performance of our novel methods on both simulated and real-world data.

Levina, Liza

University of Michigan

Session: Modern Statistical Approaches to Network Data

Talk: *Comparing groups of networks*

Network-structured data arise in many applications; joint analysis of multiple networks is especially relevant to applications like neuroimaging, where each network corresponds to a patient's brain connectome. Models for multiple networks have typically focused on estimating their shared structure. Two-sample tests have also been developed, testing some version of the hypothesis of two samples of networks being indistinguishable across all node pairs (the global null). However, scientifically relevant hypotheses rarely take this form: for example, in neuroimaging it is rarely of interest to compare whole brains of patients and healthy controls, and the focus is more often on a particular brain region. Beyond comparisons, it is also of interest to estimate structures that correspond to a specific subgroup of networks, for example, for patients with a certain trait. One could always do that using just the subgroup of interest, but using all available samples allows us to better estimate structures that are shared by all, which in turn helps separate out the structure associated with a trait. This talk will introduce two methods that help address these challenges: mesoscale testing on networks, which conducts formal hypothesis testing on a subset of edges (like a brain region) while leveraging the rest of the network to increase power; and group MultiNeSS, a method that takes a sample of networks and estimates structures

Li, Hongzhe

University of Pennsylvania

Session: Recent advances in functional data analysis and metric statistics

Talk: *Test of partial effects for Fr\'echet regression on Bures-Wasserstein manifolds*

We propose a novel test for assessing partial effects in Fr\` echet regression on the Bures–Wasserstein manifold. Under the null hypothesis, we show that the proposed test statistic admits a degenerate V-statistic approximation, whose limiting distribution is a weighted mixture of chi-squared random variables, with weights given by the eigenvalues of an integral operator associated with a kernel. We establish the asymptotic validity and consistency of the test and evaluate its finite-sample performance through simulation studies. Finally, we apply the proposed method to investigate the effect of age, adjusting for other covariates, on gene co-expression network in single-cell data.

Lin, Xihong

Department of Biostatistics and Department of Statistics, Harvard University

Session: Statistical Foundations of Diffusion Models

Talk: *Harnessing Synthetic Data from Generative AI for Statistical Inference*

Integration of statistics and generative AI plays a pivotal role for accelerating trustworthy cross-domain scientific discovery. Recent advances in generative models have dramatically increased the availability and use of synthetic data across scientific domains. While these developments create exciting opportunities for empowering data analysis, they also raise fundamental statistical challenges regarding how synthetic data can be used in a valid, reliable, and principled manner. In this talk, we first discuss the current landscape of synthetic data generation using generative AI models such as transformer- and diffusion-based models. More importantly, we present a principled framework for incorporating synthetic data in downstream statistical analysis that ensures valid statistical inference even when generative AI models are misspecified. We show that the proposed synthetic data assisted methods integrating observed and synthetic data are robust to misspecified black-box generative models and can improve statistical inferential power when the generative AI models are informative. We demonstrate the utility of these synthetic data assisted methods to the analysis of the UK biobank data, by performing genome-wide association studies (GWAS) of proteomic data and whole-genome sequencing (WGS) analyses of brain imaging phenotypes, both characterized by substantial missingness (about 90%).

Liu, Sifan

Duke University

Session: Post-selection inference

Talk: *Flexible Selective Inference with Flow-based Transport Maps*

Data-carving methods conduct selective inference by conditioning on the observed selection event and basing inference on the distribution of the full data given that event. A key limitation of existing approaches is their reliance on an analytically tractable characterization of the selection mechanism, which excludes many modern adaptive procedures, including data-driven hyperparameter tuning and complex model selection. We propose a new framework that enables selective inference without requiring an exact characterization of the selection event or the algorithm that produces it. The core idea is to learn a transport map that pushes forward a simple reference distribution to the conditional distribution given selection. The map is parameterized using a normalizing flow and trained directly from simulated data, leveraging the flexibility of flow-based generative modeling to approximate otherwise intractable distributions. Through extensive numerical experiments on both simulated and real data, we demonstrate that this method enables selective inference by providing: (i) valid p-values and confidence intervals for adaptively selected hypotheses and parameters, (ii) a closed-form expression for the conditional density, enabling likelihood-based and quantile-based inference, and (iii) a modular adjustment for intractable selection steps that can be easily combined with existing methods for the tractable steps in a multi-step selection procedure.

Lodewijks, Bas

University of Sheffield

Session: Random graph models

Talk: *Preferential attachment trees with vertex death: The super-linear regime*

We investigate an evolving random tree model in which, at every step, either a new alive vertex is added which connects to an existing alive vertex preferentially according to a function b, or an existing alive vertex is selected preferentially according to a function d and killed. We consider the super-linear regime, in which 1/(b(i)+d(i)) is summable in i. Here, we investigate structural properties of the limiting infinite tree, in particular whether the limiting tree contains (i) A unique vertex with infinite degree and no infinite path, (ii) A unique infinite path and no vertex with infinite degree, (iii) Infinitely many infinite paths and no vertex with infinite degree. We shall discuss how and why each case (i)-(iii) arises, and will look at a broad range of examples.

Joint work with Markus Heydenreich.

Loecherbach, Eva

Ecole Polytechnique

Session: Recent advances in statistics for stochastic processes

Talk: *Interacting point processes : How does the mean field limit help us for statistical inference on them?*

We consider mean field interacting Hawkes processes composed of N components. We suppose that we observe the whole system during some fixed time interval.

Asymptotics will be taken as N tends to infinity (and sometimes also as time grows). Such models are widely used, for example in neuroscience to describe neuron's spiking behavior, an application that I will focus on. Three statistical questions are particularly important for neuroscientists in this context: Firstly, by observing only the jumps of the system, is it possible to estimate the underlying interaction graph or at least some important features of it? Secondly, is it possible to detect synchrony in the spiking behavior of large assemblies of neurons? Thirdly, is it possible to understand how a neuron reacts to incoming stimuli; that is, to estimate the underlying jump rate function?

I will present some attempts of answers that have been given to these points in some recent works in which I had the chance to collaborate. In particular I will also discuss how passing to the mean field limit can help us to deal with these questions when we suppose that the underlying graph is the complete graph.

My talk is based on collaborations with Julien Chevallier (France), Aline Duarte (Brazil), Kadmo Laxa (Brazil), Dasha Loukianova (France), Guilherme Ost (Brazil), Patricia Reynaud-Bouret (France), Etienne Tanré (France) and Josué Tchouanti (France).

Lounici, Karim

École polytechnique

Session: Interactive decision making

Talk: *A conversion theorem and minimax optimality for continuum contextual bandits*

We study contextual continuum bandits, where a learner sequentially observes a context vector and selects an action in a convex set to minimize a context-dependent function. The objective is to minimize all functions across contexts, leading to contextual regret, a stronger notion than static regret.

Assuming the objective functions are γ-Hölder continuous in the context ($0<\gamma\le1$), we show that any algorithm with sublinear static regret can be extended to achieve sublinear contextual regret. We establish a general static-to-contextual regret conversion theorem bounding contextual regret in terms of static regret.

We then derive implications for three settings: Lipschitz, convex, and strongly convex smooth bandits. For Lipschitz bandits with $\gamma=1$, we obtain a minimax optimal contextual regret rate scaling as $T^{(\dc+\dx+1)/(\dc+\dx+2)}$ (up to logarithmic factors), where d_c and d_x are the context and action dimensions. For noisy evaluations, convex and strongly convex smooth bandits share the same rate, scaling as $T^{(\dc+\gamma)/(\dc+2\gamma)}$ up to logs.

Finally, we prove a minimax lower bound showing that sublinear contextual regret may be impossible without continuity in context, and that our algorithms achieve near-optimal rates (up to logarithmic factors) for convex and strongly convex settings.

Lugosi, Gabor

ICREA & Pompeu Fabra University

Session: sub-Gaussian properties and robustness

Talk: *Robust, sub-Gaussian mean estimators in metric spaces*

The optimal accuracy one may achieve with a given confidence for the problem of estimating the mean of a random vector from i.i.d. data is well understood. When the data take values in more general metric spaces, an appropriate extension of the notion of the mean is the Fréchet mean. We study the performance of estimators of the Fréchet mean in general metric spaces under possibly heavy-tailed and contaminated data. In such cases, the empirical Fréchet mean is a poor estimator. We propose a general estimator based on high-dimensional extensions of trimmed means and prove general performance bounds. Unlike all previously established bounds, ours generalize the optimal bounds known for Euclidean data. We apply our results for metric spaces with curvature bounded from below, such as Wasserstein spaces, and for uniformly convex Banach spaces.

The talk is based on joint work with Daniel Bartl, Roberto Imbuzeiro Oliveira, and Zoraida F. Rico.

Lunde, Robert

Washington University in St. Louis

Session: Modeling complex network-linked data

Talk: *Conformal Prediction for Dyadic Regresson*

Dyadic regression, which involves modeling a relational matrix given covariate information, is an important task in statistical network analysis. We consider uncertainty quantification for dyadic regression models using conformal prediction. We establish finite-sample validity of our procedures for various sampling mechanisms under a joint exchangeability assumption. Our proof uses new results related to the validity of conformal prediction beyond exchangeability, which may be of independent interest. We also show that, under certain conditions, it is possible to construct asymptotically valid prediction sets for a missing entry under a structured missingness assumption.

Lvov, Nikita

McGill University

Session: Combinatorial aspects of random matrix theory

Talk: *On the satisfaction frequency of spectral characterization conditions*

We give the first specific conjectures on how frequently graphs satisfy sufficient conditions for being uniquely characterized by spectral information. These conjectures arise from a theoretical framework that we develop based on abstract-algebraic random matrix statistics. Specifically, we rephrase conditions from the literature in terms of Z[x]-modules associated to the adjacency matrix, and study the distribution of those modules in analytically tractable profinite random matrix ensembles. We apply this new method to two distinct conditions. The first requires square-freeness of the determinant of the walk matrix, and the second uses the discriminant of the characteristic polynomial.

Lyu, Zhongyuan

The University of Sydney

Session: Advances in Matrix Methods

Talk: *HeteroJIVE: Weighted Spectral Estimation for Shared Subspace Recovery under Multi-View Heteroskedasticity*

Many modern datasets consist of multiple related matrices measured on a common set of units, where the goal is to recover the shared low-dimensional subspace. While the Angle-based Joint and Individual Variation Explained (AJIVE) framework provides a solution, it relies on equal-weight aggregation, which can be strictly suboptimal when views exhibit significant statistical heterogeneity (arising from varying SNR and dimensions) and structural heterogeneity (arising from individual components). We propose HeteroJIVE, a weighted two-stage spectral algorithm tailored to such heterogeneity. Theoretically, we first revisit the "non-diminishing" error barrier with respect to the number of views K identified in recent literature for the equal-weight case. We demonstrate that this barrier is not universal: under generic geometric conditions, the bias term vanishes and our estimator achieves the $O(K^{-1/2})$ rate without the need for iterative refinement. Extending this to the general-weight case, we establish error bounds that explicitly disentangle the two layers of heterogeneity. Based on this, we derive an oracle-optimal weighting scheme implemented via a data-driven procedure. Extensive simulations corroborate our theoretical findings, and an application to TCGA-BRCA multi-omics data validates the superiority of HeteroJIVE in practice.

Mansanarez, Paul

Nantes Université/ULB

Session: Structure and Convergence in High-Dimensional Probability

Talk: *Edgeworth expansion on Wiener chaos*

Edgeworth expansions approximate the distribution of normalized sums of i.i.d. random variables as the Gaussian measure corrected by polynomial perturbations, with coefficients expressed in terms of the cumulants of the underlying distribution.

In this talk, we investigate Edgeworth expansions for functionals of a Gaussian field, which are not necessarily i.i.d. sums. Using the Maliavin-Stein method and semigroup analysis, we derive a bound on the total variation distance between the law of F and its so-called Edgeworth expansion: a modified Gaussian measure. The bounds depend only on p and a power of the fourth cumulant of F, which quantifies the distance of the law of F to the normal distribution.

Masak, Tomas

Wirtschaftsuniversitaet Wien

Session: Towards Robust and Explainable Functional Data Analysis

Talk: *The Functional Graphical Lasso*

We consider the problem of recovering conditional independence relationships between p jointly distributed Hilbertian random elements given n realizations thereof. We operate in the sparse high-dimensional regime, where $n \ll p$ and no element is related to more than $d \ll p$ other elements. In this context, we propose an infinite-dimensional generalization of the graphical lasso. We prove model selection consistency under natural assumptions and extend many classical results to infinite dimensions, doing away with unnecessary structural assumptions. The plug-in nature of our method makes it applicable to heterogeneous data measured under any observational regime, whether sparse or dense, and indifferent to serial dependence between samples. Moreover, the method can be understood as naturally arising from a coherent maximum likelihood philosophy.

Mayo-Iscar, Agustín

Universidad de Valladolid

Session: Robust Statistics for Matrix and Tensor Data

Talk: *Trimming and constraints based robust clustering for matrix-variate data*

Authors:

Luis Ángel García-Escudero (Universidad de Valladolid)

Marcus Mayrhofer (TU Wien)

Agustín Mayo-Íscar (Universidad de Valladolid)

Abstract:

We have designed a robust clustering estimator for data coming from a mixture of matrix normal distributions. On the one hand, we have used the Kronecker-product covariance structure based on row and column covariances. On the other hand, the mentioned robustness of this estimator is based on the joint application of trimming, in order to avoid the influence of contaminating sources, and eigenvalue constraints, in order to get a well posed estimation problem and to avoid spurious solutions. Our proposal, MTCLUST, generalizes the Matrix Minimum Covariance Determinant, which is a robust mean and covariance estimator of matrix-valued data. We will provide empirical evidence about how this methodology works based on artificial and real data.

Mayrhofer, Marcus

TU Wien

Session: Towards Robust and Explainable Functional Data Analysis

Talk: *Robustness and Explainability in Multivariate Functional Data Analysis*

We propose a method for robust covariance estimation for multivariate functional data by establishing a connection between a stochastic process with a separable covariance structure and the corresponding matrix-variate distribution of its basis representation. Based on this connection we use the Matrix Minimum Covariance Determinant (MMCD) approach in combination with a truncated multivariate functional Mahalanobis semi-distance for robust estimation of mean and covariance.

To go beyond outlier detection, we generalize multivariate outlier explanations based on Shapley values to decompose the truncated multivariate functional Mahalanobis semi-distance of individual observations into time-coordinate-specific contributions.

Our approach provides a framework for the generalization of other distance-based functional outlier detection approaches and lays the groundwork for future extensions to robust covariance estimation and explainable outlier detection for random surfaces.

McKenna, Benjamin

Georgia Institute of Technology

Session: Random matrices in machine learning

Talk: *Universality for the global spectrum of random inner-product kernel matrices*

In recent years, machine learning has motivated the study of what one might call "nonlinear random matrices." This broad term includes various random matrices whose construction involves the *entrywise* application of some deterministic nonlinear function, such as ReLU. We study one such model, an entrywise nonlinear function of a sample covariance matrix f(X*X), typically called a "random inner-product kernel matrix" in the literature. A priori, entrywise modifications of a matrix can affect the eigenvalues in complicated ways. However, a long line of work has established that the global spectrum of such matrices actually behaves in a simple way, described by free probability, when either the randomness in X is from a restricted family (such as Gaussian) or X has proportional sidelengths. We show that this description is universal, holding both for much more general randomness in X and for nonclassical aspect ratios. Joint work with Sofiia Dubova, Yue M. Lu, and Horng-Tzer Yau.

Mészáros, András

HUN-REN Alfréd Rényi Institute of Mathematics

Session: Combinatorial aspects of random matrix theory

Talk: *A phase transition for the corank of random band matrices over F_p*

For the corank of random matrices over the p element field, there is a widespread universality phenomenon. Namely, for a large class of random matrices over the p element field, their cokrank has the same limiting distribution as the corank of uniform random n*n matrices.

In this talk, we consider uniform random band matrices over the p element field, and determine the critical band width where the above mentioned universality breaks down.

This question is motivated by the analogous phase transition for the spectrum of Gaussian band matrices, which is a well studied problem in classical random matrix theory.

Minchev, Martin

University of Zurich and Sofia University

Session: Probability Distributions via Moments and Cumulants

Talk: *Moments of Exponential Functionals of Lévy Processes and Bernstein-Gamma Functions*

Exponential functionals of Levy processes are a classical object in probability, which appear, for example, through the Lamperti time change linking self-similar Markov processes to Levy processes, and also in mathematical finance in connection with Asian options.

A standard argument shows that their moments satisfy a recurrence relation of period one, and this leads naturally to the question whether the law is determined by its moments. In this talk we first recall the early results of Bertoin and Yor from the early 2000s. We then turn to the extension from integer moments to complex exponents, that is, to the Mellin transform. Patie and Savov expressed this transform in terms of a new class of Gamma-type functions, called Bernstein-Gamma functions.

We present some basic properties of these functions and explain how they arise in the study of exponential functionals. We also discuss several open questions on moment determinacy for exponential functionals, as well as possible matrix-valued extensions of Bernstein-Gamma functions.

Joint work with Mladen Savov.

Misiakiewicz, Theodor

Yale University

Session: Structured and Expressive Models in High Dimensions

Talk: *Learning Multi-Index Models in High-Dimensions*

We study the problem of learning multi-index models (MIMs), where the label depends on the high-dimensional input vector only through an unknown low dimensional projection. Exploiting the equivariance of this problem under the orthogonal group, we obtain a sharp harmonic-analytic characterization of the learning complexity for MIMs with spherically symmetric inputs. Specifically, we derive statistical and computational complexity lower bounds within the Statistical Query (SQ) and Low-Degree Polynomial (LDP) frameworks. These bounds decompose naturally across spherical harmonic subspaces. Guided by this decomposition, we construct a family of spectral algorithms based on harmonic tensor unfolding that sequentially recover the latent directions and (nearly) achieve these SQ and LDP lower bounds. Depending on the choice of harmonic degree sequence, these estimators can realize a broad range of trade-offs between sample and runtime complexity.

Moyal, Pascal

Université de Lorraine / Inria

Session: Advances in Learning, Optimization, and Generalization in Modern Stochastic Systems

Talk: *Stochastic control and reinforcement learning for greedy dynamic matching models*

We consider a dynamic matching model on a general (non-bipartite) graph: items arrive one by one into a system, and depart by pairs of two, as

soon as they are matched, following compatibility rules that are given by the underlying graph structure. We assume that the matching policy is greedy, in the sense that an incoming item

that finds at least one compatible item in the queue, must be matched right away. Using the tools of dynamic programming and reinforcement learning, we address the optimal control of such models following two different lines:

First, we show that, for a simple graph structure, a threshold-type policy is optimal for minimizing holding costs. Second, we show how to optimize the access control to the queue, by using a Policy-Gradient Reinforcement Learning (PGRL) algorithm.

Müller, Hans-Georg

UC Davis

Session: Inference Without Borders: Methods for Random Objects Beyond Euclidean Data

Talk: *Geodesic Causal Inference*

Causal inference methods have emerged as a popular tool to adjust for confounding and imbalance in regression models with scalar responses and also distributional responses. We introduce a general framework for causal inference when responses reside in general geodesic metric spaces, drawing on a novel geodesic calculus, which facilitates the quantification of treatment effects through the concept of geodesic average treatment effect (GATE). Employing Fréchet regression, we obtain doubly robust estimation of GATE with consistency and rates of convergence. Theoretical results are complemented with various data illustrations. The geodesic framework can also be employed for other causal designs such as difference-in-differences and discontinuity designs. This talk is based on joint work with Daisuke Kuriso, Yidong Zhou and Taisuke Otsu.

Munk, Axel

Göttingen University

Session: Statistical Optimal Transport

Talk: *ANOVA-type Wasserstein tests for factorial designs*

We introduce a general framework for testing statistical hypotheses for probability measures supported on finite spaces, which is based on optimal transport (OT). These tests are inspired by the analysis of variance (ANOVA) and its nonparametric counterparts. They allow for testing linear relationships in factorial designs between discrete probability measures and are based on pairwise comparisons of the OT distance and corresponding barycenters. To this end, we derive under the null hypotheses and (local) alternatives the asymptotic distribution of empirical OT costs and the empirical OT barycenter cost functional as the optimal value of linear programs with random objective function. In particular, we extend existing techniques for probability to signed measures and show directional Hadamard differentiability and the validity of the functional delta method. We discuss computational issues, permutation and bootstrap tests, and back up our findings with simulations. We illustrate our methodology on a data set from biometric fingerprint identification.

This is joint work with Michel Groppe, and Shayan Hundrieser and Linus Niemöller.

Musio, Monica

University of Cagliari

Session: Proper scoring rules

Talk: *The Central Role of Proper Scoring Rules in Modern Statistics: Theory, Robustness, and Applications to Complex Data*

Proper scoring rules (PSRs) provide a foundational framework for both the evaluation of probabilistic forecasts and the construction of tailored inferential procedures. A SR is proper if it incentivizes honesty, ensuring that the expected score is optimized when the quoted distribution matches the true underlying process. While the logarithmic score is the standard choice due to its link with likelihood-based inference, its application can be limited by sensitivity to anomalous observations or the presence of intractable normalizing constants. This has motivated the development of alternative rules characterized by specific properties, such as robustness or homogeneity. In this talk, we review the theoretical foundations and practical relevance of PSRs, highlighting their deep connections with M-estimation and subjective probability. We discuss how these rules induce unbiased estimating equations, leading to consistent estimators whose asymptotic precision is governed by the Godambe information. Special emphasis is placed on applications involving complex data structures. For compositional data, we demonstrate how SRs can be integrated with isometric log-ratio transformations to respect the geometry of the simplex. Furthermore, we explore extreme value analysis, where the choice of the SR is critical. We compare the use of rules such as the Tsallis for handling contamination and the Hyvärinen for its homogeneity, which allows bypassing complex normalizing constant.

Nguyen, Hoi

The Ohio State University

Session: Structure and Convergence in High-Dimensional Probability

Talk: *Random polynomials: old and new*

The study of zeros of random polynomials dates back to the early twentieth century and remains an active area of research. In this talk, we discuss recent developments, focusing on robust frameworks for establishing universality. Topics include global and local correlations of zeros, the variance in the number of real zeros, concentration phenomena, central limit theorem fluctuations, and related questions.

Nguyen, Oanh

Brown University

Session: Structure and Convergence in High-Dimensional Probability

Talk: *Random polynomials with dependent coefficients*

Classical results on random polynomials largely focus on the idealized setting of independent coefficients. But what happens when this assumption is relaxed? In this talk, we explore a range of random polynomials with dependent structures, highlighting both the challenges they pose and the techniques used to analyze them. We will discuss models with weakly dependent coefficients, as well as those arising from Ising interactions. This talk is based on joint work with Jürgen Angst, Guillaume Poly, and Oren Yakir.

Nordhausen, Klaus

University of Helsinki

Session: Towards Robust and Explainable Functional Data Analysis

Talk: *Functional AMUSE for multivariate functional separable time series*

Second-order source separation for vector-valued time series assumes that an observable stationary p-variate process can be represented as a linear mixture of p independent components. We extend this framework to the setting of multivariate functional stationary time series. The corresponding model is introduced and its identifiability properties are discussed. Furthermore, we generalize the classical AMUSE estimator to the functional setting under a separable covariance operator assumption. The performance of the proposed method is investigated through a simulation study.

Ogden, Todd

Columbia University

Session: Structured Statistical Learning and Post-Learning Inference in High-Dimensional Settings

Talk: *Functional Metric Learning for Classification Problems*

Metric-based classification methods, such as K-nearest neighbors, are widely used, and their success often depends on the choice of distance measure. While metric learning has been shown to improve classification in multivariate settings, it faces challenges in high-dimensional spaces and when data are noisy. In many modern applications, including electroencephalograph (EEG) signals and brain images, data are more naturally represented as functional data, which exhibit inherent smoothness, continuity, and structural dependence. Motivated by this, we propose a functional metric learning framework that integrates functional principal component analysis (FPCA) with Large Margin Nearest Neighbors (LMNN) to define and learn an optimal semimetric that is tailored for functional data. Rather than relying on prespecified distances, our approach accounts for the structure of both one-dimensional and multidimensional functional data, accounts for noise, and adapts the semimetric to the classification task. Through simulations and an application to EEG data for alcoholism classification, we demonstrate that our method outperforms alternative approaches and demonstrates that frontal and parietal brain activity plays a crucial role in classification. These results establish functional metric learning as a flexible and powerful tool for improving classification in complex functional datasets.

Olhede, Sofia

EPFL

Session: Modeling complex network-linked data

Talk: *Modelling multiplex networks*

This talk will discuss how to model and make inferences of multiplex networks. One point-of-view starts from a special class of models based on multivariate graph limits (see Skeja and Olhede (2024)). These can be represented as a limiting object known as a decorated graphon (see Dufour and Olhede (2024)), where a more complex object is represented by a generalization of a graph limit. The choice of object will depend on the particular application at hand, and we will discuss the particular case of coma patients (Verdeyme et al (2025)). We will discuss choices of parameterization, and choices of summary statistics for particular models, highlighting both the benefits and challenges of multiplex observations.

This is joint work with Charles Dufour, Jake Grainger, Anda Skeja, Karl Sawaya and Arthur Verdeyme

Osman, Mohammed

Queen Mary University of London

Session: Random Matrices

Talk: *On the convex hull and coverage threshold of Ginibre eigenvalues*

We consider two geometric properties of Ginibre eigenvalues: the convex hull and the coverage threshold. The latter is the radius for which the union of disks of this radius centred at each eigenvalue covers a given region. We obtain limit theorems for the perimeter, area and number of vertices of the convex hull and the coverage threshold of disks.

Pammer, Gudmund

TU Graz

Session: Statistical Optimal Transport

Talk: *A weak transport approach to the Schrödinger-Bass problem*

The Schrödinger-Bass problem is a semimartingale optimal transport problem connecting the celebrated Schrödinger bridge with the Bass martingale problem. In this talk, I will explain how the dynamic Schrödinger-Bass problem admits an equivalent static formulation as a weak optimal transport problem with an explicit cost. This reformulation yields existence and duality results, provides a structural characterization of the optimal semimartingales, and naturally leads to a Sinkhorn-type iteration for computation.

Pananjady, Ashwin

Georgia Tech

Session: Recent Advances in Iterative Methods for Random Optimization

Talk: *State evolution beyond first order methods*

We consider the dynamics of iterative optimization algorithms when applied to instances with high-dimensional, random data. When the algorithm of choice is a first-order method, it is known that the dynamics of the method are well approximated by a low-dimensional deterministic recursion known as state evolution. In this paper, we move beyond first-order methods and develop a rigorous state evolution for a far larger set of algorithms. We show that this state evolution can be written in a “canonical form”, allowing us to argue existence and uniqueness of the deterministic updates. Along the way, we develop a variant of Bolthausen’s conditioning method that relies on a sequential variant of Gordon’s Gaussian comparison inequality and enables a fully non-asymptotic analysis.

Panigrahi, Snigdha

University of Michigan

Session: Post-selection inference

Talk: *Hierarchical Clustering with Confidence*

Agglomerative hierarchical clustering is one of the most widely used approaches for exploring how observations in a dataset relate to each other. However, its greedy nature makes it highly sensitive to small perturbations in the data, often producing different clustering results and making it difficult to separate genuine structure from spurious patterns. In this talk, I show how randomizing hierarchical clustering can be used not only to assess stability, but also to enable valid hypothesis testing based on the clustering results.

We introduce a simple randomization scheme together with a test at each node of the hierarchical clustering dendrogram that controls the Type I error rate. Our test works with any hierarchical linkage without case-specific derivations, is more powerful than existing selective inference methods, and can be used to estimate the number of clusters with a guarantee against overestimation.

Panorska, Anna

University of Nevada Reno, USA

Session: Recent Advances in Distribution Theory

Talk: *From Atmospheric Rivers to Flood Risk: A Multivariate Model for Extreme Precipitation*

Water resources and flood risk in the western United States are commonly driven by large scale precipitation events lasting several consecutive days. Some of the largest precipitation events are driven by the Atmospheric Rivers, the "rivers in the sky". We propose a multivariate model for the extreme precipitation events describing their duration, intensity, and maximum. We also link it to the atmospheric flow patterns, like those producing the Atmospheric Rivers. As the most important questions related to the impact of extreme precipitation are about the risk or probability of these extreme events, our ultimate goal is to employ the statistical model along with the meteorological and statistical methods leading to estimation of these probabilities.

Parry, Matthew

University of Otago

Session: Modern Approaches to Inference and Estimation

Talk: *Application of control theory to online parameter estimation*

Control theory studies feedback-driven decision-making for dynamical systems. In engineering, a great deal is known about how to design adaptive control processes to stabilise, regulate, or optimise system behaviour over time. These ideas can be applied in statistics to the problem of online parameter estimation. In this framework, the prediction-generating process is treated as the system, and the prediction error as the feedback signal used to update parameters from sequentially observed data. Prediction-error methods, for example, estimate parameters by minimising a loss function based on the one-step-ahead prediction error.

We develop a learning framework based on proportional–integral–derivative (PID) controllers for online parameter estimation. We show that the integral component removes persistent error, while the derivative component helps to damp over-correction. In addition to connecting PID learning to momentum-based and adaptive optimisation methods, we demonstrate that PID learning can remain stable with a constant learning rate and is robust to time-varying data-generating processes.

Pathak, Reese

UC Berkeley / Cornell University

Session: IMS Lawrence D. Brown Ph.D. Student Award

Talk: *Revisiting M-estimation: risk bounds, optimality results, and probabilistic connections*

M-estimation procedures, such as empirical risk minimization (ERM) and least squares estimation (LSE), are pervasive in statistics and machine learning. Surprisingly, however, a complete understanding of their statistical properties is lacking, even in the most basic settings. In this talk we discuss recent results on the LSE and ERM, including recently obtained risk bounds, deviation inequalities, and optimality and suboptimality results that improve upon the classical ones. We also discuss some new connections between the analysis of these estimators and important probabilistic and geometric quantities such as the Gaussian width.

Based on a series of joint works with: L. Aolaritei (UC Berkeley), M. I. Jordan (UC Berkeley / INRIA), A. Ulichney (UC Berkeley), N. Zhivotovskiy (UC Berkeley).

Perrin, Alexandre

CMAP Ecole Polytechnique

Session: Interdisciplinary perspectives at the interface between probability theory and cell biology

Talk: *DNA replication and cell division cycles coordination in E. coli: a mathematical approach to cell physiology*

Despite extensive research, the quantitative principles that govern the coordination between DNA

replication and cell division in bacteria remain debated. Multiple theoretical models have been pro-

posed, some postulating that a single regulatory process is sufficient to ensure replication–division co-

ordination, while others argue that two concurrent processes are required for robust control. To enable

the comparison of these approaches, we developed a unifying mathematical framework within which

models can be consistently formulated and quantitatively compared. Through theoretical analysis,

this talk will present the necessary and sufficient conditions under which independent replication and

division cycles recover physiological cell behaviours. Beyond the correlation-based analyses extensively

used to date, this talk will discuss within a comprehensive statistical framework that double-process

models more accurately recapitulate experimental data across all growth conditions. Finally, a novel

model will be introduced that robustly captures the replication-division coordination in every growth

regime, thereby providing a foundation for future mechanistic studies.

Podgorski, Krzysztof

Lund University

Session: Recent Advances in Distribution Theory

Talk: *A class of continuous-time Gaussian mean-variance mixture state-space models simultaneously accounting for asymmetry, tailedness, and peakedness*

The generalized inverse Gaussian (GIG) distribution underpins generalized hyperbolic (GH) models used in Gaussian mean–variance mixtures to capture asymmetry, peakedness, and heavy tails. However, although GIG distributions are infinitely divisible (ID), they are not closed under convolution, limiting their use in continuous-time and Lévy-based frameworks.

This paper introduces a related class of ID mixing distributions, termed IGG, defined as a convex combination of inverse Gaussian and gamma distributions. The IGG class retains the same number of parameters as GIG and, when used in Gaussian mean–variance mixtures, captures location, scale, skewness, and kurtosis, with an additional parameter governing the mixing proportion, enhancing flexibility.

A key advantage is closure under convolution, ensuring consistency under time scaling and irregular sampling. This enables direct use in continuous-time modeling via subordination of Brownian motion with drift to an IGG Lévy process, yielding Gaussian–IGG (NIGG) distributions. The class includes the normal–inverse Gaussian and generalized asymmetric Laplace as special cases.

With intuitive parameter interpretation, the model avoids ambiguities common in GH-based time models and is well-suited for state-space formulations, as illustrated by stochastic volatility applications.

Qu, Annie

University of California Santa Barbara

Session: Statistical Foundations of Diffusion Models

Talk: *SYNTAX-GUIDED DIFFUSION LANGUAGE MODELS WITH USER-INTEGRATED PERSONALIZATION*

Large language models have achieved remarkable progress in generating human-like text, yet their outputs often lack structural diversity, limiting personalized expression. Recent advances in diffusion models offer new opportunities for text generation, surpassing the limitations of autoregressive paradigms. Building on the idea of hierarchical modeling, we develop a syntax-guided diffusion language model that incorporates latent syntactic variables to enhance text quality, diversity, and personalized conditioning. The proposed framework decomposes the text distribution into a syntactic prior and a conditional text distribution, enabling interpretable control over both syntactic and lexical components. We design a cascaded diffusion process for syntax-to-text generation and extend it to a generalized noncascaded formulation with a unified attention mechanism. To achieve fine-grained personalization,

we introduce shared-latent representations that integrate information across users to capture style-specific lexical and syntactic patterns, supporting zero-shot inference. Extensive experiments demonstrate that the proposed approach substantially improves fluency, diversity, and stylistic fidelity.

Further qualitative analyses highlight its interpretability and flexibility in learning personalized patterns.

Racz, Miklos

Northwestern University

Session: Statistical Inference on Preferential Attachment Networks

Talk: *Community Detection and Modularity in the Preferential Attachment Block Model*

I will discuss community detection in the preferential attachment block model (PABM). The key to the results is a detailed understanding of modularity of partitions in PABM, which may be of independent interest. This is based on joint work with Shuwen Chai.

Radojicic, Una

TU Wien

Session: Robust Statistics for Matrix and Tensor Data

Talk: *A Minimum Covariance Determinant Approach for Matrix and Tensor Data*

This work introduces the Matrix Minimum Covariance Determinant (MMCD), a robust estimator for location and covariance tailored to matrix-valued data. Unlike classical approaches that require vectorization—often leading to high-dimensional problems—MMCD preserves the inherent matrix structure and consistently estimates the mean matrix along with rowwise and columnwise covariance matrices within matrix-variate elliptical distributions. The estimators are matrix-affine equivariant and achieve a higher breakdown point than any affine equivariant multivariate estimator applied to vectorized data.

An efficient algorithm with convergence guarantees is proposed, enabling practical implementation. MMCD-based robust Mahalanobis distances provide a reliable tool for outlier detection. Furthermore, Shapley values are extended to the matrix setting, allowing decomposition of squared Mahalanobis distances into contributions from rows, columns, or individual cells.

The framework naturally extends to tensor-valued data, enabling analogous robust estimation and interpretability across higher-order structures. Both theoretical results and simulations demonstrate that MMCD outperforms vectorization-based methods in robustness and computational efficiency, supported by real-world applications.

Rakhlin, Alexander

MIT

Session: Interactive decision making

Talk: *Elements of Interactive Decision Making*

Machine learning methods are increasingly deployed in interactive environments, ranging from dynamic treatment strategies in medicine to fine-tuning of LLMs using reinforcement learning. In these settings,

the learning agent interacts with the environment to collect data and necessarily faces an exploration-exploitation dilemma. We present a general framework for interactive decision making that subsumes

multi-armed bandits, contextual bandits, structured bandits, and reinforcement learning. We focus on both the statistical aspect of learning---aiming to develop a tight characterization of sample complexity in terms of properties of the class of models---and on the basic algorithmic primitives.

Ramanan, Kavita

Brown University

Session: Random processes on graphs

Talk: *Quantitative Unimodular Propagation of Chaos and its Consequences*

Interacting particle systems, or large collections of randomly evolving particles that interact locally with respect to an underlying network or graph, model a plethora of phenomena in physics, engineering, neuroscience, epidemiology and machine learning. When the underlying network is sparse, as is the case of many real-world networks, neighboring particles remain strongly correlated and there is no propagation of chaos in the classical sense. However, recent work has shown that there is a certain unimodular propagation of chaos and that the limit of the empirical measure can be characterized as the law of a local-field equation. We establish quantitative versions of unimodular propagation of chaos and describe some consequences, including showing how the local-field equation naturally interpolates to the mean-field equation as the graph size goes to infinity. The proofs combines techniques from stochastic analysis and random graph theory.

Ren, Zhimei

University of Pennsylvania

Session: Inference Beyond the Textbook: Robustness, Adaptivity, and Cross-Fitting

Talk: *Assumption-lean weak limits and tests for two-stage adaptive experiments*

Adaptive experiments are becoming increasingly popular in real-world applications for effectively maximizing in-sample welfare and efficiency by data-driven sampling. Despite their growing prevalence, however, the statistical foundations for valid inference in such settings remain underdeveloped. Focusing on two-stage adaptive experimental designs, we address this gap by deriving new weak convergence results for mean outcomes and their differences. In particular, our results apply to a broad class of estimators, the weighted inverse probability weighted (WIPW) estimators. In contrast to prior works, our results require significantly weaker assumptions and sharply characterize phase transitions in limiting behavior across different signal regimes. Through this common lens, our general results unify previously fragmented results under the two-stage setup. We further establish quantitative convergence rates in bounded-Lipschitz distance that reveal the fundamental trade-off between exploitation and inferential stability. To address the challenge of potential non-normal limits in conducting inference, we propose a computationally efficient and provably valid simulation-based method for obtaining critical values of the non-normal limiting distributions under the null, enabling practical hypothesis testing. Our results and approaches are sufficiently general to accommodate various adaptive experimental designs, including batched bandit and subgroup enrichment experiments. Simulatio

Rico, Zoraida

Bocconi University

Session: sub-Gaussian properties and robustness

Talk: *Robust and Sub-Gaussian Estimation of Means and Covariances*

Estimating the mean and covariance under heavy tails and adversarial contamination remains a central challenge in robust statistics.

In this talk, we revisit the classical trimmed mean estimator for one-dimensional mean estimation, providing new finite-sample insights. We show that the trimmed mean achieves optimal performance in this setting and satisfies a strong form of the CLT.

We then turn to covariance estimation for a d-dimensional random vector from an i.i.d. sample. We show that, even in the presence of relatively heavy tails and adversarial contamination, this estimator achieves the optimal dimension-free rate of convergence.

Ringeard, Timothé

Université Paris Cité

Session: Interdisciplinary perspectives at the interface between probability theory and cell biology

Talk: *A stochastic continuous-space model and its large population limit for the dynamic of actin polymerisation and depolymerisation in the cell*

The network of actin proteins in the cell is responsible for many structural characteristics and behaviour. In particular, during cell division, actin monomers form very long polymers around the equator, creating a ring that will contract itself to separate the two new cells. In order to better understand this phenomenon, we model the situation by a stochastic process where actin monomers follow Brownian diffusion, and can agglomerate into polymers at a prescribed location, corresponding to the contractile ring. The macroscopic behaviour of the process can be understood by a large population limit, in which the number of actin monomers goes to infinity. With the right scaling of the parameters, the limit is deterministic and follows a PDE with a singularity at the location of the polymers.

Rush, Cynthia

Columbia University

Session: Unified Perspectives on Regularization in Modern Statistics

Talk: *The distributionally robust prediction error of the square root LASSO and related estimators*

We study the classical problem of predicting an outcome variable using a linear combination of a d-dimensional covariate vector. We are interested in linear predictors whose coefficients are generated by the square root LASSO and related estimators. We provide conditions under which linear predictors based on these estimators minimize the worst-case prediction error over a ball of distributions determined by a type of max-sliced Wasserstein metric. A detailed analysis of the statistical properties of this metric yields a simple recommendation for the choice of regularization parameter. The suggested order of the regularization parameter, after a suitable normalization of the covariates, is typically given by the dimension d divided by the sample size, up to logarithmic factors. Our recommendation is computationally straightforward to implement, pivotal, has provable out-of-sample performance guarantees, and does not rely on sparsity assumptions about the true data generating process.

This is joint work with Jose Montiel Olea, Amilcar Velez and Johannes Wiesel.

Saksman, Eero

University of Helsinki

Session: Recent advances in probabilistic number theory

Talk: *Chaos and Riemann zeta function on the critical line*

Abstract: We will discuss some results, both old and new (in preparation), on the statistics of the Riemann zeta function on short intervals on the critical line. The talk is based on collaboration with Adam Harper (Warwick) and Christian Webb (Helsinki).

Samworth, Richard

University of Cambridge

Session: Statistical Foundations of Diffusion Models

Talk: *Learn the score*

Score estimation has recently emerged as a key modern statistical challenge, due to its pivotal role in generative modelling via diffusion models. Moreover, it is an essential ingredient in a new approach to linear regression via convex M-estimation, where the corresponding error densities are projected onto the log-concave class. Motivated by these applications, we study the minimax risk of score estimation with respect to squared squared error loss, weighted by the true underlying log-concave distribution. Such distributions have decreasing score functions, but on its own, this shape constraint is insufficient to guarantee a finite minimax risk. We therefore define subclasses of log-concave densities that capture two fundamental aspects of the estimation problem. First, we establish the crucial impact of tail behaviour on score estimation by determining the minimax rate over a class of log-concave densities whose score function exhibits controlled growth relative to the quantile levels. Second, we explore the interplay between smoothness and log-concavity, and determine the minimax rate up to poly-logarithmic factors. Our upper bounds are attained by a locally adaptive, multiscale estimator constructed from a uniform confidence band for the score function. This study highlights intriguing differences between the score estimation and density estimation problems over this shape-constrained class.

Schönlieb, Carola-Bibiane

cbs31@cam.ac.uk

Session: Statistical Optimal Transport

Talk: *Optimal Transport for Inverse Problems*

Optimal transport has emerged as a powerful mathematical framework for comparing structured data distributions and has recently shown significant promise in inverse problems. In this talk, I will discuss how transport-based discrepancies, priors, and geometries can be leveraged to improve robustness, reconstruction quality, and interpretability in inverse imaging problems, particularly in settings involving sparse, noisy, or distribution-shifted data. I will further highlight recent developments connecting Wasserstein metrics with variational regularisation and modern data-driven approaches.

Seiler, Marco

Goethe University Frankfurt

Session: Particle Systems in Random Environments

Talk: *The contact process on dynamical random trees with degree dependence*

The contact process models the spread of an infection in a structured population given through a graph, where individuals are infected with a certain rate and recover with rate 1. If this graph is a Bienaymé-Galton-Watson (BGW) tree, the contact process exhibits a well-known dichotomy in its survival behaviour. If the offspring distribution has no exponential moments, then the phase transition of survival is trivial, which means that for every infection rate there is a positive probability to survive. On the other hand, if some exponential moments exist, then the process dies out almost surely for small infection rates.

In this talk we study the contact process on a dynamically evolving BGW tree. Initially edges are assigned an open/closed state with a degree-dependent connection probability. Additionally, edges update at a degree-dependent rate, after which each edge is independently reopened or closed with the same degree-dependent probability as before. Finally, the contact process is only allowed to spread via open edges. We determine the impact of the update speed and the connection probability on the existence of a non-trivial phase transition. In particular, we provide conditions depending on the connection probability and update rate which guarantee a trivial phase transition.

This talk is based on joint work with Natalia Cardona-Tobón, Marcel Ortgiese and Anja Sturm

Sen, Subhabrata

Harvard University

Session: Honest Inference for Flexible Models

Talk: *Spectral algorithms for signal recovery from multi-modal data*

We will study spiked Wigner matrices with correlated spikes. The model is motivated by multimodal Principal Component Analysis (PCA). We derive and analyze a spectral algorithm for signal recovery in this problem. The spectral operator is obtained by linearization of Approximate Message Passing (AMP). We discover a sharp phase transition in the behavior of this spectral operator ---above a critical SNR threshold, the spectrum has an outlier, and the corresponding top eigenvector is correlated with the latent signals. This generalizes well-known phase transitions for spectral methods used in unimodal data analysis. We will discuss inference for the latent signals based on the top eigenvector of this new spectral operator.

Soloff, Jake

University of Michigan

Session: Empirical Bayes for approximate oracle inference

Talk: *Calibration in multiple hypothesis testing*

Among error criteria for large-scale hypothesis testing, the local false discovery rate (lfdr) is the one most directly tied to the reliability of an individual finding. Its interpretation, however, traditionally relies on a Bayesian two-groups model. This talk develops new perspectives on empirical Bayes and compound-decision inference in this context. The guiding idea is to seek procedures that, without prior knowledge, attain in expectation the same local guarantees that an oracle Bayes rule attains almost surely. First, I introduce a simple, nonparametric procedure that exactly controls the expected maximum lfdr among the discoveries; equivalently, the probability that the least promising discovery is false. I will then discuss recent and ongoing work connecting multiple testing to probabilistic forecasting, giving a prior-free interpretation of lfdr as a calibrated conditional error probability and suggesting a broader framework for interpretation of compound decisions.

Sondhi, Arjun

Northwell Health

Session: New Researchers Group Session

Talk: *Doubly regularized generalized linear models for spatial observations with high-dimensional covariates*

A discrete spatial lattice can be cast as a network structure over which spatially correlated outcomes are observed. A second network structure may also capture similarities among measured features, when such information is available. Incorporating the network structures when analysing such doubly structured data can improve predictive power, and lead to better identification of important features in the data-generating process. Motivated by applications in spatial disease mapping, we develop a new doubly regularized regression framework to incorporate these network structures for analysing high-dimensional datasets. Our estimators can be easily implemented with standard convex optimization algorithms. In addition, we describe a procedure to obtain asymptotically valid confidence intervals and hypothesis tests for our model parameters. We show empirically that our framework provides improved predictive accuracy and inferential power compared with existing high-dimensional spatial methods. These advantages hold given fully accurate network information, and also with networks which are partially misspecified or uninformative. The application of the proposed method to modelling COVID-19 mortality data suggests that it can improve the prediction of deaths beyond standard spatial models, and that it selects relevant covariates more often.

Sørensen, Helle

University of Copenhagen

Session: Recent advances in functional data analysis and metric statistics

Talk: *Local inference for functional time series*

Local inference is a technique in functional data analysis where hypotheses are tested along the domain of the functional data - in principle involving infinitely many tests. The multiple testing problem is addressed by performing tests over intervals and exploiting the continuity of the data. In this talk we consider local inference for the situation where data is a time series with each data unit being a curve. We approch two questions: (1) In which parts of the functional domain is there dependence in the time series, and (2) how can local inference for hypotheses concerning fixed effects be carried out in the presence of dependence? Our tests are based on permutations and are thus non-parametric of nature. As an application we consider data consisting of daily ice cover in the Baffin Bay in the period from 1979 to 2023.

Sriperumbudur, Bharath

The Pennsylvania State University

Session: Inference Without Borders: Methods for Random Objects Beyond Euclidean Data

Talk: *Spectral regularized kernel two-sample tests*

Over the last decade, an approach that has gained a lot of popularity in tackling non-parametric testing problems on general (i.e., non-Euclidean) domains is based on the notion of reproducing kernel Hilbert space (RKHS) embedding of probability distributions. The main goal of our work is to understand the optimality of two-sample tests constructed based on this approach. First, we show that the popular MMD (maximum mean discrepancy) two-sample test is not optimal in terms of the separation boundary measured in Hellinger distance. Second, we propose a modification to the MMD test based on spectral regularization by taking into account the covariance information (which is not captured by the MMD test) and prove the proposed test to be minimax optimal with a smaller separation boundary than that achieved by the MMD test. Third, we propose an adaptive version of the above test, which involves a data-driven strategy to choose the regularization parameter and show the adaptive test to be almost minimax optimal up to a logarithmic factor. Moreover, our results hold for the permutation variant of the test, where the test threshold is chosen elegantly through the permutation of the samples. Through numerical experiments on synthetic and real-world data, we demonstrate the superior performance of the proposed test in comparison to many popular two-sample tests.

Stephan, Ludovic

ENSAI - CREST

Session: Random matrices in machine learning

Talk: *Spectral methods for graph inference: the case of community detection*

Community detection is the task of recovering a hidden cluster structure from network information. Spectral methods, especially those based on the graph adjacency matrix, are a very popular family of algorithms but they can fail when the average graph degree is too low. I will present several remediations to this phenomenon, from the celebrated non-backtracking method and its variants to more efficient but less theoretically tractable methods such as the Bethe Hessian.

Stoyanov, Jordan

Bulgarian Academy of Sciences

Session: Probability Distributions via Moments and Cumulants

Talk: *Discussion on Open Questions Involving Moments and Cumulants*

The IMS meeting is an excellent opportunity for a comprehensive discussion on outstanding questions related to moments and cumulants. For non-linear Box-Cox transformations of random data, it is not clear whether the M- or C-determinacy of the transformed data is preserved or lost. To study determinacy in conjunction with infinite divisibility poses extra challenges. Numerous questions remain unanswered. More questions are related to the important qualitative result by C. Berg and co.: If a distribution F possesses finite moments and is M-indeterminate, there exist infinitely many distributions of any kind, continuous, discrete, singular, all sharing the same moments as F. There are several checkable conditions that enable us to establish M-(in)determinacy for absolutely continuous and discrete distributions. Only a few instances are known where infinite families of both continuous and discrete distributions are explicitly identified. The Log-Normal distribution serves as a notable example. The M-indeterminacy of power 3 of a r.v. with either a normal, exponential or IG distribution can be shown in a few ways, we still do not know any 'discrete' solution in these and many other cases.

I will present several open questions arising from my research. Attendees of the discussion will have the opportunity to share their own open questions. Additionally, colleagues unable to attend the meeting have graciously provided their open questions to be addressed publicly.

Sur, Pragya

Harvard University

Session: Structured and Expressive Models in High Dimensions

Talk: *Optimality of Self-distillation in Spiked Covariance Models*

In modern artificial intelligence pipelines, self-distillation has emerged as a promising technique for improving model performance without requiring additional data. Yet it remains unclear when self-distillation yields improved out-of-sample risk or better estimation compared to commonly used statistical estimators. This talk will present a rigorous framework for analyzing the performance of a broad class of estimators in high dimensions, which we term spectral-shrinkage estimators. We show that for spiked covariance models, self-distillation achieves uniquely optimal prediction and estimation performance, outperforming well-known methods such as ridge regression, minimum norm interpolation, and principal components regression. This stands in contrast to the behavior observed under isotropic covariances, where suitably tuned ridge regression is optimal among these shrinkage estimators. Time permitting, we will also present a federated learning setting in which multiple data centers share their self-distilled estimators and a central server seeks to aggregate them optimally. We characterize optimal aggregation strategies in this context. Together, these results elucidate why self-distillation improves predictive performance and provide a broader statistical framework that connects it to other shrinkage-based methods.

Taeb, Armeen

University of Washington

Session: Advances in Unsueprvised Learning & Causal Discovery

Talk: *Model-oriented distances between causal graphs via posets*

A well-defined distance on the parameter space is key to evaluating estimators, ensuring consistency, and building confidence sets. While there are typically standard distances to adopt in a continuous space, this is not the case for combinatorial parameters such as graphs that represent statistical models. Defined on the graphs alone, existing proposals like the structural Hamming distance ignore the structure of the model space and can thus exhibit undesirable behaviors. We propose a model-oriented framework for defining the distance between graphs that is applicable across different graph classes. Our approach treats each graph as a statistical model and organizes the graphs in a partially ordered set based on model inclusion. This induces a neighborhood structure, from which we define the model-oriented distance as the length of a shortest path through neighbors, yielding a metric in the space of graphs. We apply this framework to probabilistic undirected graphs, causal directed acyclic graphs, probabilistic completed partially directed acyclic graphs, and causal maximally oriented partially directed acyclic graphs. We analyze theoretical and empirical behaviors of the model-oriented distance and draw comparison with existing distances. By exploiting the underlying poset structures, we develop algorithms for computing and bounding the proposed distance that scale to moderate-sized graphs.

Talyigas, Zsofia

BOKU University

Session: Interdisciplinary perspectives at the interface between probability theory, ecology and evolution

Talk: *Selection of the fittest or selection of the luckiest: the emergence of Goodhart's law in evolution*

Natural selection is commonly assumed to become more effective as it becomes stronger. However, selection acts on phenotypes rather than directly on genotypes, and phenotypic success is inherently noisy. Here we study how this mismatch shapes long-term evolutionary dynamics. Using a minimal stochastic model in which individuals inherit genetic fitness while selection acts on noisy phenotypic expressions, we show that increasing selection strength accelerates adaptation only up to a critical threshold. Beyond this point, stronger selection paradoxically slows evolution and erodes genetic diversity by favoring the luckiest individuals rather than the genetically fittest.

We identify two distinct evolutionary regimes—selection of the fittest and selection of the luckiest—separated by a sharp transition. This transition corresponds to a previously unrecognized change in the structure of traveling fitness waves, from semipulled to fully pulled fronts, with profound consequences for adaptation speed and genealogical structure. Our results reveal a biological instance of Goodhart’s law: when phenotypic measures become overly optimized targets, they cease to reliably promote genetic improvement. These findings highlight intrinsic limits to the effectiveness of strong selection and suggest that optimal evolutionary outcomes require intermediate selection strength in noisy environments.

Joint work with Colin Desmarais Bastien Mallein, Francesco Paparella and Emmanuel Schertzer.

Tang, Tiffany

University of Notre Dame

Session: Advances in Unsueprvised Learning & Causal Discovery

Talk: *Consensus dimension reduction via multi-view learning*

A plethora of dimension reduction methods have been developed to visualize high-dimensional data in low dimensions. However, different dimension reduction methods often output different and possibly conflicting visualizations of the same data. This problem is further exacerbated by the choice of hyperparameters, which may substantially impact the resulting visualization. To obtain a more robust and trustworthy dimension reduction output, we advocate for a consensus approach, which summarizes multiple visualizations into a single consensus dimension reduction visualization. Here, we leverage ideas from multi-view learning in order to identify the patterns that are most stable or shared across the many different dimension reduction visualizations, or views, and subsequently visualize this shared structure in a single low-dimensional plot. We demonstrate that this consensus visualization effectively identifies and preserves the shared low-dimensional data structure through both simulated and real-world case studies. We further highlight our method's robustness to the choice of dimension reduction method and hyperparameters---a highly-desirable property when working towards trustworthy and reproducible data science.

Tang, Yanbo

Imperial College London

Session: Structured Statistical Learning and Post-Learning Inference in High-Dimensional Settings

Talk: *Post-reduction inference for confidence sets of models*

In high-dimensional regression problems, especially in areas like genomics, we often face thousands of potential explanatory variables but only a limited number of observations. In these settings, it is often more informative to construct a set of plausible explanatory models on which follow-up studies can be based, rather than selecting a single model. Previous approaches to constructing such confidence sets of models have relied on sample splitting, but here we present an approach based on Fisher's ideas of conditional inference to avoid this. We illustrate the method with the normal linear regression model and show how it extends to time-to-event data and more general regression models.

Terveer, Sara

LMU Munich

Session: Random walks and interacting particles on random graphs

Talk: *Return probabilities on Bienaymé-Galton-Watson trees*

Consider a simple random walk on a Bienaymé-Galton-Watson tree. For its annealed return probability, an approach by Piau (1998) established upper bounds with subexponential decay in time. The bounds obtained this way for offspring distributions resulting in leaf-free trees are optimal and it has been conjectured that these are also the optimal bounds for general offspring distributions.

In this talk, we derive such optimal upper bounds for all offspring distributions with a finite first moment. We discuss the main techniques of the proof, specifically how to control regions of the tree with 'bad' isoperimetric properties, as well as consequences of the result, including an application to the spectral theory of Erdős–Rényi random graphs.

Tibshirani, Ryan

UC Berkeley

Session: Modern Optimization for Statistical Learning

Talk: *Gradient equilibrium, Blackwell approchability, and ergodicity*

Gradient equilibrium is a property of a sequence of iterates in online learning such that the average of gradients along the sequence converges to zero. In general, this property is achievable by standard online learning methods, such as gradient descent, under weak conditions; furthermore, in various problem settings it translates into statistically meaningful properties such as sequential notions of calibration and unbiasedness. This talk will review some of the basic theory of gradient equilibrium, and describe new connections to Blackwell approachability, and ergodic theory.

Tokdar, Surya

Duke University

Session: Empirical Bayes for approximate oracle inference

Talk: *Selective and marginal selective inference for exceptional groups*

Statistical analyses of multipopulation studies often use the data to select a particular population as the target of inference. For example, a confidence interval may be constructed for a population only in the event that its sample mean is larger than that of the other populations. We show that for the normal means model, confidence interval procedures that maintain strict coverage control conditional on such a selection event will have infinite expected width. For applications where such selective coverage control is of interest, this result motivates the development of procedures with finite expected width and approximate selective coverage control over a range of plausible parameter values. To this end, we develop selection-adjusted empirical Bayes confidence procedures that use information from the data to approximate an oracle confidence procedure that has exact selective coverage control and finite expected width. In numerical comparisons of the oracle and empirical Bayes procedures to procedures that only guarantee selective coverage control marginally over selection events, we find that improved selective coverage control comes at the cost of increased expected interval width.

Tuzhilina, Elena

University of Toronto

Session: Advances in Matrix Methods

Talk: *Smooth Zero-Inflated Modeling on Counting Tensors*

Inferring the three-dimensional organization of the genome from single-cell Hi-C data presents a major statistical challenge due to the extreme sparsity and noise inherent in these measurements. Each single-cell Hi-C experiment yields a contact matrix recording the frequency of spatial proximity between genomic loci. However, many entries are zeros, corresponding either to true biological absence of contact or to technical dropouts from incomplete ligation or limited sequencing depth. Distinguishing these sources is essential for accurate characterization of chromatin architecture at the single-cell level. We represent single-cell Hi-C data as a three-dimensional tensor encoding loci-by-loci contact maps across multiple cells. Building on this, we develop a tensor-based imputation framework using a zero-inflated Poisson model to distinguish true and false zeros and address data sparsity. Efficient parameter estimation procedures are derived, and the method is applied to real datasets to recover missing contacts and improve data quality across cell types.

van der Hofstad, Remco

Eindhoven University of Technology

Session: Statistical Inference on Preferential Attachment Networks

Talk: *Change-point detection in preferential attachment models*

Many phenomena in the real world can be phrased in terms of networks. Examples include the World-Wide Web, social interactions and Internet, but also the interaction patterns between proteins, food webs and citation networks. A very popular model for such networks is the preferential attachment model, in which vertices arrive in the network sequentially, and attach to the older vertices in the network such that higher-degree vertices are more likely chosen. In this talk, the focus is on the so-called affine preferential attachment model, where vertices attach according to the degree plus a constant. Such models have power-law degree distributions, alike those observed in real-world networks.

We discuss the problem of identifying a change point in the attachment mechanism, where the constant in the attachment rule changes at some time in the evolution of the network. We give conditions for when such a change can be detected, the precise condition being close to optimal.

[This joint work with Gianmarco Bet, Kay Bogerd and Rui Castro.]

van Engelenburg, Diederik

Technische Universität Wien

Session: Statistical mechanics

Talk: *One-arm exponents of the Ising model*

I will focus on behavior of the Ising model in high dimensions ($d \geq 4$).

Widom proposed that thermodynamic quantities follow power laws governed by critical exponents, and above the upper critical dimension $d_c = 4$, these exponents reduce to the mean-field values (matching those on trees or complete graphs). I will talk about a recent work about the so-called one-arm event (the origin connects to distance n) in the FK-Ising model,

We observe that this exponent, as well as the upper critical dimension, depends on the boundary condition: for wired boundary conditions, we prove that this probability decays up to constants as n^{-1} for $d \geq 4$ (by the Edwards-Sokal coupling, this probability is the magnetization at the origin), whereas in infinite volume we prove that it decays as n^{-2} for $d \geq 6$.

Based on joint work with Christophe Garban, Romain Panis and Franco Severo.

Van Keilegom, Ingrid

KU Leuven

Session: Recent advances in Dependence Modeling

Talk: *On an extension of the Cox model for time-varying covariates under dependent censoring with unknown association*

This paper proposes a copula-based Cox model to take dependent censoring into account in the presence of time-varying covariates. Unlike many existing approaches, the copula parameter defining the copula function is assumed to be unknown and is estimated from the observed survival data without requiring prior information. Although the standard partial likelihood method remains valid for estimating regression coefficients in Cox models with time-varying covariates, it is not suitable when the survival time and censoring time are dependent. To overcome this issue, we expand the nonparametric hazard function in the Cox model using M -spline basis functions. Then, the proposed model is estimated using both maximum likelihood and penalized likelihood methods. The latter method employs a penalty function to regularize the baseline hazard estimates, giving a numerically stable estimator of the variance–covariance matrix. Another advantage of our penalized likelihood approach is its reduced sensitivity to both the

number and placement of knots. The finite-sample performance of the proposed estimators is evaluated via simulation studies, and the methodology is further illustrated using a real data example.

Van Werde, Alexander

University of Muenster

Session: New Researchers Group Session

Talk: *On the spectral determinacy of random graphs*

How much about a network can be deduced from the eigenvalues of associated matrices? Is the graph uniquely characterized by its adjacency spectrum? Both positive and negative examples are known, but it remains poorly understood what happens in the typical case given by an Erdős–Rényi random graph. I will discuss some old and new conjectures as well as some new results for sparse random graphs.

Vantini, Simone

MOX - Dept of Mathematics, Politecnico di Milano

Session: Conformal prediction for non-exchangeable data

Talk: *Real-time anomaly detection in space-time processes via conformal prediction and functional data analysis*

Temporally variant data observed on two-dimensional domains arise naturally across several disciplines. Functional data analysis proves inherently suitable for representing and modelling this kind of data. Within this framework, discrete-time evolving surfaces can be effectively modelled as functional time series of random real-valued functions defined on a two-dimensional domain. Building upon this approach, an anomaly detection method for handling such data is here developed. The central focus revolves around conformal prediction for functional time series. The proposal extends a functional autoregressive process of order one to this context and incorporating conformal prediction bands for functional data recently introduced in Diquigiovanni et al (2025, Statistica Sinica) and Ajroldi et al (2023, CSDA). This methodology allows the real-time construction of a spatially and temporally varying prediction range for each point of the spatial domain ensuring control of the joint coverage probability. An anomaly is identified every time the observed surface deviates from the prediction range at a particular point in the domain. This approach inherently guarantees the exact control of the probability of encountering one or more false warnings across the spatial domain. Finally the procedure is applied to weekly satellite interferometric measures of land elevation speed in the Phlegraean Fields and Etna Volcano (Bortolotti et al 2024, International Astronautical Conference 2024).

Véber, Amandine

Université Paris Cité

Session: Random processes on graphs

Talk: *A "piecewise Gaussian" limit for the fluctuations of a stochastic spatial chemical reaction network*

In this presentation, we will introduce a measure-valued Markov process modelling a finite set of biochemical reactions taking place in a compact, continuous space. In general, molecules react when they are close to one another; certain reactions or chemical species may be localised in space; some chemical species are abundant (with an O(N) number of molecules) whilst others may be rare (with O(1) molecules). We shall begin by showing that, as N tends to infinity, a properly normalised version of the counting measure describing the state of the system converges to a piecewise deterministic Markov process with measure values. We shall then describe a central limit theorem associated with this convergence, in which the limiting process is a measure-valued semi-martingale whose noise part can be seen as a (conditionally) "piecewise" Gaussian process. Work in collaboration with Lea Popovic (Concordia University).

Vuursteen, Lasse

Duke

Session: Probability Foundations for Robust and Private Learning

Talk: *How Differential Privacy Reshapes Optimal Statistical Design*

Much of the differential privacy literature focuses on optimal inference within a fixed statistical model and design. But researchers typically face a choice among multiple study designs that address the same scientific question. For instance, to estimate the half-life of a drug, one must decide how many patients to enroll, how many measurements to take per patient, and when to take them.

I will present results showing that designs which are statistically equivalent in classical settings can perform drastically differently under differential privacy: some designs permit accurate private inference, while others fundamentally do not. This distinction is especially consequential for functional data, which arises naturally in EEG monitoring, wearable biosensors, and GPS tracking. We characterize the trade-offs between the number of individuals, the number and placement of measurements per individual, and whether the design is random or fixed. Our results reveal that privacy constraints fundamentally alter the optimal allocation of sampling resources — sometimes amplifying phenomena already present in the non-private setting, and sometimes breaking entirely from the classical picture.

Wang, Guanghui

Nankai University

Session: Recent developments in change-point analysis

Talk: *Distribution-Free and Model-Agnostic Changepoint Detection with Finite-Sample Guarantees*

We introduce ART, a distribution-free and model-agnostic framework for changepoint analysis with finite-sample guarantees. ART transforms independent observations into real-valued scores via a symmetric function; under the null hypothesis of no changepoint these scores are exchangeable. Ranking and aggregating the scores yields test statistics whose null distribution is known exactly from the permutation law of ranks, enabling exact finite-sample Type I error control without repeated refitting under permutations. ART extends naturally to a multi-scale setting: by locally ranking scores over a family of intervals and aggregating them, it supports multiple changepoint testing, localization with inference, and post-detection inference, while retaining distribution-free calibration. The approach is model-agnostic: it imposes minimal structural or distributional assumptions and accommodates diverse score constructions, including features learned by statistical or machine-learning models. Across simulations and real-data applications, ART delivers valid error control and competitive power across a range of models and distributions. These properties make ART a reliable and versatile tool for modern changepoint analysis.

Wang, Runmin

Texas A&M University

Session: Recent developments in change-point analysis

Talk: *High-dimensional Change-point Detection Using Generalized Homogeneity Metrics*

In this talk, we study the problem of detecting abrupt changes in the data-generating distributions of a sequence of high-dimensional observations beyond the first two moments. This problem has remained substantially less explored in the existing literature, especially in the high-dimensional context, compared to detecting changes in the mean or the covariance structure. We develop a nonparametric methodology to (i) test the existence of a change-point, and (ii) identify the change-point locations in an independent sequence of high-dimensional observations. Our approach rests upon recent nonparametric tests for the homogeneity of two high-dimensional distributions. We construct a single change-point test statistic based on a cumulative sum process in an embedded Hilbert space. We shall derive its limiting null distribution and present the asymptotic consistency under the high dimension medium sample size framework. We also combine our statistics with wild binary segmentation to recursively estimate and test for multiple change-point locations. The superior performance of our methodology compared to other existing procedures will be illustrated via extensive simulation studies and the application to the stock return data observed during the period of the global financial crisis in the United States.

Wang, Sven

EPFL

Session: Recent advances in statistics for stochastic processes

Talk: *On polynomial-time computation in PDE models*

In complex statistical models involving PDEs or stochastic process, likelihood-based methods may suffer from non-convex and possibly multimodal loss landscapes. Thus, optimization and Markov chain Monte Carlo (MCMC) methods may mix exponentially slowly. We discuss recent progress in devising statistical and polynomial-time computational guarantees in such settings. In particular, we will discuss a class of computationally tractable estimators which can be deployed in models where the likelihood function is implicitly defined via a PDE. For some prototypical non-linear inverse problems arising from elliptic PDEs, we prove that these estimators attain the best currently known statistical convergence rates while being globally computable in polynomial time. Our analysis is based on new generalized stability estimates, extending classical stability beyond the range of the forward operator, combined with tools from nonparametric M-estimation. Our estimators also provide principled warm-start initializations for polynomial-time Bayesian computation.

Wegkamp, Marten

Cornell University

Session: Structured and Expressive Models in High Dimensions

Talk: *A New Regression Lens on Multi-Class Classification*

Linear Discriminant Analysis (LDA) is a foundational method for classification. Its linear structure makes it easy to interpret, and it naturally accommodates multi-class problems. LDA is also closely connected to classical multivariate techniques, including Fisher's discriminant analysis, canonical correlation analysis, and linear regression.

In this talk, we deepen LDA's link to multivariate response regression by deriving an explicit relationship between discriminant directions and regression coefficients. This insight leads to a new regression-based procedure for multi-class classification that can flexibly incorporate structured, regularized, or even non-parametric regression methods. Unlike existing regression-based approaches, our formulation is particularly well-suited for establishing provable performance guarantees.

Wei, Waverly

University of Southern California

Session: Causality Meets AI: Generative Structure, LLM‚ÄëAssisted Trials, and Spatial Heterogeneity

Talk: *Can large language models boost the power of randomized experiments without statistical bias?*

Randomized experiments or randomized controlled trials (RCTs) are gold standards for causal inference, yet cost and sample-size constraints limit power. Meanwhile, modern RCTs routinely collect rich, unstructured data that are highly prognostic of outcomes but rarely used in causal analyses. We introduce CALM (Causal Analysis leveraging Language Models), a statistical framework that integrates large language models (LLMs) predictions with established causal estimators to increase precision while preserving statistical validity. CALM treats LLM outputs as auxiliary prognostic information and corrects their potential bias via a heterogeneous calibration step that residualizes and optimally reweights predictions. We prove that CALM remains consistent even when LLM predictions are biased and achieves efficiency gains over augmented inverse probability weighting estimators for various causal effects. In particular, CALM develops a few-shot variant that aggregates predictions across randomly sampled demonstration sets. The resulting U-statistic-like predictor restores i.i.d. structure and also mitigates prompt-selection variability. In simulations, CALM delivers lower variance relative to other benchmarking methods, is effective in zero- and few-shot settings, and remains stable across prompt designs. By principled use of LLMs to harness unstructured data and pretraining, CALM provides a practical path to more precise causal analyses in RCTs.

Wei, Yuting

University of Pennsylvania

Session: Inference Beyond the Textbook: Robustness, Adaptivity, and Cross-Fitting

Talk: *Statistical Inference under Adaptive Sampling with LinUCB*

Adaptively collected data has become ubiquitous in modern practice. Yet even seemingly benign adaptive sampling schemes can introduce severe biases, rendering traditional statistical inference tools inapplicable. Focusing on the linear bandit problem, a fundamental and influential framework in reinforcement learning and the bandit literature, we characterize the performance of LinUCB, a canonical upper-confidence-bound algorithm that balances exploration and exploitation, and derive inferential procedures that remain valid despite the challenges posed by adaptive data collection.

Willett, Rebecca

University of Chicago

Session: Statistical theory and algorithms for AI

Talk: *How do simple rotations affect the implicit bias of Adam?*

Adaptive gradient methods such as Adam and Adagrad are widely used in machine learning, yet their effect on the generalization of learned models -- relative to methods like gradient descent -- remains poorly understood. Prior work on binary classification suggests that Adam exhibits a "richness bias," which can help it learn nonlinear decision boundaries closer to the Bayes-optimal decision boundary relative to gradient descent. However, the coordinate-wise preconditioning scheme employed by Adam renders the overall method sensitive to orthogonal transformations of feature space. We show that this sensitivity can manifest as a reversal of Adam's competitive advantage: even small rotations of the underlying data distribution can make Adam forfeit its richness bias and converge to a linear decision boundary that is farther from the Bayes-optimal decision boundary than the one learned by gradient descent. To alleviate this issue, we show that a recently proposed reparameterization method -- which applies an orthogonal transformation to the optimization objective -- endows any first-order method with equivariance to data rotations, and we empirically demonstrate its ability to restore Adam's bias towards rich decision boundaries.

Wu, Wei

University of Chicago

Session: Statistical learning for spatial-temporal data

Talk: *Gaussian Approximation and Concentration of Constant Learning-Rate Stochastic Gradient Descent*

We establish a comprehensive finite-sample and asymptotic theory for stochastic gradient descent (SGD) with constant learning rates. First, we propose a novel linear approximation technique to provide a quenched central limit theorem (CLT) for SGD iterates with refined tail properties, showing that regardless of the chosen initialization, the fluctuations of the algorithm around its target point converge to a multivariate normal distribution. Our conditions are substantially milder than those required in the classical CLTs for SGD, yet offering a stronger convergence result. Furthermore, we derive the first Berry-Esseen bound – the Gaussian approximation error – for the constant learning-rate SGD, which is sharp compared to the decaying learning-rate schemes in the literature. Beyond the moment convergence, we also provide the Nagaev-type inequality for the SGD tail probabilities by adopting the autoregressive approximation techniques, which entails non-asymptotic largedeviation guarantees. These results are verified via numerical simulations, paving the way for theoretically grounded uncertainty quantification, especially with non-asymptotic validity

Wu, Yuchen

Cornell University

Session: Advances in Stochastic Gradient Descent: Algorithms and Asymptotics

Talk: *Provably Reliable Classifier Guidance via Cross-Entropy Control*

Classifier-guided diffusion models generate conditional samples by augmenting the reverse-time score with the gradient of the log-probability predicted by a probabilistic classifier. In practice, this classifier is usually obtained by minimizing an empirical loss function. While existing statistical theory guarantees good generalization performance when the sample size is sufficiently large, it remains unclear whether such training yields an effective guidance mechanism. We study this question in the context of cross-entropy loss, which is widely used for classifier training. Under mild smoothness assumptions on the classifier, we show that controlling the cross-entropy at each diffusion model step is sufficient to control the corresponding guidance error. In particular, probabilistic classifiers achieving conditional KL divergence ε2 induce guidance vectors with mean squared error O˜(dε), up to constant and logarithmic factors. Our result yields an upper bound on the sampling error of classifier-guided diffusion models and bears resemblance to a reverse log-Sobolev--type inequality. To the best of our knowledge, this is the first result that quantitatively links classifier training to guidance alignment in diffusion models, providing both a theoretical explanation for the empirical success of classifier guidance, and principled guidelines for selecting classifiers that induce effective guidance.

Wylomanska, Agnieszka

Wroclaw University of Science and Technology

Session: Recent Advances in Distribution Theory

Talk: *A modified Greenwood statistic for goodness-of-fit testing in heavy-tailed multivariate samples*

This presentation details an enhanced version of the Greenwood statistic, specifically tailored for analyzing complex, multivariate datasets. The study addresses the challenges of "heavy-tailed" data, i.e. where extreme outliers are frequent, common in distributions such as Pareto, Student’s t, and alpha-stable. By exploring the stochastic properties of this modified statistic, the research demonstrates its superior performance in goodness-of-fit testing. A major breakthrough of this work is the statistic’s ability to determine if a sample originates from a distribution with infinite variance, a recurring hurdle in modern data science. The methodology's reliability is confirmed via Monte Carlo simulations and further proven through the analysis of real data sets.

Xia, Lucy

The Hong Kong University of Science and Technology

Session: The many facets of score matching

Talk: *Is the F-test doubly robust?*

We study the robustness of the F-test in random design linear models, and reach a somewhat nuanced conclusion. On the positive side, one of our main results is that the size of the test is close to its nominal level as soon as either the distribution of the normalised error vector is close to uniform on the unit sphere, or the distribution of the design matrix, after applying any column space-preserving orthogonalisation scheme (e.g.~Gram--Schmidt), is close to a Haar distribution. This provides a sense in which the F-test is doubly robust. Our conclusion is reached by establishing a H\"older continuity property of the Kolmogorov distance between the distribution of the F-statistic and its notional F-distribution under the null. Writing n, p and p_0 for the sample size and the dimensions of the full and null models respectively, we prove that the H\"older exponent is $1/3$ when $\min(n-p,p-p_0) = 1$ and $1/2$ when $\min(n-p,p-p_0) \geq 2$. On the other hand, these exponents are relatively small and cannot be improved in general, revealing that the size of the test may depart from its nominal level quite quickly as we move away from settings where the test is exact. In particular, the regression t-test for the significance of a single predictor corresponds to the case $p - p_0 = 1$, so this test may be especially vulnerable to model misspecification.

Xie, Minge

Rutgers University

Session: Modern Approaches to Inference and Estimation

Talk: *Model-Free Inference for High-Dimensional Binary Classification Using Repro Samples*

Performing inference is particularly challenging in high-dimensional settings when the underlying data-generating regression structure is completely unknown and unspecified. In this talk, we develop a model-free inference approach for high-dimensional binary classification, leveraging a class of (likely misspecified) sparsity GLMs. Our method builds on the repro samples framework, which generates artificial samples that mimic the true data or noises. The proposed approach facilitates inference by targeting both the model support and the regression coefficients of the oracle model, defined as the working model closest to the unknown true model. The method offers three key advantages: (1) it is fully model-free, requiring neither correct model specification nor sparsity of the true model; (2) it constructs a candidate set of influential covariates with guaranteed coverage under weak signal conditions; and (3) it provides confidence sets for any linear transformation of the oracle coefficients. When the oracle model coincides with the true underlying model, the inference results directly apply to the true model, allowing simultaneous quantification of uncertainty in (discrete) model selection and (continuous) parameter estimation. Simulation studies illustrate the effective performance of the proposed method. An application to single-cell RNA-seq immune response data demonstrates that it identifies key genes, offering new insights into cellular immune response mechanisms.

Xie, Yao

Georgia Institute of Technology

Session: Bridging Statistics and Modern AI: Foundations for Deep Learning and Generative Models

Talk: *Local Flow Matching Generative Models: Stepwise construction of flows via diffusion-inspired dynamics*

A central goal in statistics is to represent and construct complex probability distributions from data, for uncertainty quantification as well as conditional, predictive, and generative inference. Generative models offer a flexible framework for this task by learning maps from a simple reference law to a target distribution. Among them, flow matching is particularly appealing because it learns a continuous and invertible transport between distributions, combining computational tractability with a clear probabilistic interpretation. In this talk, I introduce Local Flow Matching (LFM), a stepwise extension of flow matching motivated by the gradient-flow structure of diffusion processes. In contrast to a single global flow from noise to data, LFM decomposes a difficult long-range transport problem into a sequence of local flow-matching subproblems, each connecting distributions that are closer to one another. This yields a more localized approximation of the overall transport, simplifies estimation at each stage, permits smaller function classes, and enables sharper analysis of the discrepancy between the generated and target distributions. By exploiting the contraction property of the diffusion process, we establish a generation guarantee in terms of the chi-square-divergence. Empirical results on tabular, image, and conditional generation tasks demonstrate competitive performance and improved training efficiency.

Xu, Gongjun

University of Michigan

Session: Statistical Learning with Noisy Labels, Privacy Perturbations, and Latent Structures

Talk: *Identifiability and Inference for Generalized Latent Factor Models*

Generalized latent factor analysis not only provides a useful latent embedding approach in statistics and machine learning, but also serves as a widely used tool across various scientific fields, such as psychometrics, econometrics, and social sciences. Ensuring the identifiability of latent factors and the loading matrix is essential for the model's estimability and interpretability, and various identifiability conditions have been employed by practitioners. However, fundamental statistical inference issues for latent factors and factor loadings under commonly used identifiability conditions remain largely unaddressed, especially for correlated factors and/or non-orthogonal loading matrix. In this work, we focus on the maximum likelihood estimation for generalized factor models and establish statistical inference properties under popularly used identifiability conditions. The developed theory is further illustrated through numerical simulations and an application to a personality assessment dataset.

Xu, Min

Rutgers University

Session: Statistical inference for data science and AI

Talk: *Optimal Convex M-Estimation via Score Matching*

In the context of linear regression, we construct a data-driven convex loss function with respect to which empirical risk minimisation yields optimal asymptotic variance in the downstream estimation of the regression coefficients. At the population level, the negative derivative of the optimal convex loss is the best decreasing approximation of the derivative of the log-density of the noise distribution. This motivates a fitting process via a nonparametric extension of score matching, corresponding to a log-concave projection of the noise distribution with respect to the Fisher divergence. At the sample level, our semiparametric estimator is computationally efficient, and we prove that it attains the minimal asymptotic covariance among all convex M-estimators. As an example of a non-log-concave setting, the optimal convex loss function for Cauchy errors is Huber-like, and our procedure yields asymptotic efficiency greater than 0.87 relative to the maximum likelihood estimator of the regression coefficients that uses oracle knowledge of this error distribution. In this sense, we provide robustness and facilitate computation without sacrificing much statistical efficiency. Numerical experiments using our accompanying R package "asm" confirm the practical merits of our proposal.

Xu, Yunbei

National University of Singapore

Session: Advances in Learning, Optimization, and Generalization in Modern Stochastic Systems

Talk: *Pointwise Generalization in Deep Neural Networks*

We address the fundamental question of why deep neural networks generalize by establishing a pointwise generalization theory for fully connected networks. This framework resolves long-standing barriers to characterizing the rich, nonlinear feature-learning regime and builds a new statistical foundation for representation learning. For each trained model, we characterize the hypothesis via a pointwise Riemannian Dimension, derived from the eigenvalues of the learned feature representations across layers. This establishes a principled framework for deriving hypothesis-dependent, spectrum-aware generalization bounds. These bounds offer a systematic upgrade over approaches based on model size, products of norms, and infinite-width linearizations, yielding guarantees that are orders of magnitude tighter in both theory and experiment. Analytically, we identify the structural properties and mathematical principles that explain the tractability of deep networks. Empirically, the pointwise Riemannian Dimension exhibits substantial feature compression, decreases with increased over-parameterization, and captures the implicit bias of optimizers. Taken together, our results indicate that deep networks are mathematically tractable in practical regimes and that their generalization is sharply explained by pointwise, spectrum-aware complexity.

Yi, Grace

University of Western Ontario

Session: Statistical Learning with Noisy Labels, Privacy Perturbations, and Latent Structures

Talk: *Causal Learning of Paired Vectors with Label Noise: Impact and Correction Methods*

Causal inference involves determining whether a cause-effect relationship exists between two sets of interest, a task that can be framed as a binary classification problem. When dealing with a sequence of independent and identically distributed paired vectors, the kernel mean embedding of the probability distribution can be utilized to map the empirical distribution to a feature space. Subsequently, a classifier is trained in this feature space to predict causation for future vector pairs. However, this approach is susceptible to mislabeling of causal relationships, a common challenge in causation studies. In this talk, I will discuss the impact of mislabeled outputs on the training results. Moreover, I will present robust learning methods that take into account the mislabeling effects and offer theoretical justifications for the validity of these proposed methods.

Young, Elliot

University of Cambridge

Session: Royal Statistical Society Session

Talk: *Outrigger local polynomial regression*

Standard local polynomial estimators of a nonparametric regression function employ a weighted least squares loss function that is tailored to the setting of homoscedastic Gaussian errors. We introduce the outrigger local polynomial estimator, which is designed to achieve distributional adaptivity across different conditional error distributions. It modifies a standard local polynomial estimator by employing an estimate of the conditional score function of the errors and an 'outrigger' that draws on the data in a broader local window to stabilise the influence of the conditional score estimate. Subject to smoothness conditions, and only requiring consistency of the conditional score estimate, we establish that even under the least favourable settings for the outrigger estimator, the asymptotic ratio of the worst-case local risks of the two estimators is at most 1, with equality if and only if the conditional error distribution is Gaussian. Moreover, we prove that the outrigger estimator is minimax optimal over Hölder classes up to a multiplicative factor depending only on the smoothness of the regression function and the dimension of the covariates. For Lipschitz functions this factor is shown to be at most 1.69, and in low smoothness or high dimension regimes we approach exact constants. A further attraction of our proposal is that we do not require structural assumptions such as independence of errors and covariates, or symmetry of the conditional error distribution.

Zaffran, Margaux

Inria, Laboratoire de Mathématiques d'Orsay

Session: Conformal prediction for non-exchangeable data

Talk: *On the hardness of group-conditional distribution-free predictive inference*

Predictive uncertainty quantification is crucial in decision-making problems.In this talk, we will focus on distribution-free uncertainty quantification by considering predictive intervals for the target Y enjoying validity (i.e. nominal coverage) with no assumptions on the underlying data generating process nor the sample size.

After introducing the framework, we will detail the nuance between marginal validity and conditional---on the test point---validity. We will review the existing (impossibility) results on conditional validity.

This will lead us to our main question: how can we relax the goal of conditional validity to make it achievable? We will present new hardness results, that characterize the limits of group conditional coverage (e.g., achieving nominal coverage not only on average but also among women on the one hand, and among men on the other hand), a weaker goal extensively used in the literature in place of the impossible perfect conditional validity.

Finally, we will dive into applying these results in the context of prediction with missing values. There, one wants to obtain not only marginally valid intervals despite missing values, but also intervals that achieve the nominal coverage regardless of which values are missing at test time. We provide an algorithm reaching this goal by constructive informative predictive intervals in the light of our hardness results.

Zhang, Jiechen

EPFL

Session: Probability Distributions via Moments and Cumulants

Talk: *Moments and Cumulants in Probability Theory: Recent Results and Perspectives*

Moments and cumulants are two fundamental sequences of a random variable, closely related but often playing different roles. They are central in probability theory, especially in questions of determinacy and in limit theorems. Classical topics include moment determinacy versus indeterminacy, the Gaussian law characterized by vanishing higher cumulants, the Fréchet--Shohat theorem, and Wigner's semicircle law. There are also parallel cumulant-based methods, which have proved very effective.

I will present two recent and somewhat unexpected results. The first is a nonconventional limit theorem which is genuinely cumulant-based and cannot be accessed effectively by moment methods alone. The second says that a moment-indeterminate random variable can be decomposed as a sum of two random variables each satisfying Carleman's condition, hence each moment-determinate.

My own work focuses on what more, beyond the classical formulas, can be said when one passes from moments to cumulants, and vice versa. In direction moments⇝cumulants', I will present recent explicit and distribution-free bounds with factorial rather than superexponential growth. In direction cumulants⇝moments', I will explain both what is possible and what is not, and discuss a certain `gap'.

I will also mention another perspective, related to M-indeterminate distributions and their Stieltjes classes, showing that moment equivalence may be invisible to polynomial tests yet still matter for other natural functionals.

Zhang, Xiaozhu

University of Washington

Session: Modern Optimization for Statistical Learning

Talk: *Convex Mixed-Integer Programming for Causal Additive Models with Optimization and Statistical Guarantees*

Learning causal structure from nonlinear data remains computationally and statistically challenging. We address this problem for additive, nonlinear structural equation models with Gaussian noise. By expressing each non-linear function through a basis expansion, we formulate a maximum likelihood estimator with a group l0 penalty that directly controls graph sparsity. Despite the combinatorial nature of the problem, the estimator can be solved efficiently via a convex mixed-integer program combined with a novel outer approximation scheme, yielding globally optimal solutions within seconds. Our framework can incorporate prior knowledge such as edge constraints and partial orders. We prove consistency for graph recovery when the number of variables grows with sample size, and connect optimization guarantees to statistical error, yielding an early stopping rule that preserves consistency while reducing computation. Compared to methods based on linearity, equal noise variances, or heuristics, our approach provides a unified framework with provable optimality, efficiency, and statistical guarantees. Experiments demonstrate substantial improvements in graph recovery on both simulated and real data.

This is joint work with Nir Keret, Ali Shojaie, and Armeen Taeb.

Zhou, Kangjie

Columbia University

Session: Generative Models for Statistical Inference

Talk: *Statistical Inference in an Interactive Learning Paradigm*

The proliferation of generative artificial intelligence has given rise to an interactive learning paradigm, where model parameters are continuously updated using not only data generated by natural processes, but also synthetic outputs produced by other models. This paradigm introduces two major challenges: (1) training data are no longer drawn exclusively from the target population, undermining a core assumption of classical statistical learning theory, and (2) model training processes become inherently correlated, as models influence one another through repeated exposure to each other's synthetic outputs. Establishing reliable statistical inference in such interactive environments therefore remains an important open problem. In particular, there is growing concern about model collapse, a phenomenon in which the performance of generative models progressively degrades as they are trained on synthetic data produced by earlier model generations.

In this talk, we study the behavior of generative models under general interaction patterns.

We formalize these interactions using directed graphs and show that the occurrence of model collapse depends critically on the graph's topology. Within this framework, we derive an explicit necessary and sufficient condition characterizing when model collapse occurs.

Ziedins, Ilze

University of Auckland

Session: Stochastic Networks

Talk: *Priority queues with differing customer selection rules*

We consider priority queues where each class may have its own selection rule (such as first in first out, or last in first out). We show that the customer waiting time distribution for a given class is insensitive to the customer selection rules used for the other priority classes, and depends only on the selection rule for that class. The sample path argument used to show this also yields a general approach for finding waiting time distributions in priority queues. This is joint work with Peter Taylor (University of Melbourne) and David Stanford (Western University).